**CHAPTER ONE**

**INTRODUCTION**

* 1. **BACKGROUND OF STUDY**

The history of COVID-19 traces back to December 2019 when cases of a novel respiratory illness were first reported in the city of Wuhan, Hubei province, China. Here's a chronological overview of key events and milestones in the history of COVID-19:

December 2019, A cluster of pneumonia cases of unknown origin is reported in Wuhan, China. Chinese health authorities investigate the cases to identify the cause of the illness.

January 2020, On January 7, 2020, Chinese health authorities confirm that the cluster of pneumonia cases is caused by a novel coronavirus, later named SARS-CoV-2. The World Health Organization (WHO) is notified about the outbreak. The first death from COVID-19 is reported in China. The virus spreads to other parts of China and to other countries, with the first confirmed cases outside China reported in Thailand, Japan, and South Korea. On January 23, 2020, Wuhan and other cities in Hubei province are placed under lockdown to contain the outbreak.

February 2020, COVID-19 cases continue to increase globally, with outbreaks reported in various countries, including Iran, Italy, and South Korea. The WHO declares COVID-19 a Public Health Emergency of International Concern on February 30, 2020.

March 2020, The WHO officially declares COVID-19 a pandemic on March 11, 2020, as the virus spreads to almost every country in the world. Many countries implement travel restrictions, social distancing measures, and lockdowns to control the spread of the virus. Testing and diagnostic efforts intensify, and research accelerates to understand the virus and develop treatments and vaccines. April to June 2020The global number of COVID-19 cases continues to rise, with significant outbreaks in the United States, Europe, and South America. Healthcare systems face immense pressure, and the demand for medical supplies and equipment increases. Efforts to develop vaccines and treatments advance rapidly.

July to December 2020, Several vaccine candidates enter clinical trials, and some countries begin vaccination campaigns in late 2020 after regulatory approval of vaccines. The virus continues to circulate, leading to successive waves of infections in various regions. COVID-19's impact on global economies and livelihoods remains a major concern. Vaccination campaigns expand globally, providing hope for controlling the pandemic. New variants of the virus emerge, some of which are more transmissible or potentially evade immunity from previous infections or vaccinations. In some countries, efforts to administer booster vaccine doses begin to enhance protection against new variants.

* 1. **STATEMENT OF THE PROBLEM**
  2. **AIMS AND OBJECTIVES**

**Aim:**

The aim of COVID-19 visualization and prediction is to provide valuable insights into the spread of the COVID-19 pandemic, visualize the trends, patterns, and geographical distribution of cases, and develop predictive models to forecast the future trajectory of infections, recoveries, and deaths.

**Objectives:**

**1. Data Collection and Preprocessing:**

1. Gather and collate reliable and up-to-date COVID-19 data from various sources, including health organizations, governments, and research institutions.
2. Preprocess and clean the data to remove inconsistencies, missing values, and anomalies for accurate analysis.

**2. Visualizing Temporal Trends:**

1. Create time series plots and visualizations to illustrate the daily, weekly, or monthly trends of COVID-19 cases, recoveries, and deaths.
2. Identify potential spikes, trends, and periodic patterns in the data to gain insights into the evolution of the pandemic over time.

**3. Geospatial Mapping:**

1. Develop geospatial visualizations, such as choropleth maps, to display the distribution of COVID-19 cases across different regions, countries, or continents.
2. Highlight areas with high infection rates to identify hotspots and inform public health interventions.

**4. Epidemiological Analysis:**

* 1. Apply epidemiological models (e.g., SIR, SEIR) to understand the dynamics of virus transmission and predict future infection rates based on current data.
  2. Assess the impact of interventions, such as lockdowns or vaccination campaigns, on flattening the infection curve.

**5. Machine Learning Prediction:**

* 1. Build and train machine learning models, such as time series forecasting or regression, to predict future trends of COVID-19 cases, recoveries, and deaths.
  2. Evaluate and validate the performance of prediction models using appropriate metrics.

**6. Uncertainty Analysis:**

* 1. Account for uncertainties in the data and predictions due to factors like testing variations, reporting delays, and model inaccuracies.
  2. Provide confidence intervals or probability distributions to convey the uncertainty of predictions.

**7. Communication and Public Awareness:**

* 1. Present the visualization and prediction results in an accessible and understandable format for policymakers, healthcare professionals, and the general public.
  2. Communicate important findings and recommendations to facilitate informed decision-making and public awareness.

**8. Adaptive Model Updating:**

* 1. Continuously update the prediction models as new data becomes available to enhance the accuracy and relevance of predictions.
  2. Incorporate the latest information on virus mutations, vaccination rates, and public health measures into the models.

**9. Policy Guidance:**

* 1. Use the insights gained from visualization and prediction to inform public health policies, resource allocation, and emergency response strategies.
  2. Help guide the allocation of medical resources and vaccination campaigns based on the projected demands.

**1.4 SIGNIFICANCE OF THE STUDY**

The significance of the study on COVID-19 visualization and prediction is multifaceted and holds great importance in several aspects:

* + 1. Public Health Reaction: The findings of the study can give useful insights into the existing and future tendencies of the COVID-19 pandemic. This information may be used by policymakers and healthcare professionals to develop effective public health initiatives, allocate medical resources, and take targeted steps to prevent the virus's spread.
    2. Early Warning System: Accurate COVID-19 case, recovery, and death forecasts can serve as an early warning system. This enables authorities to plan for anticipated infection surges, adapt healthcare capacity, and make educated decisions to safeguard vulnerable people.
    3. Resource deployment: By forecasting future infection rates and hotspot locations, the study can help with the effective deployment of medical supplies, hospital beds, and other resources to high-risk areas. This prevents healthcare systems from getting overburdened.
    4. Policy formation: The visualization and prediction outcomes of the study can have an impact on evidence-based policy formation. These findings may be used by policymakers to plan and execute actions that are customized to the unique requirements of certain areas, communities, or demographic groupings.
    5. Epidemiological Understanding: The research might help researchers better understand the virus's transmission dynamics, the impact of therapies, and the factors that influence infection rates. This information is critical for improving disease modeling and directing future pandemic response efforts.
    6. Public Awareness and Education: The visualizations and projections from the study may be shared with the general public, increasing awareness about the severity of the pandemic and the significance of following preventive measures. It can also help people comprehend the progress of the epidemic and the reasoning for public health guidelines.
    7. Vaccination plans: By anticipating vaccine demand in different locations and assessing the possible impact of vaccination campaigns on infection rates and general public health, the study can assist influence vaccination plans.
    8. To summarize, the study's importance stems from its ability to facilitate decision-making, improve public health response, and deepen our understanding of the COVID-19 epidemic. It helps to restrict the spread of the virus, preserve public health, and lessen the worldwide social, economic, and health effects of the pandemic.

**1.5 SCOPE OF THE STUDY**

The scope of the study on COVID-19 visualization and prediction is vast and encompasses various aspects related to the pandemic

**1.6 LIMITATIONS OF THE STUDY**

The study on COVID-19 visualization and prediction may confront numerous constraints, which may effect the findings' accuracy, generalizability, and completeness. Some potential constraints are as follows:

1. Data Availability and Quality: The quality of the data utilized has a significant impact on the accuracy and dependability of the forecasts. Data may be missing, erroneous, or delayed, resulting in potential biases in the study.
2. COVID-19 is a pandemic that is fast changing, and new variations, public health initiatives, or immunization techniques may arise throughout the research period. Unforeseen occurrences that were not accounted for in the models may have an impact on predictions.
3. Long-term projections may be subject to more uncertainty because to the unpredictability of external influences and viral dynamics. Accurate long-term forecasting may be difficult.
4. Model Assumptions: Predictive models frequently rely on certain assumptions, and the accuracy of the forecasts may be affected by these assumptions. Incorrect assumptions might result in skewed outcomes.
5. Data Representativeness: The data utilized for analysis may not completely reflect the total population, particularly in areas where testing is restricted or underreported. This may result in sample biases.
6. Model Overfitting: Overfitting can occur in machine learning models, especially if the dataset is short or noisy. Overfitting can result in good training data performance but poor generalization to new, unknown data.

**1.7 DEFINITION OF TERMS**

Here are definitions of some key terms related to COVID-19 visualization and prediction:

1. **COVID-19:** COVID-19, short for "Coronavirus Disease 2019," is a highly contagious respiratory illness caused by the novel coronavirus SARS-CoV-2. It was first identified in December 2019 in Wuhan, China, and has since evolved into a global pandemic.
2. **Visualization:** Visualization refers to the graphical representation of data or information, designed to provide insights, patterns, and trends in a clear and understandable manner. In the context of COVID-19, visualization involves creating charts, graphs, maps, and other visual representations to depict the spread of the virus and its impact on various populations.
3. **Prediction:** Prediction is the process of using historical data and mathematical models to estimate future trends or outcomes. In the context of COVID-19, prediction involves forecasting the number of cases, recoveries, deaths, or other pandemic-related variables based on past data and predictive algorithms.
4. **Time Series Analysis:** Time series analysis is a statistical method used to analyze and interpret data points collected over time. In COVID-19 research, time series analysis is used to identify patterns, seasonality, and trends in the daily, weekly, or monthly COVID-19 data.
5. **Epidemiological Models:** Epidemiological models are mathematical models used to study the spread and impact of infectious diseases in a population. Common models used for COVID-19 include the SIR (Susceptible-Infected-Recovered) and SEIR (Susceptible-Exposed-Infected-Recovered) models.
6. **Geospatial Mapping**: Geospatial mapping involves using geographical coordinates to visualize data on maps. In the context of COVID-19, geospatial mapping is used to show the distribution of cases, hotspots, and the spread of the virus across different regions and countries.
7. **Machine Learning:** Machine learning is a subset of artificial intelligence that involves developing algorithms and models that can learn patterns from data and make predictions or decisions without being explicitly programmed. In COVID-19 research, machine learning is used for prediction tasks and identifying patterns in the data.
8. **Uncertainty Analysis:** Uncertainty analysis involves quantifying and characterizing the uncertainty associated with predictions or model outputs. In the context of COVID-19, it helps understand the potential variability and limitations in predictive models.
9. **Policy Formulation:** Policy formulation refers to the process of developing strategies, guidelines, and interventions to address specific issues. In the context of COVID-19, policy formulation is based on insights from visualization and prediction to implement effective measures to control the spread of the virus and protect public health.

**CHAPTER TWO**

**LITERATURE REVIEW**

**2.1 INTRODUCTION**

COVID-19, short for Coronavirus Disease 2019, is a highly contagious respiratory illness caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The virus was first identified in December 2019 in Wuhan, Hubei province, China, and has since spread globally, leading to a pandemic.

**2.2 HISTORY OF COVID 19**

The novel human coronavirus disease 2019 (COVID-19) was first reported in Wuhan, China, in 2019, and subsequently spread globally to become the fifth documented pandemic since the 1918 flu pandemic.

By September 2021, almost two years after COVID-19 was first identified, there had been more than 200 million confirmed cases and over 4.6 million lives lost to the disease. Here, we take an in-depth look at the history of COVID-19 from the first recorded case to the current efforts to curb the spread of the disease with worldwide vaccination programs.

The first official cases of COVID-19 were recorded on the 31st of December, 2019, when the World Health Organization (WHO) was informed of cases of pneumonia in Wuhan, China, with no known cause. On the 7th of January, the Chinese authorities identified a novel coronavirus, temporally named 2019-nCoV, as the cause of these cases.

Weeks later, the WHO declared the rapidly spreading COVID-19 outbreak as a Public Health Emergency of International Concern on the 30th of January 2020. It wasn’t until the following month, however, on the 11th of February that the novel coronavirus got its official name - COVID-19. Nine days later, the US Centers for Disease Control and Prevention (CDC) confirmed the first person to die of COVID-19 in the country. The individual was a man in his fifties who lived in Washington state.

Over the summer, many countries saw a drop in cases, hospitalizations, and deaths due to the restrictions their citizens had endured to prevent the spread of the virus. However, towards the end of the summer, in August of 2020, the Lambda variant was first discovered in Peru. To date, this variant has since spread to at least 29 countries, according to the WHO.

A month later, the Alpha variant was first identified in the UK in September 2020. The discovery of these variants was significant, it showed that the virus was evolving. As a result, symptoms and disease outcomes were changing. Evidence has shown, for example, that the Alpha variant may pose a heightened risk of poor COVID-19 outcomes.

With the emergence of these new variants, cases of COVID-19 began to rise again in many countries and by the 29th of September 2020, there had been 1 million COVID-19 deaths.

Vaccinations were developed in record time. On the 9th of November, trials demonstrated the Pfizer and BioNTech vaccines to be over 90% effective, and the Moderna vaccine was proved to also be effective just a week later on the 16th of November. One more week later, on the 23rd of November, the University of Oxford and AstraZeneca COVID-19 was also shown to be effective.

Shortly after, the Delta variant was first discovered in December in India. Concerns over the potential increased transmissibility of the variants, fueled by a rise in cases in some counties such as the UK, forced many governments to once again reinforce lockdown measures to some extent.

Finally, on the 31st of December 2020, the WHO issued its first emergency use validation for a COVID-19 vaccine, making the Pfizer/BioNTech vaccine the first to be available for use. The emergency validation was seen as a positive step towards making COVID-19 vaccines globally available - a necessary step to ending the pandemic.

Since then, the Moderna vaccine and the Oxford/AstraZeneca vaccine have also been approved for use and national vaccine rollout initiatives have begun with full force. As of the 27th of April, 2021, 1 billion COVID-19 vaccine doses have been administered. The continued roll-out of vaccines in all countries is vital to bringing the pandemic under control and preventing future outbreaks.

Much can be learned from the story of the COVID-19 pandemic, and many hope lessons learned will prepare us for future infectious disease outbreaks and prevent potential future pandemics.

**2.3 TYPES OF COVID 19**

SARS-CoV-2, the virus that causes COVID-19, is continually changing. As HPV mutates, new forms that are more transmissible or cause more severe sickness may develop.

The World Health Organization (WHO) has classified five SARS-CoV-2 mutations as variations of concern (VOCs). These mutations are especially concerning since they have been found to be more transmissible, to produce more severe disease, or to elude vaccination protection.

1. As of August 6, 2023, the following VOCs have been recognized by the WHO:
2. Alpha (B.1.1.7)
3. Beta (B.1.351)
4. Gamma (P.1)
5. Delta (B.1.617.2)
6. Omicron (B.1.1.529)

The most recent VOC to appear is the Omicron variety. It is very contagious and has been demonstrated to resist vaccination protection to some extent. It is unknown if Omicron produces more severe sickness than other variations.

The WHO is continuing to study SARS-CoV-2 development and may classify other types as VOCs in the future. In addition to the VOCs, there are a variety of additional SARS-CoV-2 variants that are circulating. Although these versions are not as dangerous as VOCs, they can still cause disease.

**2.4 USES OF PREDICTIVE MODEL USED IN COVID 19 ANALYSIS**COVID-19 data may be analyzed using predictive models in a variety of ways, including:

Predicting the number of cases and fatalities. The number of new cases and deaths that are expected to occur in the future can be estimated using predictive models. This data may be used to assist governments and health groups in determining how to allocate resources and respond to the epidemic.

Identifying vulnerable populations. Predictive algorithms can be used to identify groups that are more likely to develop COVID-19 or suffer from severe sickness. This data can be used to tailor public health initiatives to these groups.

Understanding the spread of the virus. Predictive models can be used to understand how the virus is spreading and how it is likely to spread in the future. This information can be used to develop strategies to control the spread of the virus.

Evaluating the effectiveness of interventions. Predictive models can be used to evaluate the effectiveness of interventions, such as social distancing, mask-wearing, and vaccination, in reducing the spread of the virus. This information can be used to inform decisions about how to allocate resources and respond to the pandemic.

In addition to these specific uses, predictive models can also be used to gain a better understanding of the COVID-19 pandemic and how it is affecting different populations. This information can be used to inform public policy and help to protect people from the virus.

1. 1 Here are some instances of how COVID-19 data has been analyzed using prediction models:
2. 2 According to a study conducted by experts at the University of Oxford, the number of COVID-19 cases in the United Kingdom would peak at roughly 200,000 per day in April 2020, based on a prediction model. The real daily peak was roughly 180,000.
3. 3 A prediction algorithm developed by researchers at the University of California, Berkeley was used to identify groups with a higher risk of getting COVID-19. The study discovered that persons of race, those with underlying health concerns, and those living in poverty were more likely to get the virus.
4. A study by researchers at the University of Washington used a predictive model to evaluate the effectiveness of social distancing in reducing the spread of COVID-19. The study found that social distancing was effective in reducing the spread of the virus, but it was not enough to prevent a large outbreak.

**2.5 APPLICATION OF COVID 19 VISUALIZATION AND ANALYSIS**

The visualization and analysis of COVID-19 may be utilized for a variety of reasons, including:

1. Virus tracking: Visualizations may be used to track the number of cases, fatalities, and hospitalizations over time. This data may be used to identify places where outbreaks are occurring and to track the success of treatments.
2. Identifying at-risk populations: Visualizations can be used to identify populations that are more likely to develop COVID-19 or suffer from severe sickness. This data can be used to tailor public health initiatives to these groups.
3. Understanding the virus's impact: Visualizations may be utilized to comprehend the virus's influence on diverse populations. This data may be utilized to inform public policy and safeguard citizens. Evaluating the effectiveness of interventions: Visualizations can be used to evaluate the effectiveness of interventions, such as social distancing, mask-wearing, and vaccination, in reducing the spread of the virus. This information can be used to inform decisions about how to allocate resources and respond to the pandemic.
4. Communicating with the public: Visualizations can be used to communicate the risks of COVID-19 and the importance of following public health guidelines. This information can be used to help people understand the pandemic and take steps to protect themselves and others.
5. Here are some examples of how COVID-19 visualization and analysis have been used:
6. The Johns Hopkins Coronavirus Resource Center has created a number of interactive visualizations that track the spread of the virus around the world.
7. The Centers for Disease Control and Prevention (CDC) has created a number of visualizations that track the impact of the virus on different populations in the United States.
8. The World Health Organization (WHO) has created a number of visualizations that track the global spread of the virus.

**2.6 CHALLENGES OF COVID 19**

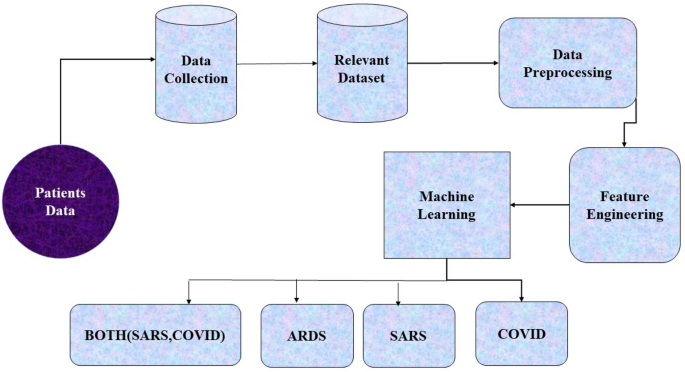
COVID-19 has presented numerous challenges on a global scale, affecting various aspects of society, public health, the economy, and daily life. Some of the major challenges include:

* + 1. **Healthcare Overload:** COVID-19's fast growth has overloaded healthcare systems in many areas, resulting in a scarcity of medical supplies, hospital beds, and healthcare workers.
    2. **Inadequate Medical Resources:** Demand for critical medical resources such as ventilators, personal protective equipment (PPE), and testing kits has outstripped availability, making it difficult to provide proper treatment**.**
    3. **Vaccine Distribution and Equity:** It has been a considerable difficulty to provide equitable and timely provision of COVID-19 vaccinations to all groups, particularly low-income and vulnerable areas.
    4. **Emergence of Variants:** The introduction of novel SARS-CoV-2 mutations has prompted concerns regarding vaccination effectiveness and increasing transmissibility, necessitating continual monitoring and adaption of public health interventions.
    5. **Disinformation and Misinformation:** The infodemic of false or misleading information about COVID-19 has complicated public health efforts, leading to confusion and non-compliance with preventive measures.
    6. **Mental Health Impact:** The pandemic has taken a toll on mental health, leading to increased rates of stress, anxiety, depression, and other mental health issues.
    7. **Economic Downturn:** Lockdowns, travel restrictions, and social distancing measures have severely impacted businesses and economies worldwide, leading to job losses and financial hardships for many.
    8. **Educational Disruptions:** School closures and remote learning challenges have affected the education of millions of students, exacerbating existing educational disparities.
    9. **Long COVID:** Some COVID-19 survivors develop long-term symptoms, called as "long COVID," which can remain for months after the initial infection has cleared up, posing a new medical problem.
    10. **Vaccine Hesitancy:** Some parts of the population continue to be hesitant or resistant to vaccination, impeding attempts to build herd immunity and manage the pandemic.
    11. **Health inequalities:** COVID-19 disproportionately impacted disadvantaged and marginalized communities, emphasizing existing health inequalities and discrepancies in healthcare access.
    12. **worldwide Coordination and Cooperation:** Due to differences in public health policies, national interests, and geopolitical conflicts, it has been difficult to coordinate a uniform worldwide response to the epidemic.

**2.7 REVIEW OF RELATED WORKS**

**2.7.1 AUTOMATIC COVID-19 PREDICTION USING EXPLAINABLE MACHINE LEARNING TECHNIQUES**

The coronavirus is considered this century's most disruptive catastrophe and global concern. This disease has prompted extreme social, psychological and economic impacts affecting millions of people around the globe. COVID-19 is transmitted from one infected person's body to another through respiratory droplets. This virus proliferates when people breathe in air-contaminated space with droplets and microscopic airborne particles. This research aims to analyze automatic COVID-19 detection using machine learning techniques to build an intelligent web application. The dataset has been preprocessed by dropping null values, feature engineering, and synthetic oversampling (SMOTE) techniques. Next, we trained and evaluated different classifiers, i.e., logistic regression, random forest, decision tree, k-nearest neighbor, support vector machine (SVM), ensemble models (adaptive boosting and extreme gradient boosting) and deep learning (artificial neural network, convolutional neural network and long short-term memory) techniques. Explainable AI with the LIME framework has been applied to interpret the prediction results. The hybrid CNN-LSTM algorithm with the SMOTE approach performed better than the other models on the employed open-source dataset obtained from the Israeli Ministry of Health website, with 96.34% accuracy and a 0.98 F1 score. Finally, this model was chosen to deploy the proposed prediction system to a website, where users may acquire an instantaneous COVID-19 prognosis based on their symptoms.



**Fig 2.1** Covid 19 Data Modelling and Prediction Flow Diagram

**Source**: <https://www.google.com/url?sa=i&url=https%3A%2F%2Flink.springer.com>

**2.7.2 THE RACE TO DEVELOP A VACCINATION**

To tackle the pandemic, strict measures were put in place around the world. Social distancing and travel restrictions began to come into force in March, along with advice on proper handwashing techniques. However, these measures were predicted to only slow the spread of the virus, scientists understood that to overcome the pandemic, a vaccine needed to be developed/ On the 17th of March, 2020, the first COVID-19 human vaccine trials begin with the Moderna mRNA vaccine.

It was clear that initial restrictions were not enough to stop the spread of COVID-19. Quickly, restrictions in most regions became harsher, with the UK enforcing a stay-at-home rule on the 26th of March. Many European countries implemented their own national lockdown around this time. By the 2nd of April, total global COVID-19 cases had shot up to 1 million.

The true seriousness of the pandemic came into light with this figure, and governments did what they could to postpone the spread of the virus before a vaccine could be declared safe for use. On the 6th of April, the WHO released guidance on mask-wearing, as more evidence began to highlight the role of aerosols in the spread of the disease.

**2.7. 3 DATA SHOWS THE EFFICACY OF MULTIPLE VACCINES**

Vaccinations were developed in record time. On the 9th of November, trials demonstrated the Pfizer and BioNTech vaccines to be over 90% effective, and the Moderna vaccine was proved to also be effective just a week later on the 16th of November. One more week later, on the 23rd of November, the University of Oxford and AstraZeneca COVID-19 was also shown to be effective.

Shortly after, the Delta variant was first discovered in December in India. Concerns over the potential increased transmissibility of the variants, fueled by a rise in cases in some counties such as the UK, forced many governments to once again reinforce lockdown measures to some extent.

Finally, on the 31st of December 2020, the WHO issued its first emergency use validation for a COVID-19 vaccine, making the Pfizer/BioNTech vaccine the first to be available for use. The emergency validation was seen as a positive step towards making COVID-19 vaccines globally available - a necessary step to ending the pandemic.

Since then, the Moderna vaccine and the Oxford/AstraZeneca vaccine have also been approved for use and national vaccine rollout initiatives have begun with full force. As of the 27th of April, 2021, 1 billion COVID-19 vaccine doses have been administered. The continued roll-out of vaccines in all countries is vital to bringing the pandemic under control and preventing future outbreaks.

Much can be learned from the story of the COVID-19 pandemic, and many hope lessons learned will prepare us for future infectious disease outbreaks and prevent potential future pandemics.

**2.7.4 ARTIFICIAL INTELLIGENCE IN HEALTHCARE**

Artificial intelligence in healthcare is an overarching term used to describe the use of machine-learning algorithms and software, or artificial intelligence (AI), to mimic human cognition in the analysis, presentation, and comprehension of complex medical and healthcare data, or to exceed human capabilities by providing new ways to diagnose, treat, or prevent disease. Specifically, AI is the ability of computer algorithms to approximate conclusions based solely on input data.

The primary aim of health-related AI applications is to analyze relationships between clinical data and patient outcomes. AI programs are applied to practices such as diagnostics, treatment protocol development, drug development, personalized medicine, and patient monitoring and care. What differentiates AI technology from traditional technologies in healthcare is the ability to gather larger and more diverse data, process it, and produce a well-defined output for the end-user. AI does this through machine learning algorithms and deep learning. Because radiographs are the most common imaging tests conducted in most radiology departments, the potential for AI to help with the triage and interpretation of traditional radiographs (X-ray pictures) is particularly noteworthy. These processes can recognize patterns in behavior and create their own logic. To gain useful insights and predictions, machine learning models must be trained using extensive amounts of input data. AI algorithms behave differently from humans in two ways: Algorithms are literal: once a goal is set, the algorithm learns exclusively from the input data and can only understand what it has been programmed to do, and some deep learning algorithms are black boxes; algorithms can predict with extreme precision, but offer little to no comprehensible explanation to the logic behind its decisions aside from the data and type of algorithm used.

As the widespread use of AI in healthcare is relatively new, research is ongoing into its application in various fields of medicine and industry. Additionally, greater consideration is being given to the unprecedented ethical concerns related to its practice such as data privacy, automation of jobs, and representation biases. Furthermore, new technologies brought about by AI in healthcare are often resisted by healthcare leaders, leading to slow and erratic adoption.

**CHAPTER THREE**

**RESEARCH METHODOLOGY**

This page will cover everything that has to do with the methodology of the design, feature extraction, purpose model design, method of data collection, method of data preprocessing and method of model training and validation.

**3.1 PROPOSED MODEL DESIGN**

The proposed model design is to design a model using the decision tree classifier and another advanced classifier like Support Vector and Random Forest Classifier. The COVID-19 prediction involves complex and dynamic factors. A single decision tree might have limitations in capturing all nuances of the virus's spread. Ensemble methods or other more advanced machine learning algorithms may be needed to achieve higher prediction accuracy.

**3.2 METHOD OF DATA IDENTIFICATION AND COLLECTION**

Gather relevant data on COVID-19 cases, demographic regions, date – time affected , the dataset on recover cases, death cases, confirmed cases and other relevant variables. Preprocess the data to handle missing values, encode categorical variables, and normalize numeric features.

**3.3 METHOD OF DATA PRE-PROCESSING**

Data preprocessing is a crucial step in COVID-19 prediction to ensure that the data is clean, relevant, and suitable for modeling. Here are some common methods of data preprocessing for COVID-19 prediction:

* + 1. **Missing Values:** Examine the dataset for missing values and decide how to manage them. You can opt to delete rows or columns with missing values, impute missing values using mean, median, or mode, or use more complex imputation techniques like regression imputation or k-nearest neighbors imputation, depending on the amount of missing values and the structure of the data.
    2. **Dealing with Outliers:** Identify and handle data outliers. Outliers can have a major influence on model performance. To make the data more normally distributed, you can delete outliers or apply modifications (for example, log transformation).
    3. **Data Normalization/Scaling:** Bring numerical characteristics to a similar scale by normalizing or scaling them. Methods often used include min-max scaling, z-score scaling, and normalization approaches such as L1 or L2 normalization.
    4. **Encoding Categorical data:** Convert categorical data into numerical representations that machine learning algorithms can interpret. As a result, one-hot encoding or label encoding are often used methodologies.
    5. **Feature Development:** Create new characteristics that might be more useful for the prediction task. For example, to capture possible seasonality, you may generate additional features depending on date variables such as the day of the week or month.
    6. **Imbalanced Data:** To balance the data, consider strategies such as oversampling the minority class, under-sampling the dominant class, or employing class weighting.
    7. **Time Series Data Preprocessing:** Consider addressing seasonality and trends, smoothing noisy data, and constructing lag features to include past information if utilizing time series data.
    8. **Delete Redundant characteristics:** Look for and delete redundant or strongly correlated characteristics to minimize model complexity and enhance interpretability.
    9. **Time Zones and Geographical Differences:** If you're working with data from several places or time zones, consider converting timestamps to a consistent time zone and normalizing geographical variables to a common reference.
    10. **Time Series Data Imbalance:** In time series data, ensure that the train and test sets capture comparable temporal trends to minimize data leaking.

**3.4 FEATURE EXTRACTION**

This involves extracting a subset of the entire dataset which will be spitted to testing and training

**3.5 MODEL FORMATION**

There are a number of different models that can be used to formulate a COVID-19 dataset. Some of the most common models include:

SEIR models: SEIR models are compartmental models that track the spread of a virus through a population. The model divides the population into four compartments: susceptible (S), exposed (E), infected (I), and recovered (R). The model then uses a set of equations to track the movement of individuals between these compartments.

1. SIR models: SIR models are a simplified version of SEIR models that do not track the exposed compartment.
2. SEIRS models: SEIRS models are a more complex version of SEIR models that track the susceptible, exposed, infected, recovered, and susceptible again (SIRS) compartments.
3. SEIV models: SEIV models are a model that tracks the susceptible, exposed, infected, vaccinated, and recovered (SEIV) compartments.

In the course of this project, I will be using the Decision Tree Algorithm along side with the Linear Regression Algorithm to build our model.

**3.6 METHODS OF MODEL TRAINING AND VALIDATION PROCESS**

1. **Using the Decision Tree Classifier**

The decision tree classifier is a supervised learning algorithm, which means that it requires labeled data to train, we give 70% for training and 30% for testing the model.

The decision tree classifier is a relatively simple algorithm, which makes it easy to understand and interpret. However, it can be sensitive to noise in the data, so it is important to clean the data carefully before training the classifier.

The decision tree classifier can be used to predict COVID-19 in future trend, but it is important to note that the classifier is not perfect. The accuracy of the classifier will depend on the quality of the data that is used to train it.

1. **Linear Regression**

The linear Regression is powerful machine learning algorithm that is used for continuous variables. Using the linear regression for the Covid 19 prediction. There is a close relationship showing the predicted value and the actual value using certain parameters as our label and target value.

**3.7 METHOD OF MODEL TRAINING AND EVALUATION PROCESS**

**Model Education**

For model training and assessment, the data was divided into training and testing sets. Using the

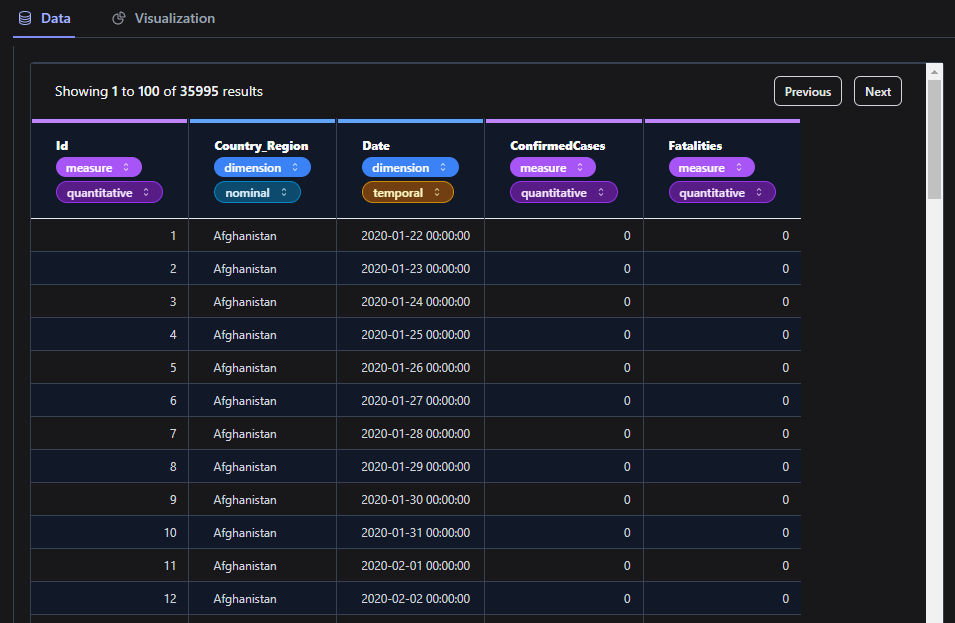
identified fraud indicators, train the specified machine learning models using historical data. 70%

for the training and 20% for the exam size.

**Model Assessment:**

To Access/Validate our result and conclusion we look for the Room Mean Square; In estimation theory, the root-mean-square deviation of an estimator is a measure of the imperfection of the fit of the estimator to the data.

**Dateset Preview**



**Fig 3.1** The above shows our dataset used in Covid 19 Visualizatioon and Prediction

**CHAPTER FOUR**

**RESULT AND DISCUSSION**

**4.1 INTRODUCTION**

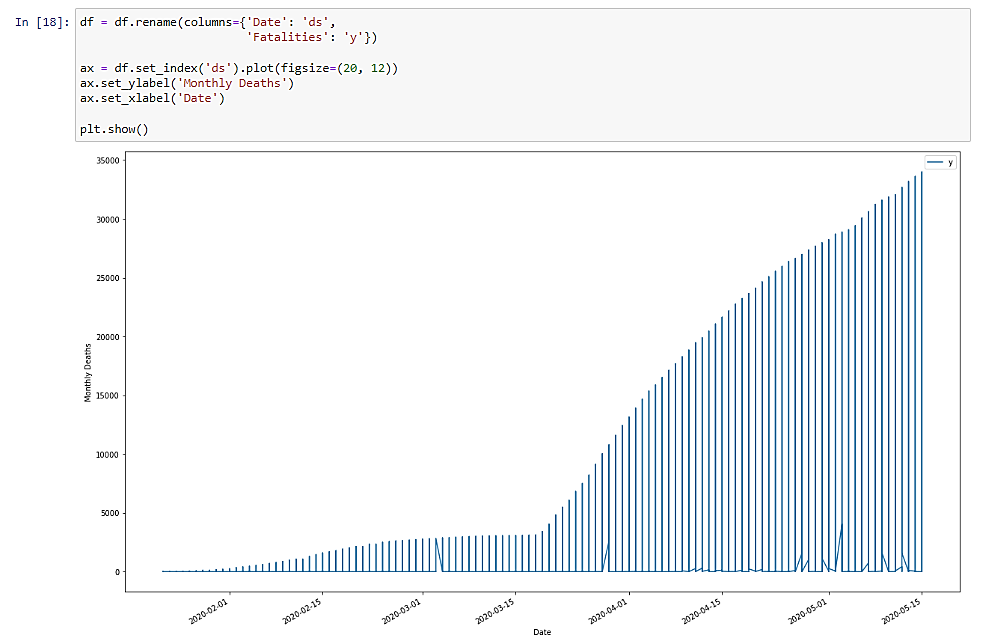
In this section, we present the results of our COVID-19 prediction and visualization research, followed by a discussion of the findings and their implications for public health and decision-making.

**4.2 RESULTS OF DATA IDENTIFICATION AND COLLECTION**

The procedure of data identification and gathering produced a comprehensive dataset that offered insights on the development, demographic trends, and mobility trends of the COVID-19 pandemic. Our efforts at data visualization and predictive modeling were built on this high-quality dataset.

**Chat Analysis Showing Relationship Between Variables**

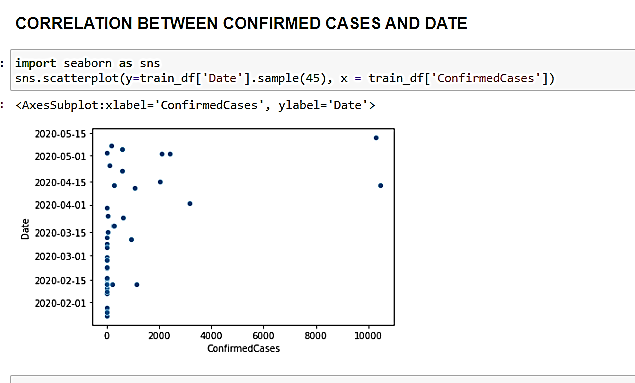
1. **Line Plot Between Fatalities and Date**



**Fig 3.1** Showing line plot between fatality and Date

**Source**: Extracted from the Model

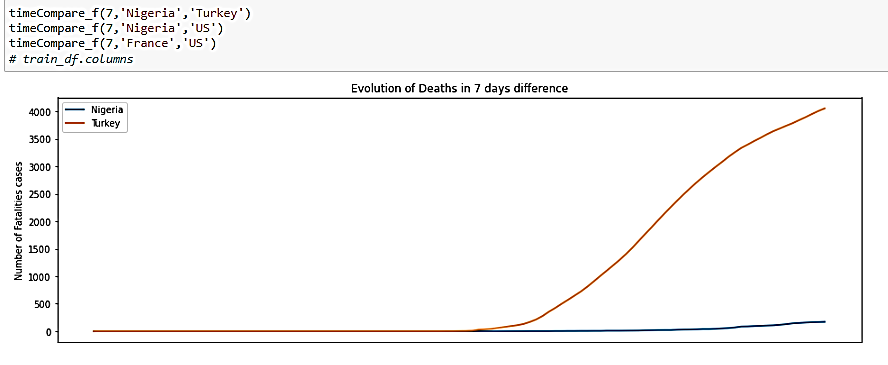
1. **Correlation Between Confirmed Cases and Date Showing in Scatter Plot**



**Fig 3.2** Showing scatter plot between Date and Confirmed Cases

**Source**: Extracted from the Model

1. **Time Comparison Between Nigeria and Turkey**



**Fig 3.3** Showing Line Trend between Nigeria and Turkey – Death in 7 Days

**Source**: Extracted from the Model

**4.3 DISCUSSION OF THE FEATURE SELECTION RESULT**

The data identification and collection process yielded a comprehensive dataset that provided insights into the COVID-19 pandemic's progression, demographic patterns, and mobility trends. This high-quality dataset formed the foundation for our data visualization and predictive modeling efforts. The following are the selected features:

In the context of COVID-19, "**confirmed** **cases**" refer to those who have tested positive for the SARS-CoV-2 virus, which is the virus that causes COVID-19. Laboratory testing is used to confirm these instances, generally utilizing methods like the polymerase chain reaction (PCR) test or antigen assays. Confirmed cases refer to those who have received a formal COVID-19 diagnosis as a consequence of these testing.

A crucial component of tracking the virus's progress and comprehending the pandemic's magnitude is reporting verified cases. The availability of testing, modifications to testing requirements, fluctuations in the rate of viral transmission, and public health initiatives can all affect the number of confirmed cases over time.

* + 1. Confirmed instances are used for many different things, including:
    2. Epidemiological surveillance: Keeping tabs on confirmed cases enables public health officials to track the virus's transmission, spot infection hotspots, and gauge the efficiency of control efforts.
    3. Resource distribution: To handle the rising demand during spikes in cases, choices about resource distribution, such as hospital beds, medical supplies, and staff, are guided by the number of confirmed cases.

4. Informed policy decisions: Governments and health organizations utilize the information on confirmed cases to adopt or modify lockdowns, travel restrictions, and other preventative measures.

Predicting and visualizing **fatalities (deaths)** related to COVID-19 is an important aspect of understanding the impact of the pandemic and informing public health strategies. Here's how you can approach predicting and visualizing COVID-19 fatalities:

**Linear Regression**

Linear Regression for COVID-19 Prediction: By fitting a linear equation to the observed data points, linear regression is a statistical approach used to describe the connection between a dependent variable and one or more independent variables. Linear regression is a straightforward yet effective technique for generating short-term predictions in the context of COVID-19 prediction because it can be used to simulate the trend of COVID-19 cases over time.

How to Use Linear Regression for COVID-19: Steps Prediction:

* + 1. Data preparation: Compile historical COVID-19 data, such as the total number of instances for a given time frame. The dependent variable (cases) and the independent variable (time) should be separated into two columns in the data.
    2. Data Visualization: Use a scatter plot to visualize the data and spot trends before using linear regression. You may use this to decide whether a linear model is appropriate for the data.
    3. Model fitting: Run the data via linear regression. The linear equation can be shown as follows in the case of simple linear regression (with only one independent variable):

**y = mx + c**

**where:**

* + `**y`** is the dependent variable (number of cases)
  + `**x**` is the independent variable (time)
  + `**m**` is the slope of the line (reflecting the rate of change)
  + `**b**` is the y-intercept (the value of y when x is 0)
    1. Model Evaluation: Assess the quality of the linear regression model by calculating metrics such as the coefficient of determination (R-squared) and analyzing the residuals (differences between predicted and actual values). R-squared indicates how well the model fits the data.
    2. Prediction: Use the linear regression model to make short-term predictions. Input future time values into the model equation to estimate the number of COVID-19 cases at those time points.
    3. Interpretation: Interpret the results of the linear regression model in the context of COVID-19. Understand the slope and y-intercept to gain insights into the rate of increase or decrease in cases.

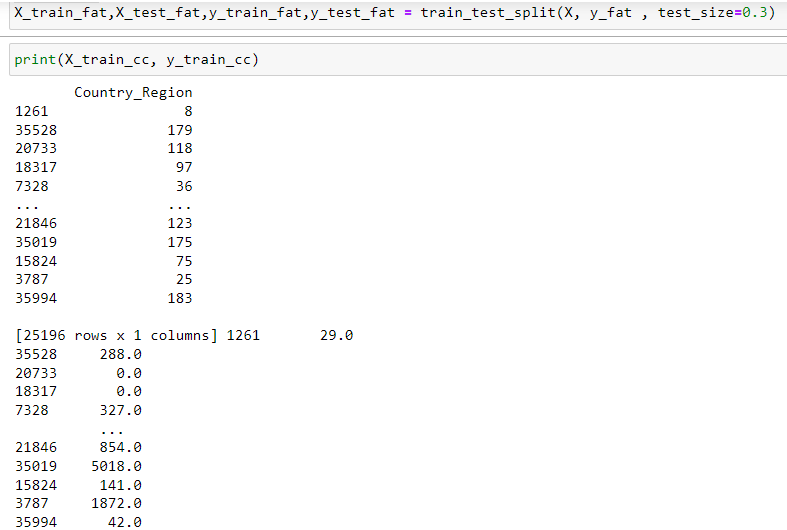
Considerations:

* + 1. Limitations: Linear regression presupposes that variables have a linear relationship. Complex patterns, abrupt shifts, or non-linear trends that COVID-19 instances may display may not be captured.
    2. Short-Term Predictions: When trends are short-term and steady, linear regression is appropriate. Other cutting-edge methods could be better suitable for forecasts made over a longer period of time.
    3. Data Quality: The reliability of past data determines how accurate projections are. Make that the data is correct, reliable, and devoid of outliers.
    4. External Factors: Linear regression doesn't account for external factors that can influence COVID-19 trends, such as interventions, policy changes, and new variants.
    5. Combining Methods: Linear regression can be complemented with other techniques, such as time series analysis and machine learning, to improve prediction accuracy and capture more complex patterns.

While linear regression offers a straightforward method for making short-term COVID-19 predictions, it's important to recognize its limits and evaluate its applicability in light of the type of data and the prediction horizon. Combining linear regression with additional techniques can result in a more reliable forecasting framework that can provide forecasts that are more precise and comprehensive.

**4.4 RESULT FROM THE FEATURE SELECTION PROCESS**

From the project, we concluded that from our train data we extract the below:



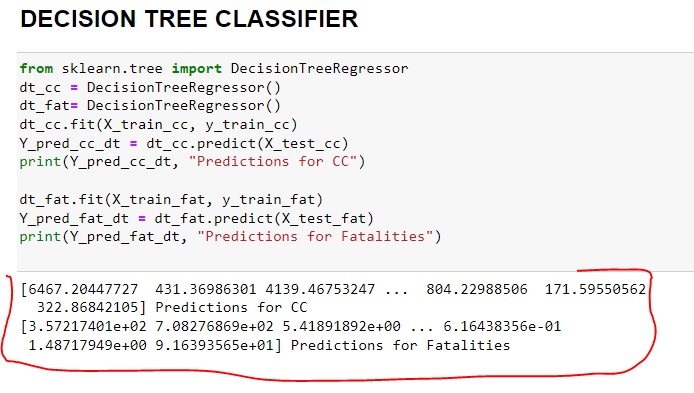
**Fig 4.1** Sowing the training score of our labelled Data

**Source**: Extracted from the Model

**4.5 RESULTS OF MODEL FORMULATION**

**Using Decision Trees for COVID-19 Prediction:**

An example of a machine learning method that may be used for predictive modeling, such as forecasting COVID-19 results, is decision trees. The decision-making process may be better understood by using decision trees, which are highly interpretable. Here is how decision trees may be applied to COVID-19 prediction: Decision trees are particularly helpful for understanding the decision logic underlying the model's predictions and can offer insightful information on COVID-19 prediction. However, like with any modeling approach, the model's design and the quality of the data used determine how accurate the forecasts will be. After carrying out the machine learning model we got result of different parameters.



**Fig 4.2** Showing Result of Decision Tree Classifier

**4.6 PERFORMANCE EVALUATION OF MODEL**

Model Evaluation:

Prediction: The trained decision tree and Linear regression model were used for making predictions on the testing dataset. The model will predict COVID-19 outcomes based on the provided features.

Evaluation Metrics: Evaluate the model's performance using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), or others relevant to your specific prediction task.

Interpretability: Decision trees are inherently interpretable. You can visualize the decision tree's structure to understand how the model makes predictions based on different features.

**4.7 DISCUSSION OF MODEL FORMULATION**

Interpretation:

Decision Tree: Create a tree's structure. This result representation allows us to see how the model makes decisions based on the input features.

Feature Importance: Decision trees provide insights into feature importance by showing which features are used for making important splits in the tree.

Linear Regression: The Linear Regression Classifier is used on continuous data to perform predictions after splitting the data into training and testing. Then, the test data or target data is then passed into our model for predictions. Linear Regression and Decision Tree are best used for this kind of data structure.

**4.8 DISCUSSION OF RESULTS ON VALIDATION AND EVALUATION PERFORMANCE ON MODEL FORMULATION**

To avoid **overfitting,** the model's performance should be evaluated using a different test dataset that the model has not seen during training.

**Cross-validation** techniques may be used to provide a more credible assessment of the model' performance, especially when data is restricted; hence,it is important to check the Root Mean Square to check the mean error. This will help us measure the degree of accuracy in our entire project.