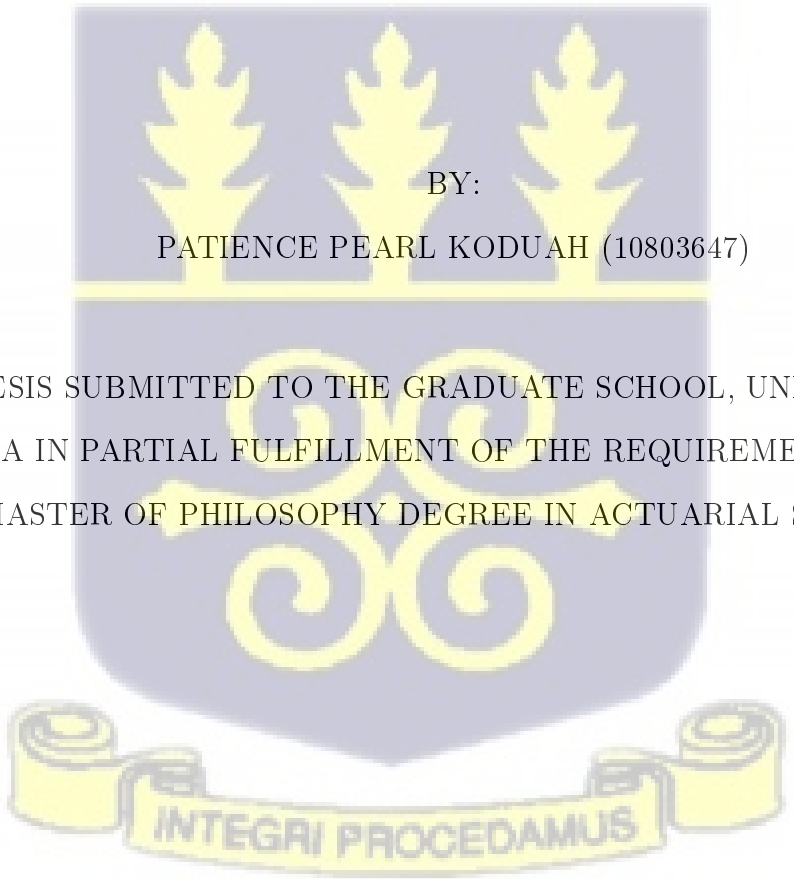
UNIVERSITY OF GHANA, LEGON

MODELLING COVID-19 TRANSMISSION IN GHANA USING A DISCRETE-TIME MARKOV MODEL AND MACHINE LEARNING TIME-SERIES FORECASTING ALGORITHMS

BY:

PATIENCE PEARL KODUAH (10803647)

A THESIS SUBMITTED TO THE GRADUATE SCHOOL, UNIVERSITY OF GHANA IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE MASTER OF PHILOSOPHY DEGREE IN ACTUARIAL SCIENCE



September, 2022

DECLARATION

Candidate's declaration:

I hereby declare that this submission is my own work towards the award of the Master of Philosophy degree and that, to the best of my knowledge, it contains no material previously published by another person nor material which had been accepted for the award of any other degree of the university, except where due acknowledgement had been made in the text.

SIGNATURE: ..... DATE: **17-10-2022**

PATIENCE PEARL KODUAH

(10803647)

Supervisors' declaration:

We hereby certify that this thesis was prepared from the candidates own work and supervised in accordance with guidelines on supervision of thesis laid down by the university of Ghana.

SIGNATURE: ..... DATE: **17-10-2022**

DR. LOUIS ASEIDU

(Principal Supervisor)

SIGNATURE: ..... DATE: 17-10-2022

DR. PERPETUAL ANDAM BOIQUAYE

(Co-Supervisor)

ABSTRACT

The COVID-19 pandemic has and continue to have a severe impact on the health sectors, businesses, economies, and the world at large, despite many healthcare interventions, with much still yet to be learnt regarding its infection dynamics. In addition, researchers have developed classical compartmental or epidemiological models and other advanced mathematical models to better explain COVID-19 infection dynamics across many countries. Critical information, such as the likelihood of first infection and recovery, average infection duration before this infection dies out entirely, COVID-19 infected people's life expectancy, and generalised transition probabilities, is understudied at any given future time. Using nationwide aggregated COVID-19 datasets and a discrete-time Markov model (to estimate these key disease metrics), the current study adds to our understanding of COVID-19 infection dynamics in Ghana. Additionally, the predictive power of some existing state-of-the-art machine learning (ML) algorithms such as K-Nearest Neighbor regression (KNN), Neural Network Auto-Regressive (NNAR), Generalized Regression Neural Network (GRNN), Multi-Layer Perceptron (MLP), and Extreme Learning Machines (ELM) in forecasting daily cases of COVID-19 infection (over the study period) is investigated using an out-of-sample rolling-origin evaluation by exploring the trade-off between computational speed and accuracy. It was estimated that there would be a prolonged COVID-19 transmission for at least 150 years before infection could die out. The study supports the idea that with a high overall recovery rate, a low infection rate, and a longer infection period, there is a possibility of herd immunity (as evident in the 2021 infection period despite the relatively high overall rate of infection). Finally, the K-Nearest Neighbour (KNN) regression was found to be the most cost-effective ML algorithm to predict the daily cases of COVID-19 in Ghana via the rolling-origin evaluation strategy.

DEDICATION

To my parents, Dr. and Mrs. Imbeah, Ms. Elizabeth Yankah, and Mr and Mrs Koduah

ACKNOWLEDGMENT

Many people have contributed in diverse ways to helping me succeed in my MPhil studies and complete this thesis. I want to express my sincere gratitude and appreciation to my supervisor, Dr Louis Asiedu, whose help, stimulating suggestions, and words of encouragement made me reach this pedestal in my academic work. I also want to sincerely thank my co-supervisor, Dr Perpetual Andam, for spending time proofreading and correcting my many mistakes. With much appreciation, I would also like to acknowledge the crucial role of Clement Twumasi of Oxford and Cardiff University (UK), who served as a secondary supervisor to provide valuable support. His helpful guidance and immense knowledge helped me in my research and thesis writing. I could not have imagined having a better advisor and mentor for my MPhil thesis. My sincere thanks go to my sitting mate, Fred Mawuli, my parents, and my husband for the sleepless nights we worked together to achieve this goal. Last but not least, I would like to thank my family: Dr and Mrs William Imbeah, Ms Elizabeth Yankah (Registrar at the University of Cape Coast) and Mr and Mrs Koduah for supporting me spiritually and financially. I am grateful to all the loved ones who contributed their quota in diverse ways to my success.

Contents

List of Tables

List of Figures

List of Abbreviation

ARIMA Autoregressive Integrated Moving Average

KNN K-Nearest Neighbor regression

NNAR Neural Network Auto-Regressive

GRNN Generalized Regression Neural Network

MLP Multi-Layer Perceptron

ELM Extreme Learning Machines

NN Neural Network

ML Machine Learning

MLE Maximum Likelihood Estimation

RBF Radial Basis Functions

CART Classification and Regression Trees

SVR Support Vector Regression

RNN Recurrent Neural Network

BNN Bayesian Neural Network

LSTM Long Short Term Memory

AANN Automated Artificial Neural Network

GP Gaussian Processes

S-I-D Susceptible-Infected-Dead

RMSE Root Mean Squared Error

MdRMSE Median Root Mean Squared Error

FELTP Field Epidemiology and Laboratory Training Program

PPE Personal Protective Equipment

AIC Akaike Information Criterion

WHO World Health Organisation

Chapter 1

Introduction

This chapter presents the background of the study by giving some salient information on the underlying disease of the study. The researcher also discusses the problem of the study, the objectives of the study, the significance of the study, and how the study is organised.

1.1 Background of the Study

The coronavirus is a zoonotic virus, an RNA virus in the order Nidovirales in the family Coronaviridae (WHO, 2020b). It is a family of viruses that cause respiratory infections, first isolated in 1937 and named coronaviruses in 1965 because they have a microscopic crown-like appearance (WHO, 2020b). Seven coronaviruses can produce infections in people around the world. The forms of coronaviruses known to date are i) HCoV-229E and HCoV-NL63 alpha coronaviruses, ii) HCoV-OC43 and HCoV-HKU1 beta coronaviruses, iii) SARS-CoV (which causes extreme acute respiratory syndrome), iv) MERS-CoV (which causes respiratory syndrome in the Middle East), and v) SARS-CoV-2 (a new variant) identified in China at the end of 2019 after reported cases in China (WHO, 2020b); Although the new variant is infectious, most individuals are infected with other human variants: 229E, NL63, OC43, and HKU1. They often lead to respiratory infections ranging from the common cold to more severe diseases such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS). Nonetheless, the current pandemic is caused by a coro-

navirus dubbed COVID-19. Initially, WHO called this infectious disease Novel Coronavirus-Infected Pneumonia (NCIP) and the virus was labelled 20 (2019-nCoV) but was formally renamed the clinical disorder COVID-19 on February 11, 2020 (a shortening of Corona Virus Disease-19). In Wuhan, Hubei Province, China, an outbreak of COVID-19 triggered by the 2019 novel coronavirus (SARS-CoV-2) started in December 2019, with the latest outbreak officially considered a pandemic (WHO, 2020b).

Droplets of various sizes can spread respiratory infections: they are referred to as respiratory droplets when the droplet particles are >5-10μm in diameter, and they are referred to as droplet nuclei when they are > 5μm in diameter (WHO, 2014). The COVID-19 virus is primarily transmitted between individuals via respiratory droplets and communication routes, according to current evidence (Liu et al., 2020; C. Huang et al., 2020; Team, 2020; WHO, 2020b). Droplet transmission occurs when a person is in close contact (within 1 meter) with someone who has respiratory symptoms (such as coughing or sneezing) and is thus at risk of exposure to potentially infectious respiratory droplets from an infected person's mucosa (mouth and nose) or conjunctiva (eyes). Transmission can also occur via fomites in the immediate environment around the infected individual (Ong et al., 2020). Consequently, COVID-19 virus transmission can occur through direct contact with infected people and indirect contact with surfaces in the immediate environment or with items used by the infected person (e.g., stethoscope or thermometer).

The World Health Organization announced the novel COVID-19 a pandemic on March 11, 2020 (WHO, 2020a). On March 12, 2020, the first two cases in Ghana were confirmed when two infected citizens, one from Norway and the other from Turkey, arrived in Ghana (Ducan, 2014). In consultation with other stakeholders, the government rapidly introduced the measure to curb its dissemination. The government imposed bans on schools, events, and social gatherings. Addition-

ally, a temporary lockdown and restrictions on the movement of people in the Greater Accra and Ashanti regions of Ghana were steps imposed by the government beginning on March 15, 2020 (GHS, 2020). These two cases introduced the first method of contact tracing in Ghana. Of the first two cases registered in Ghana, one was a senior officer returning from Norway at the Norwegian Embassy in Ghana. At the same time, the other was a staff member returning from Turkey at the United Nations (UN) offices in Ghana (Ducan, 2014; Anyoriga, 2020; GhraphicOnline, 2020). According to COVID-19 data from Worldometer (2020), the current number of confirmed cases in Ghana as of December 31 was 54,771, with 335 deaths and 53,594 recoveries.

1.2 Problem Statement

In late December 2019, a wave of unexpected pneumonia-type cases with an unknown cause was found in Wuhan, China. A new zoonotic coronavirus (COVID-19) was thus identified as the pathogenic agent of this unexplained pneumonia a few days later, which caused a public health crisis and was eventually declared by WHO as a global pandemic in early 2020 due to the exponential growth of the number of infected individuals per country. Despite several healthcare interventions (e.g., production of vaccines, lock-down measures, and other forms of preventive measures), the COVID-19 pandemic has had a devastating impact on the health sector, businesses, economies, and the world at large, with a lot still yet to be uncovered. Several accelerated attempts have been made by governments, clinicians, and health experts through clinical trials to develop vaccines such as Pfizer, AstraZeneca, and Moderna, etc., to protect humans from developing COVID-19 pathogenesis (Haque & Pant, 2020). Also, disease modellers have developed classical compartmental or epidemiological models and other sophisticated mathematical models to understand the COVID-19 infection dynamics across different countries. Consequently, these studies have estimated epidemic

thresholds for the outbreak, including but not limited to the basic reproductive number, Malthusian parameter, as well as the herd immunity threshold (which quantifies the proportion of the population needed to be vaccinated to control the infection), amongst others (see eg. Demongeot et al., 2021; Khoshnaw et al., 2020; Khrapov & Loginova, 2020; Sookaromdee et al., 2020).

However, other important information concerning the probability of first infection and recovery, average infection duration before COVID-19 infection dies out from a given population, the life expectancy of COVID-19 infected individuals, and generalised transition probabilities at any given future time are under-studied. Estimates of these parameters can also help inform management decisions and control infectious diseases. Zipkin et al. (2010) and Twumasi et al. (2019) have successfully shown that the first-order discrete-time Markov chain model can be adopted to provide useful disease summaries and estimates of these relevant epidemiological quantities for a given infectious disease within a population. Thus, aggregated data of the numbers of susceptible, infected, recovered, and dead individuals over an observation period are sufficient to estimate these disease metrics, with the help of the discrete-time Markov model (without any prior knowledge about the per-capita infection, recovery, and mortality rates as in the case of fitting traditional compartmental models). Moreover, due to the exponential growth of COVID-19 infected cases (regardless of the presence of existing health-care interventions) across the world (especially among developing nations), it is also imperative to develop other robust predictive models to forecast these cases of infections over time.

The current study provides additional insights into the infection dynamics of COVID-19 in Ghana by adopting a discrete-time Markov model (to estimate other relevant disease metrics) with the help of national aggregated COVID-19 datasets (obtained in 2020 and 2021). Additionally, the predictive power of some existing state-of-the-art machine learning algorithms: K-Nearest Neighbor regres-

sion (KNN), Neural Network Auto-Regressive (NNAR), Generalized Regression Neural Network (GRNN), Multi-Layer Perceptron (MLP), and Extreme Learning Machines (ELM) in forecasting daily cases of COVID-19 infection (over the same observed periods) via an out-of-sample rolling-origin evaluation strategy is investigated (to learn more from the empirical data).

1.3 Research Objectives

The primary goal of the study is to forecast daily COVID-19 cases in Ghana using the first-order Markov chain model and machine learning algorithms. The specific objectives of the study are:

1. To develop a Susceptible-Infected-Dead (S-I-D) discrete-time Markov model and estimate other relevant disease metrics for COVID-19 transmission based on aggregated data for 2020 and 2021.
2. To derive and compare the exact mathematical expression for the n-th step transition matrix (P^n) at each infection period (2020 and 2021) to make predictions using the derived P^n matrix, and compare the model performance at the two infection periods.
3. To compare the predictive performance of the fitted machine learning algorithms.

1.4 Research questions

1. What is the overall probability of COVID-19 infection and recovery in Ghana?

2. What is the expected time to infection or expected duration of COVID-19 in Ghana?

3. What is the average life expectancy for susceptible and infected individuals during the COVID-19 outbreak in 2020 and 2021?

4. Is the n-th step transition probability matrix during the COVID-19 pandemic dependent on the period of infection?

5. Which machine learning algorithm(s) has the best forecasting power via the rolling-origin evaluation strategy (based on model accuracy-speed tradeoff)?

6. Is there a significant difference between the error distributions of the fitted ARIMA time-series model and the machine learning time series models (GRNN, NNAR, MLP, and ELM) in predicting the daily COVID-19 infection cases in Ghana?

1.5 Significance of the Study

The study will be imperative in several ways;

1. The study explains how aggregated counts per disease outcome (without prior knowledge of per-capita rates of the disease outcomes) can be computed. On the back of this, the current study demonstrates the application of discrete-time Markov modelling in estimating other important disease metrics (such as the probability of first transition per each disease outcome, expected duration of infection, and life expectancies due to COVID-19 transmission). The knowledge of this will help in controlling and managing the spread of the pandemic.

2. Academically, the research demonstrates the application of ML algorithms and a classical time-series model for effectively forecasting daily cases of

COVID-19 infection. It elaborates on how this forecasting can be done via a more robust evaluation strategy (out-of-sample rolling-origin) compared to the fixed-origin approach.

1.6 Organization of the Study

The study entails five major chapters. Chapter One introduces the background of the study, problem statements, research objectives, research questions, the significance of the study, and the organization of the study. Chapter Two reviews relevant literature and discusses other underlying theories. This chapter will also add to knowledge by reviewing scholarly information on the subject matter and helping broaden the researcher's knowledge. Chapter Three highlights the methodologies employed in the current study (i.e., the discrete-time S-I-D Markov model and the time-series models). It also includes a description of the empirical datasets. Chapter Four presents the results and discusses the key research findings. Finally, Chapter Five covers the study's summary, conclusion, and recommendations.

Chapter 2

Literature Review

This chapter discusses the chronology of COVID-19 infections and how the pandemic has evolved since its outbreak. The chapter also presents an empirical review of literature relating to modelling of the pandemic across countries. Any theoretical view on the modelling of COVID-19 is also explored in this chapter.

2.1 The chronology of COVID-19 infections

In December of the year 2019, the primary cases were accounted for (Du Toit, 2020). Five individuals were admitted to the medical clinic with intense respiratory pain conditions from December 18 to December 29, 2019, and one of them kicked the bucket (Ren-LL et al., 2020). By January 2, 2020, the research center confirmed COVID-19 contamination in 41 admitted emergency clinic patients; just under half of these patients had ongoing illnesses like diabetes, hypertension, or cardiovascular infection (C. Huang et al., 2020). In all probability, patients were thought to have gotten tainted while in that medical clinic, in all probability because of nosocomial sickness. The COVID-19 infection was resolved not to be a super-hot spreading infection (one patient tainting numerous others) but rather spread because of a few patients getting contaminated at various locales throughout the medical clinic through unexplained strategies. Moreover, just patients who turned out to be clinically sick were tried, inferring that a lot more individuals were probably contaminated. An aggregate of 571 instances of the 2019-new COVID-19 had been recorded across China's 25 regions (areas and ur-

ban communities) as of January 22, 2020, (H. Lu, 2020). The China National Health Commission has released the current realities of the initial 17 passings, which happened between January 22 and January 22, 2020. On January 25, 2020, an aggregate of 1975 cases in terrain China were confirmed to be contaminated with COVID-19, with 56 passings (Wang et al., 2020). On January 24, 2020, another source surveyed the aggregate rate in China and found 5502 cases (Nishiura et al., 2020). 7734 occurrences have been confirmed in China as of January 30, 2020, with 90 additional cases revealed from Taiwan, Thailand, Vietnam, Malaysia, Nepal, Sri Lanka, Cambodia, Japan, Singapore, Republic of Korea, and the United Arab Emirates, United States, Philippines, India, Australia, Canada, Finland, France, and Germany. The case casualty rate (170/7824) was determined to be 2.2 percent (Bassetti et al., 2020). The depiction, ID, finding, clinical history, and the executives of this case depended on the primary instance of COVID-19 disease confirmed in the United States. This covers the patient's underlying gentle manifestations at a show, as well as the movement of the sickness to pneumonia on day 9 (Team, 2020). The Centers for Disease Control and Prevention (CDC) has screened more than 30,000 individuals showing up at US air terminals for the new Covid. Following this underlying screening, 443 individuals in 41 states in the United States were tried for Covid disease. Just 15 (3.1%) were discovered to be positive, while 347 were discovered to be negative and the discoveries for the excess 81 are as yet forthcoming.

The World Health Organization (WHO) broadcast the COVID-19 spread as a Public Health Emergency of International Concern (PHEIC) on January 30, 2020(WHO, 2020b). The WHO Director-General declared the COVID-19 episode a pandemic on March 11, 2020, because of an undeniable degree of overall spread and lethality (WHO, 2020b).

2.2 Symptoms

After a brooding time of around 5.2 days, COVID-19 disease side effects happen(Li et al., 2020). The time it took for COVID-19 indications to show into death went from 6 to 41 days, with a middle of 14 days (Wang, Tang, and Wei, 2020). The period relies upon the patient's age and the condition of their insusceptible framework. At the point when patients past the age of 70 were contrasted with those younger than 70, it was more limited (Wang et al., 2020). Fever, hack, and depletion are the most common manifestations of COVID-19 contamination, albeit different indications incorporate sputum creation, migraine, hemoptysis, looseness of the bowels, dyspnea, and lymphopenia (Ren-LL et al., 2020; C. Huang et al., 2020; Wang et al., 2020). A chest CT filter affirmed pneumonia, however, there were likewise variant qualities like RNAaemia, intense respiratory trouble disorder, intense heart harm, and an expanded pervasiveness of terrific glass opacities, all of which added to death (C. Huang et al., 2020). Numerous fringe ground-glass opacities were seen in subpleural areas of the two lungs in specific cases (Lei, Li, Li, and Qi, 2020), which probably created a fundamental and confined insusceptible reaction, bringing about raised aggravation. Sadly, treatment with interferon inward breath at times had a minimal clinical effect and seemed to bother the ailment by making lung opacities advance (Lei, Li, Li, and Qi, 2020).

Significantly, the manifestations of COVID-19 and past beta coronaviruses are comparable, including fever, dry hack, dyspnea, and two-sided ground-glass opacities on chest CT filters (C. Huang et al., 2020). Coronavirus, then again, showed certain particular clinical attributes, like the focusing of the lower aviation route, as confirmed by upper respiratory lot side effects, for example, rhinorrhoea, sniffling, with sore throat (Assiri et al., 2013). Besides, a portion of the cases shows an invasion in the upper projection of the lung, which is connected with expanded

dyspnea with hypoxemia, in light of the aftereffects of chest radiographs taken upon affirmation (Lee et al., 2003).

Significantly, while COVID-19 patients showed gastrointestinal indications like loose bowels, just a little level of MERS-CoV or SARS-CoV patients displayed comparative GI misery. Thus, it's basic to analyze defecation and pee tests to preclude a potential backup way to go of transmission, for example, through medical care faculty, patients, etc (Assiri et al., 2013). Thus, creating devices to recognize the different components of the transmission, for example, fecal and pee tests, is basic to foster techniques to forestall or potentially decline transmission, just as medicines to control the illness (Rothan & Byrareddy, 2020).

2.3 Transmission

The huge number of contaminated patients presented to the wet creature market in Wuhan City, where live creatures are regularly sold, suggests that this is the COVID-19's reasonable zoonotic beginning (Rothan & Byrareddy, 2020). The quest for a repository host or middle transporters from whom the sickness might have been communicated to people has been launched. Starting examination recognized two snake species as possible sources for the COVID-19 infection (Rothan & Byrareddy, 2020). All things considered, other than warm-blooded animals and birds, there has been no reliable proof of Covid supplies to date Bassetti et al. (2020). Coronavirus shares 88% of its genomic succession with two bat-determined serious intense respiratory disorder (SARS)- like Covids (R. Lu et al., 2020), demonstrating that vertebrates are the most probable wellspring of COVID-19. A few investigations have uncovered that COVID-19 contamination spreads, for the most part, through individual-to-individual contact. Cases among families and among people who didn't visit Wuhan's wet creature market back up this affirmation (Carlos et al., 2020). Individual-to-individual transmission

occurs primarily through direct touch or drops sent by a contaminated individual while hacking or wheezing (Rothan & Byrareddy, 2020). A small study of women in their third trimester who were found to be COVID-infected notwithstanding, because the entirety of the pregnant ladies had cesarean methods, it's unclear if the transmission can occur during vaginal birth. This is huge because pregnant ladies are more defenseless to respiratory microbe contamination and extreme pneumonia H. Chen et al. (2020). The principal period of viral disease is the limiting of a receptor communicated by cells, trailed by its combination with the cell layer. The infection's essential objective is believed to be the lung epithelial cells. The limit between the receptor-restricting space of infection spikes and the cell receptor, which has been distinguished as the angiotensin-changing over catalyst 2 (ACE2) receptor, has been found to prevent human transmission of SARS-CoV. (Wan et al., 2020). The receptor-restricting space of COVID-19 spikes is tantamount to that of SARS-CoV, which is critical. This proof emphatically suggests that the ACE2 receptor is the most probable pathway into the host cells (Wan et al., 2020).

2.4 COVID 19 measures in Ghana

At the beginning of the COVID-19 pandemic, Ghana's reaction was hailed as truly outstanding in Africa. The country's entire government approach was organized around five targets: to shorten the importation of cases; recognize and contain them; care for the wiped out; pad the effect of COVID-19 on Ghana's financial and public activity; lift homegrown creation as a method for developing independence (Ofori-Atta, 2020).The first COVID-19 cases in Ghana were identified on March 12, 2020, when two cases were confirmed in the capital city of Accra. The two cases were distinguished as individuals who had gotten back to the country from Norway and Turkey. These imported cases started the primary contact following interaction in Ghana, identifying a few many cases in a brief timeframe. Following

the disclosure that the pandemic had effectively begun to spread in Ghana, the public authority organized measures to stop the neighborhood's spread of the contamination and any further import of the infection into the country. Before the month's over there had been 152 affirmed cases, 5 passings, and 22 recuperated patients, leaving 125 dynamic cases going into April (WHO, 2020b).

A day before the reporting of the primary case in Ghana, the president declared $100 million in subsidies from the Ministry of Finance to upgrade COVID-19 readiness and reaction, to be specific, foundation, materials, and hardware, and government-funded instruction (Presidency, 2020). In ensuing addresses, the president focused on the significance of securing medical service laborers and reported the acquisition of individual defensive hardware (PPE) and hand sanitizers for the wellbeing of faculty (Presidency, 2020). The president also stated that he would work with neighborhood manufacturing organizations to support homegrown PPE creation and provide improved protection bundles and stipends for medical services workers (Presidency, 2020). Parliament passed the Novel Coronavirus (COVID-19) National Trust Fund Act, 2020 (Act 1013) which established an asset to get and oversee commitments and gifts from people, gatherings, and corporate bodies to help in the battle against the infection. Different estimates included closing all land borders to the country along with the conclusion of the principal worldwide air terminal in Accra. Mediator drives saw people entering the country between the hour of the declaration and the closing of the lines, being isolated in 4- and 5-star inns to the detriment of the Ghanaian government. Identification administrations have also been halted (GhraphicOnline, 2020). Two organizations, specifically the Kumasi Center for Collaborative Research into Tropical Medicine (KCCR) in Kumasi and the Noguchi Memorial Institute for Medical Research (NMIMR) in Accra, filled in as introductory testing communities for every presumed case. Tests from the northern part of the country were sent to the KCCR, while those from the southern part were sent to the NMIMR.PCR testing of around 1,000 examples each day, due to pooling of

tests, large numbers of which were negative. To supplement endeavors, Zipline, an organization with a mission to "furnish each human on Earth with moment admittance to indispensable clinical supplies," gathered examples from more than 1,000 wellbeing offices situated in hard-to-reach rustic spaces of Ghana and sent them to the testing habitats utilizing drones. This was huge as it assisted with working with speedy case discovery and inception of treatment for every single positive case.

Starting in May 2020, the public authorities of Ghana engaged other wellbeing offices to set up testing communities. Some private wellness offices have likewise begun to embrace tests for COVID. Although testing was first unified in the two previously mentioned organizations, hospitalization occurred at wellbeing offices inside the region where the positive cases had been found. To deal with COVID-19 cases, disengagement focuses were established in almost every area.

Furthermore, beginning in July 2020, a portion of the restrictions were gradually eased, including the opening of air terminals and land borders on September 1, 2020. Individuals had the option to move considerably more unreservedly between urban communities. All global voyagers flying into Ghana needed to be tested in their nation of origin before embarkation. On appearance, a test is likewise conducted at the air terminal in Ghana before anybody is permitted into the country. Those testing positive are made to isolate at an assigned inn for 12 days before being permitted to go, over their objections.

A fractional lockdown was put on the two greatest urban communities, Accra and Kumasi, which were the focal points for the pandemic, from the end of March 2020 for three weeks. There were limitations on the movement of individuals, with the police and military mounting barriers to complete checks. Only specialists of fundamental administrations (in areas like medical services, media, food sellers/eateries, and security offices) were permitted to move around in these urban communities. Individuals in the remainder of the nation could move around

since no cases had been accounted for at that point.

The government of Ghana additionally restricted all open get-togethers, including meetings, workshops, burial services, celebrations, political assemblies, church and mosque services, and other related occasions, to diminish the spread of the infection. Seashores, bars, and nightclubs were additionally closed. Notwithstanding, with the impact from the start of June, there were some unwinding of limitations. Strict administrations were permitted to begin on Friday, June 5, with the mandatory use of nose veils and gatherings of no more than 100 people. Private entombments with the most extreme participation of 100 people were permitted. Furthermore, weddings and other parties could take place with fewer than 100 people.

All instructive organizations from kindergarten to tertiary establishments, including colleges, both public and private, were shut. For kids in some kindergartens and elementary schools, educators will instruct in a study hall, take recordings and send them to the youngsters through WhatsApp, while their parents and gatekeepers will get their tasks from the schools every week for them to do. Applicants who were going to complete their last terms of the year at the primary and secondary levels were brought in May 2020 to finish up with their academic programs. Any remaining understudies were to proceed with their online stage performances. The president declared that schools would resume on January 18, 2021. He consoled the public that the grounds were going through sanitization and that actions were set up for contamination control (Presidency, 2020). School administration should exercise caution when dealing with class size issues that are specific to their school or region.

Public vehicle administrators were likewise offered mandates to uphold vehicle physical separating, while workers were relied upon to utilize face covers inside vehicles. Consistency to these orders differed, with a larger part (98%) of transport administrators adhering to the rules, albeit a huge number of suburbanites

were not sticking to the rules on face covers (Dzisi & Dei, 2020). The Local Government Minister likewise declared the sterilization of 137 business sectors in the Greater Accra Region beginning from Monday 23rd of March 2020 which was adequately done. The Ministry of Local Government and Rural Development collaborated with Moderpest Company and Zoomlion Ghana for the activity. Additionally, the Ghana Health Service (GHS) set out on wellbeing training locally. A contact following application, and an advanced instrument that assisted contaminated individuals with gaining admittance to wellbeing administrations and self-report side effects, was dispatched. The Government of Ghana likewise made arrangements for poor people, particularly individuals inside ghettos, and in-country networks to have free access to water and power. During lockdown periods in the enormous urban areas, the public authority gave food to weak individuals influenced.

The Ghanaian media has assumed a huge part in keeping the overall population educated about the pandemic, the public authority's administration plans, and preventive measures. Other than giving state-of-the-art data from true sources, some news sources run instruction projects and back local area commitment and execution of control measures (CNR, 2021).

As for immunizations, Ghana plans to inoculate 20 million of its 32 million population before the end of October 2021 and has sourced the necessary portions through multilateral arrangements and two-sided bargains. Ghana was the principal nation to get an antibody shipment from the COVAX office; somewhere in the range of 600,000 Oxford-AstraZeneca vaccines were followed through on February 24th, 2021. Vaccinations were made a priority for segments of the medical services labor force and security workforce, persons over 60 years old, those with comorbidities, and some administration officials beginning March 1,2022. Pregnant women and children under the age of 16 are currently prohibited due to limited information from immunization preliminary tests.

2.5 The impact of COVID 19 on Ghanaian economy

Since the first reported case, the effect of COVID-19 pandemic has been profoundly felt across the length and breadth of countries and landmasses throughout the planet. The lockdowns and different limitations set up influenced numerous organizations since they couldn't work. A large number of individuals lost their positions and vocations as numerous associations requested that their talented labor force telecommute. The government of Ghana carried out different monetary policies to assist with enduring the adverse consequences of the pandemic.

2.6 Mathematical Modeling

The best advantage of mathematical displaying is its capacity to in part supplant the costly and tedious cycle of building a test framework (Dechsiri, 2004). Typically, an interaction model of a genuine framework is created on a more limited size so the researcher may watch the conduct in a reproduction where every one of the factors can be controlled. The mathematical model that rises out of these limited scale tests gives a reasonable and quantitative portrayal of the discoveries, and in certain circumstances, it can even supplant the limited scale actual model (Dechsiri, 2004).

When all is said and done, a mathematical model can't capture the entirety of the particulars of a certifiable interaction, yet it can describe the framework's generally fundamental and huge detectable properties, which is adequate information (Dechsiri, 2004). Displaying is also a versatile cycle in which the outcomes of each exploratory endeavor are used to advance the following age model. This adaptability is one of the highlights of the displaying cycle. The demonstrating

cycle typically includes a few phases,

- select a mathematical model,
- match the model to information,
- redesign the model, and
- potentially extra tests.

Until the researcher is satisfied that she/he comprehends the framework (Dechsiri, 2004). The figuring limit of the present PCs, just as advances in the field of mathematical calculations, has diminished the time it takes to do a mathematical calculation from hours to seconds. Thus, the mathematical displaying of cycles has filled a noticeable quality in designing science. The most significant advantage of mathematical modelling is its ability to largely replace the costly and time-consuming interaction of developing a test framework (Dechsiri, 2004).

2.7 Classifications of mathematical models

Deterministic and stochastic mathematical models can be inexactly characterized into two gatherings (Dechsiri, 2004). Deterministic models are commonly composed as differential conditions that, when joined with start and limit conditions, precisely expect the conduct of a framework (Dechsiri, 2004). Arbitrary factors with obscure results are utilized in stochastic models, and we can just figure the likelihood of elective results (Dechsiri, 2004). The differentiation among deterministic and stochastic displaying approaches, however helpful, is a long way from clear. In the field of nonlinear dynamic frameworks, analysts have found that deterministic models can act in a stochastic way. When analyzed at the right scales or when figuring total amounts, stochastic models, then again, can have deterministic conduct.

Mathematical models are instruments for portraying actual reality, and the differentiation between stochastic and deterministic events isn't just about as clear here as it shows up from the outset (Dechsiri, 2004). It's a distortion to say that irregular examinations have a dubious end while traditional deterministic trials have an anticipated end, as they're ordinarily educated in essential courses. We'll utilize two guides to delineate our point. Tossing reasonable dice is a typical illustration of an irregular test that might be portrayed by an arbitrary variable that has a similar likelihood of taking the numbers in the set $\{1, 2 \cdots, 6\}$. A pass on, then again, is a strong body whose movement is characterized by an arrangement of differential conditions and is represented by Newton's laws of mechanics. If we know the starting conditions and the outside powers are following up on the dice, we can precisely expect the result of a throw (Dechsiri, 2004). The way that the last is, in every case, hard to foresee with sufficient precision is the significant reason for the far-reaching utilization of stochastic models for dice throwing. We'll look at the versatility of atoms in a liquid as a subsequent model. The movement of a solitary atom is administered by traditional mechanical standards. In any case, there are so many countless powers working on any atom because of crashes with different particles that anticipating the course of an individual molecule is practically incomprehensible (Dechsiri, 2004). Therefore, the only sensible way is to utilize a stochastic model, in this case, the Brownian movement. At the point when we see molecule movement at a full-scale level, we see deterministic wonders like dissemination transport, which can be handily addressed by utilizing differential conditions (Dechsiri, 2004).

Stochastic models model a solitary molecule's excursion through the reactor because of arbitrary elements (like vertical convection in wakes and isolation impacts), though deterministic models center around densities and, consequently, colossal quantities of particles. The law of enormous numbers guarantees the unsurprising conduct of, for instance, molecule densities, when countless particles are included, and this cutoff along these lines fills in as a connection among tiny

and perceptible models. In the long run, the long run produces similar outcomes in this situation. An infinitesimal model gives more data about the vacillations in the mean conduct if the quantity of particles picked is small. By and large, an infinitesimal model is simpler to conceptualize since it's anything but a solitary molecule's way, which is a straightforward and genuinely instinctive thing. Stochastic cycles address the advancement of an irregular framework across time. They get their name from the Greek expression 'stochastics,' which signifies 'speculating.' A stochastic cycle is a bunch of arbitrary factors with a record $X_{t\geq 0}$ and a boundary t. In our applications, t indicates time, and X_t depicts the condition of the framework, or the same part of it, at time t. The arrangement of potential upsides of X_t is called "state space", and will be meant by S. The state can, e.g., by the area of a given molecule in the reactors, wherein case S is some limited subset of R^3 (Dechsiri, 2004)

A deterministic model is characterized by a bunch of conditions that decisively clarify the framework's advancement through time. In a stochastic model, the advancement is essentially, to a great extent, irregular, and the interaction won't deliver indistinguishable results whenever rehashed on various occasions. The acknowledgment of a stochastic cycle alludes to various runs of the interaction. Stochastic models are harder to investigate than deterministic models.

2.8 Ideas of a Stochastic interaction

2.8.1 Stochastic cycle

A stochastic process is a family of random variables X_t defined on a probability space (W, L, P), where W is the sample space, L denoting a sigma field and P, a probability measure; with t denoting a parameter running over a suitable index set T(depending upon comfort, $X(t)$ can also be written rather than X_t).

A stochastic processes, then again, is a collection of random variables $X(t), t \geq 0$ index by time, t. The family of random variables may be continuous or discrete. The set of values taken by the states are called the State-space, and the individual values realized are called states. It is additionally workable for the state space to be discrete or persistent. The state space, or the scope of potential qualities for the arbitrary variable X_t, just as the time or boundary space, recognizes stochastic cycles. At the point when the ordering boundary is discrete, it is advantageous to indicate the ordering time as n and address the interaction as $\{X_n, n = 0, 1 \cdots\}$ (Bhat, 1984).

The potential upsides of t and X_t can be utilized to portray stochastic cycles as follows:

1. Discrete state, discrete time stochastic cycles
2. Discrete state, nonstop time stochastic cycles
3. Nonstop state, discrete time stochastic cycles
4. Constant state, nonstop time stochastic cycles

Independent and dependent stochastic cycles are the two kinds of stochastic cycles. For frameworks like rehashed tests, where future conditions of the framework are independent of various time states, freedom would be a helpful model (Dechsiri, 2004).

The decision between a deterministic and a stochastic model is impacted by an assortment of variables, including the idea of the interaction to be mimicked (Dechsiri, 2004). A few examples of completed paperwork for the use of a stochastic demonstrating procedure are as follows:

The cycle has a solid component of arbitrary movement of material particles. Stochastic displaying is naturally engaging for such frameworks, and it is viable with the idea of the cycle. In contrast with a deterministic model, a stochastic

model gives more data on factual vulnerability all the time (Dechsiri, 2004) The cycle is mind-boggling and includes numerous discrete occasions. To make deterministic models, the interaction should be changed to a constant cycle in both existences. This is frequently incomprehensible, or it necessitates broad suspicions. This isn't needed for a stochastic model, and the model can be communicated unequivocally. The cycles referenced in this proposition are instances of what should be possible (Dechsiri, 2004).

Different approaches to the same problem might lead to the same result in a variety of ways. They're all aiming for the same thing: to mimic and comprehend the real system as closely as possible. The final evaluation of which model to use comprises assessing which is more likely to result in the element of interest, as well as how much time and money are available for a system description (Dechsiri, 2004).

2.9 Markov Chain

Markov chains are a fundamental part of stochastic processes. A Markov chain is a model that explains the probability of sequences of random variables, or states, each of which can take on values from a range of possibilities.

A stochastic process is a Markov chain if it possesses Markovian properties. The Markovian property is simply that for the process, the future and past states of the process are independent if its present state is known (Leon-Garcia, 1994).

Various mathematical and normal epidemiological models have been proposed for investigating infectious illness dynamics. We can use such models to figure out how disease infections spread. However, most of these models are unable to forecast other critical disease metrics for both susceptible and infected individuals, such as the likelihood of first infection and recovery, as well as the projected time

for infection and recovery, to some extent.

Words, tags, or symbols representing anything, such as the weather, can be used to create these sets. When we need to compute a probability for a series of observable events, we can utilize a Markov chain. A Markov chain makes the extremely strong assumption that all that matters is the current state if we wish to predict the future in the sequence. The states that came before the current state have no bearing on the future unless they are influenced by the current state. It's as if you could look at today's weather to anticipate tomorrow's weather, but you couldn't look at yesterday's. Markov processes (or Markov chains) are well-known methods for modeling a wide range of phenomena in which a random variable's changes over time are represented by a sequence of future values, each of which is dependent only on the immediately preceding state and not on previous states (Bai et al., 2018). Markov processes are used to represent and analyze a wide range of phenomena in a variety of fields, including linguistics and biology, as well as political science, medicine, economics, and computer science (Bai et al., 2018).

2.9.1 Markov Chain Defined

A Markov chain, according to Suryaningrat et al. (2021), is a stochastic model that describes a succession of possible events in which the probability of each state or event is only dependent on the previous state. Serfozo (2009) defined a Markov chain as a stochastic process $X = Xn : n \geq 0$ on a countable set S is a Markov Chain if, for any $i, j \in S$ and $n \geq 0$,

$$P(X_{n+1} = j|X_0, \ldots, X_n) = P(X_{n+1} = j|X_n), \tag{2.1}$$

$$P(X_{n+1} = j|X_n = i) = P_{ij}^n \tag{2.2}$$

p_{ij} is the probability that the Markov chain jumps from state i to state j. These transition probabilities satisfy $p_{ij} = 1, i, j \in S$, and the matrix $P = (p_{ij})$ is the transition matrix of the chain.

The condition in equation (2.1), regularly known as the Markov property, expresses that, given the present status ,X_n, the following state ,X_{n+1} is restrictively free of the past $X_0, ..., X_{n-1}$. Markov chains, just like related ceaseless time Markov measures, are henceforth normal models or building blocks for applications. The condition in equation (2.2) simply states that the progress probabilities are independent of the time boundary n, demonstrating that the Markov chain is "time-homogeneous". If the change in probabilities were time elements, the cycle X_n would be a non-time-homogeneous Markov chain. These chains are like time-homogeneous chains, yet the time reliance adds additional bookkeeping components that we will not cover here.

Suryaningrat et al. (2021) define a Markov chain as a stochastic model that depicts a sequence of probable events in which the state-existence file may be used to identify the overall Markov chain. Using a grid and a vector, a Markov chain describes and predicts the behavior of a framework that starts with one state and then moves on to the next based on the current state (Suryaningrat et al., 2021). A discrete-time Markov chain (DTMC) is a discrete-time interaction with discrete boundaries or times that is referred to as a "Markov chain" (Suryaningrat et al., 2021).

2.10 Markov Processes

A stochastic augmentation of a limited-state robot is a Markov cycle. State changes in a Markov interaction are probabilistic, and in contrast to a limited state machine, there is no contribution to the framework. Moreover, at each time step, the framework is just in one state. Markov measures, to put it another

way, are models for the advancement of arbitrary events whose future conduct is unaffected by their current state.

A Markov process is a stochastic process with the following characteristics:

- There are a finite number of possible outcomes or states
- The outcome of any stage is solely dependent on the previous stage's outcome
- Over time, the probabilities remain constant

2.11 Applications of Markov modeling

Standard epidemiological and numerical models have been developed by statistician to explore infectious elements, some of which can be carried out using deterministic or stochastic approaches. Because most processes, such as disease elements, are stochastic, analyzing them deterministically may not be appropriate.

In most cases, prior data on epidemiological amounts, such as contact rate, illness occurrence, power of contamination, regenerative number, transmission speed, and so on, is also necessary. In nations like Sub-Saharan Africa and other locations where infection impacts are predominantly investigated using numerical models, these numbers may not be available. As opposed to Markov chain models, complex methods and reenactment approaches are employed without a fraction of these amounts, which may be difficult to learn and implement by different modelers who are not mathematically disapproved (Zipkin et al., 2010).

There are other critical illness indicators that neither the deterministic nor alternative stochastic S-I-R models can assess. The risk of initial disease, as well as the potential for contamination for both powerless and polluted persons, are

among the other important factors to consider. As a result, Markov models can be used to address the concerns outlined above.

Beck & Pauker (1983) introduced Markov chain models to clinical writing and provided a demonstrated design that has been used in numerous subsequent studies to aid judgments between risky clinical drugs with costs and outcomes occurring now, as well as in the near and distant future.

In his examination, Kadhem & Kadhim (2022) proposed a dynamical model strategy for assessing COVID-19 diseases called engrossing Markov chains. He needed to target and expect two phases of ingestion from their demonstrating approach: recovery and passing because those two conditions were viewed as fundamental signs in surveying the wellbeing status. As per the investigation, there is a slow expansion in the expected mortality number while a lessening in the quantity of recuperated people, in light of the engrossing Markov model.

J. Chen et al. (2020) developed another discrete-time Markov chain model based on their findings, which directly mixes stochastic lead and requires limited evaluation using available data. Using data from China's Hubei domain, the model proved to be adaptable, intense, and careful (for which Wuhan is the ordinary capital city and which addresses around 82% of the full-scale definite COVID-19 cases in the country). The major Shanghai help clinical gathering in Wuhan's Jinyintan Facility, which was the world's most recently assigned clinical center to recognize COVID-19 patients, has recognized it in this method. The estimate was utilized to help with clinical organization decisions, as well as to design clinical staff, ICU beds, ventilators, and other clinical resources for consideration.

Ghosh & Chakraborty (2021) developed a coordinated deterministic–stochastic approach for predicting the drawn-out orientations of the COVID-19 instances in Italy and Spain in their investigation. The deterministic component of the everyday case univariate time series was investigated using an upgraded variation of the

SIR [Susceptible–Infected–Recovered–Protected–Isolated (SIRCX)] model, while the stochastic part was dissected using an autoregressive (AR) time series model. The proposed coordinated SIRCX-AR (ISA) approach relies on two functionally unmistakable showcasing standards that use the predominance of both the deterministic SIRCX and stochastic AR models to locate the drawn-out directions of the plague bends. Compared to the ODE-based SIRCX model, trial testing using the suggested ISA model revealed significant improvements in long-term COVID-19 example determination for Italy and Spain.

To track COVID-19's progress in Morocco from March 14 to October 5, 2020, Marfak et al. (2020) proposed using the Hidden Markov Chain, a factual framework that models changes from one state (confirmed cases, recovered, dynamic, or death) to another as per a progress likelihood lattice. They emphasized that information for the total number of confirmed, recovered, active, and passing cases should aid Moroccan professionals in developing appropriate methods for dealing with the consequences of COVID-19 detention.

Wu, Zheng, & Chen (2020)used a state change framework model to predict COVID-19 progression in overall basic plague zones. The preparation set was created using data from the National Health Commission of the People's Republic of China (December 31, 2019, to March 5, 2020) and World Health Organization (January 20, 2020, to March 5, 2020) reports, while the approval set was created using data from March 6 to 9. They gathered and examined new close contacts, newly confirmed cases, aggregated confirmed instances, non-extreme cases, serious cases, basic cases, relieved cases, relieved cases, and demise. The information was examined using the State Transition Matrix model. The optimistic situation (non-Hubei model, day-by-day increase pace of 3.87 percent), the mindfully idealistic situation (Hubei model, day-by-day increase pace of 2.20 percent), and the somewhat critical situation (change, day-by-day increase pace of 1.50 percent) were all deduced and demonstrated using the results. South

Korea's IFP is scheduled for March 6 to 12, Italy's is scheduled for March 10 to 24, and Iran's is scheduled for March 10 to 24. In the best-case scenario, the total number of confirmed cases in South Korea, Italy, and Iran would be roughly 20,000, 209,000, and 226,000, respectively. They did say, however, that if optional conclusion criteria were used, the variety of new occurrences could cause a variety of repercussions in the predictive model. They discovered that their protesting demonstrated that the conclusion of the episode was not unavoidable and that the number of critically wiped-out people would continue to rise.

Cauchemez et al. (2004) used a Bayesian MCMC approach to investigate the aspects of flu in longitudinal data, focusing on the contamination risk in the home and surrounding region. During a 15-day visit by an overall specialist to 334 homes, subjects were identified as both susceptible and irresistible. One of their significant flaws was that the transmission model's boundary gauge was unclear, given that a big chunk of the irresistible cycle was not visible and only dates of new irresistible disclosure were recorded. The transmission model, which had a three-level progressive structure, fully exploited the observational, transmission, and prior levels. To explore the joint back conveyance of model boundaries and improve information from a Bayesian perspective, the Markov chain Monte Carlo (MCMC) approach was used. In any event, the sufficiency of the Bayesian deduction is also influenced by the circumstances in which the information or noted examples were generated randomly. The average irresistible period was found to be 3.8 days, with the risk of flu spreading between an infected and a vulnerable person decreasing as the size of the family grows (0.32 individuals each day). The immediate risk of disease from the territory was calculated to be 0.0056 per day. As a result, they discovered that while children (under the age of 15) were more likely to transmit the virus than adults, the interval of contamination for both children and adults was quite similar. As a result, the back probability that children were more helpless than adults was 76%, and the back probability that they were more vulnerable than adults was 79%. Various examinations are de-

duced using Bayesian methods. Most of these approaches rely on Markov chain Monte Carlo (MCMC) algorithms to generate tests from back-appropriated data. The Markov chain's assembly to a given distribution determines the accuracy or execution season of these calculations. In any case, there is no reason to believe that the Markov chain will join rapidly or gradually in a random situation. This emphasizes the importance of developing plausible hypothetical systems for Bayesian decisions.

Wu, Darcet, et al. (2020) used the strategic development model, the summed up calculated development model, the summed up development model, and the summed up Richards model to demonstrate the COVID-19 flare-up throughout 29 regions in China and around the world. Wu, Darcet, et al. (2020) aligned the calculated development model, the summed up strategic development model, the summed-up development model, and the summed-up Richards model to the detailed number of tainted cases for the entire nation of China, 29 regions in China, four seriously influenced nations, and Europe as a whole from January 19 to March 10 for the entire nation of China, 29 regions in China, four seriously influenced nations, and Europe as a whole (2020). To provide upper and lower boundaries for specific situation projections, multiple models were employed. They showed high risk in Japan as a result of the discovery, with absolute confirmed cases expected to be 1574 (95 percent CI: [880, 2372]) as of March 25 and 5669 (95 percent CI: [988, 11340]) by June. They predicted that the number of infected cases in South Korea will reach 7928 in 20 days (95 percent CI: [6341, 9754]). In a positive scenario, they estimated that 0.15 percent of the Italian population would be polluted (95 percent CI: [0.03 percent, 0.30 percent]). In the worst-case scenario, 114867 persons in Europe would be polluted in ten days, accounting for 0.015 percent of the population. According to their findings, China's harsh regulatory efforts were extremely effective, resulting in certain informative variations among districts. In the following months, the spread of the pandemic to other countries appears to be virtually unavoidable. During the epidemic, Japan

and Italy were in critical waterways, with not a single fast finish to be found. There was a big risk associated with the 2020 Summer Olympics in Tokyo, which would be held in July. They say Iran's condition is particularly hazardous, with unknown and terrifying future scenarios, whereas South Korea appeared to be nearing the end of the event. Both Europe and the United States were in the early phases of the episode without major improvements, posing significant health and financial concerns to the rest of the world.

Zaki et al. (2012) investigated the time-fluctuating generation number of the COVID-19 episode in China. You et al., (2020) used three methodologies to estimate the critical and controlled proliferation numbers: Poisson probability-based procedure (ML), dramatic development rate-based strategy (EGR), and stochastic Susceptible-Infected-Removed powerful model-based technique (SIR). As a result of the findings, a total of 198 chains of transmission were discovered among 14,829 confirmed cases outside Hubei Province as of March 31, 2020, including dates of onset of side effects and 139 dates of illnesses. The overall sequential stretch was 4.60 days, with a standard deviation of 5.55 days, the brooding time frame was 8.00 days, with a standard variance of 4.75 days, and the irresistible length was 13.96 days, with a standard deviation of 5.20 days. The assessed controlled generation numbers, Rc, generated by each of the three techniques are much lower than the essential multiplication numbers, R0, in all of the studied areas of China. The number of regulated propagations in China is now much fewer than one across all regions of the country, according to You et al (2020). It fell below one within 30 days of receiving exceptional control measures, demonstrating that the Chinese government's concerted efforts to limit the disease were successful. Regardless, steps to halt the current outbreak are still needed, as imported cases from various countries pose a significant risk of a flare-up.

Musa et al. (2020) looked at the main phase of the COVID-19 episode in Africa from March 1 to April 13, 2020, using the essential dramatic development model.

They used publicly available materials from the WHO situation report to demonstrate how COVID-19 could spread if comprehensive health insurance were not maintained. The Poisson probability system is used for fitting data and determining boundary conditions. They calculated the critical proliferation number by extrapolating the COVID-19 age span (GI) circulation as Gamma appropriations with a mean of 4.7 days and a standard deviation of 2.9 days from before work. Because remarkable development will begin on March 1, 2020, the projected daily rate of dramatic development is 0.22 (95% confidence interval: 0.20-0.24), and the fundamental multiplication number, R_0, is 2.37. (95% CI: 2.22-2.51). With a R_0 of 2.37, we determined the episode's prompt contagiousness utilizing the time-changing compelling regenerative number to show COVID-19's capacity to spread across Africa. As per Musa et al. (2020), the yearly increment of COVID-19 cases in Africa was speedy and fluctuated enormously between countries. And they opined that their projections ought to demonstrate the importance of planning for the development of the COVID-19 pestilence in Africa.

Zhang et al. (2021) investigated the pandemic status of new coronary pneumonia (COVID-19) and the clinical danger features of patients based on cardinal data and clinical side effects using an epidemiological Markov model. Between January and May 2020, they collected 500 patients with COVID-19 who were diagnosed by nucleic corrosive testing in the X emergency clinic using their method. They were divided into two categories based on the severity of the illness: general (200 examples) and extreme basic (100 cases) (300 cases).To estimate the amount of COVID-19 contamination, a Markov model was used. The patients' overall data, clinical qualities, and safety procedures were thoroughly investigated. The findings revealed that the average length of stay for COVID-19-infected patients was 14 days, as demonstrated by Markov model data. Sexual orientation, age, hypertension, coronary artery disease, windedness, myocardial damage, and thrombocytopenia were all significantly different between the two groups of patients ($P<0.05$). Sex, age, and the presence of hypertension, coronary sickness, wind-

edness, myocardial injury, or thrombocytopenia, as determined by a strategic multivariate relapse examination, were all clinical danger factors for COVID-19 patients. Zhang et al. (2021) believe that the Markov model can be used to predict the COVID-19's time course at various stages of development. Furthermore, the COVID-19 virus, which has spread swiftly, is extremely dangerous. Dynamic anticipation can improve the beneficial impact just as much as the patient's respiratory capacity and, more broadly, personal happiness in clinical settings.

The geographic dispersion features of COVID-19 and its forecast were analyzed by Dehghan Shabani & Shahnazi (2020). The Markov chain and the spatial Markov chain were used in this study. In comparison to the stochastic piece strategy, the Markov chain technique has the advantage of providing data on the movement of areas within the appropriation Le Gallo (2004). The findings suggest that COVID-19 in Asia did not always align with current policies and that COVID-19's spread was influenced by its surroundings. Policymakers should instead embrace cross-public collaborative tactics rather than native limiting infrastructure systems.

Giuliani et al. (2020) suggested a multivariate time-series linear model comprehend and measure the spatio-fleeting dispersion of the COVID-19 pestilence in Northern Italy. When compared to the most recent information accessible in Italy, the proposed model can predict with a 3% error rate, according to the experts. The spatio-fleeting distribution of COVID-19 contaminations at the neighborhood level was estimated using formally available data on the number of contaminated at the most significant level of spatial area (Italian territories). An endemic-plague time-series blended impact model for areal ailment tallies was used to explain and predict the wonder's spatio-transient propagation. Three subcomponents of the fitted model are reliant on the gathered outcomes. The first explained how the disease spread within territories, the second explained how it spread between regions, and the third explained how the disease progressed over

time. Control measures at the local level initially influenced territories that were not influenced by the effect of spatial neighbors. In territories that were heavily influenced by diseases, however, the segment expressing spatial contact with adjoining regions was prominent. Furthermore, the proposed model generates excellent forecasts for the number of illnesses at the local level after allowing for delayed disclosure. Following extensive research, it was determined that strict control measures implemented in specific territories effectively broke viruses and limited their spread to surrounding areas. While area-specific regulation plans may be the more effective, successful, and uniform authorization of control measures at the public level is required to prevent infectious disease prevention from being postponed or ignored in general. This could also be true on a worldwide scale if internal border checks between states have been liberally eliminated, as they have been in the European Union and the United States.

Kucharski et al. (2020) developed a numerical model to better understand the early stages of the COVID-19 pandemic. The entire population was divided into four groups: vulnerable, exposed, irresistible, and extinguished.

Catak & Duran (2020) propose a nonlinear Markov bound model to deconstruct and analyze the novel (COVID-19) pandemic in their study. The model that was presented was created using information from China. The suggested nonlinear Markov chain model was then utilized to consistently assess fresh COVID-19 instances in a few countries. The results reveal that the suggested model can accurately predict the overall course of the COVID-19 pandemic dissemination. They showed that all revealed data, including the number of new COVID-19 cases per day, the number of passings per day, and the number of COVID-19 test outcomes per day, was considered precise.

2.12 Application of Machine Learning for Time-series Forecasting

In many domains, such as business, energy, and the environment, time series forecasting is a typical activity. Forecasts that are accurate can save a lot of time and money by making planning, scheduling, and other tasks easier (Kim, 2003; Dudek, 2016; Araújo et al., 2017). There are times when a large number of time series must be forecasted fast, such as when selling various retail products. Fast forecasting algorithms that may be implemented automatically are extremely important in this scenario. Some common techniques are prohibited by these restrictions; for example, the Autoregressive Integrated Moving Average (ARIMA) methodology is typically used under the supervision of an expert, whereas other forecasting tools are not yet at the expense of high computational demands (Box et al., 2015; Hyndman & Athanasopoulos, 2018). These automated forecasting tools are termed machine learning time-series models.

Machine learning (ML) is a type of computational intelligence that uses pre-programmed algorithms to analyze input data and learn from it via supervised or unsupervised processes in order to anticipate output values that are within a reasonable range (Twumasi & Twumasi, 2021). Over the last decade, machine learning models have proven to be a superior alternative to traditional statistical models for forecasting and other research tasks (such as regression and classification). The development of ML algorithms began with the neural network (NN). Deep learning methods known as NNs were developed as mathematical models of the brain. Complex non-linear relationships between the response variable and its predictors can be modelled using ML algorithms. Other researchers used neural networks to produce Decision trees, Random forests, Gradient Boosting Machines, Support vector machines, and other machine learning methods for dealing with regression analysis problems (Alpaydin, 2020; Hastie et al., 2009)

Parallel efforts have been made to empirically validate current ML models, compare models, and construct new ones. The enormous importance of these ML breakthroughs to modellers provides a wide range of options and a thorough grasp of the strengths and limitations of available models for various forecasting situations. Multi-Layer Perceptron (MLP), Bayesian Neural Network (BNN), Radial Basis Functions (RBF), Generalized Regression Neural Network (GRNN), K-Nearest Neighbour regression (KNN), CART regression trees (CART), Support Vector Regression (SVR), Recurrent Neural Network (RNN), Long Short Term Memory neural network (LSTM), Automated Artificial neural network (AANN), and Gaussian Processes (GP) are the most popular (Ahmed et al., 2010; Makridakis et al., 2018)

Coronavirus (COVID-19) has remained a global public health issue. In reality, keeping track of the virus's spread necessitates unrestrained effort. However, it is assumed that using an effective prediction of the prevalence is the most important condition for efficiently controlling the spreading rate. Time series models have long been regarded as one of the most practical ways for predicting disease prevalence or dissemination rates. Many empirical research have been undertaken in various areas to forecast the probability of COVID-19 cases using the ARIMA model (Gupta & Pal, 2020; Yonar et al., 2020; Shi & Fang, 2020; Takele, 2020). In this study ML time-series methods are compared to the traditional ARIMA time-series model in forecasting COVID-19 infection cases in Ghana. Within the forecasting domain, machine learning (ML) algorithms for time-series forecasting have improved dramatically over time and are now considered strong competitors to traditional methods (Twumasi & Twumasi, 2021)

Chapter 3

Methodology

3.1 Introduction to Chapter

The methodology section contains a detailed summary of the study's empirical datasets and the models employed for statistical analysis. These models include the discrete-time Markov Model and the Machine Learning (ML) time-series forecasting algorithms. These models are imperative to satisfy both the initial and secondary objectives of the current study. Here, we propose an S-I-D discrete-time Markov model to estimate other relevant disease metrics for COVID-19 transmission, which are difficult to determine using the traditional epidemic models at two infection periods: years 2020 and 2021 in Ghana. A detailed framework of the modified Susceptible-Infected-Dead (S-I-D) Markov model and other theoretical justifications have been developed. In addition, the state-of-the-art ML models for time-series forecasting are outlined.

3.2 Data Source and Structures

A publicly accessible national aggregated data on the counts of susceptible, infected, recovered, and COVID-19 induced deaths (for the years 2020 and 2021) and the time-series data on the daily cases of COVID-19 (from March 14, 2020, to October 19, 2021) in Ghana, were obtained from the Ghana Health Service, Worldometer, World Population Review websites, and Central Intelligence Agency (CIA) factbook. For each year (2020 and 2021), the total number of

susceptible individuals were found by subtracting the sum of the total number of COVID-19 infected individuals and dead individuals (per infection period) from the mid-year population (as at the time of the study). The mid-year populations for the years 2020 and 2021 were respectively 31,072,940 and 31,732,129 (according to United Nations population-based estimates, which were retrieved via the URL link: `www.worldometers.info/world-population/ghana-population`). The number of dead individuals in Ghana was estimated for 2020 and 2021, based on the per-capita death rate (provided in the CIA factbook) as a function of the mid-year population; retrieved from the URL links: `https://www.indexmundi.com/g/g.aspx?c=gh&v=26` and `www.indexmundi.com/ghana/death_rate.html`. Ghana's COVID-19 cases (just like other countries) were influenced by several factors (such as diagnostic, screening coverage, and complete reporting), not ignoring the possibility of under-reported cases.

3.3 Markov Chain Model

3.3.1 Concepts of Stochastic processes

Let's have a look at a coin tossing trial. It can result in a heads (H) or tails (T) outcome. These are referred to as outcomes, and they all add up to $S = \{T, H\}$, which is referred to as a sample space. If we define a function that gives $H = +1$ and $T = -1$, then $X : S \to R, X(T) = -1, X(H) = 1$ is referred to as a random variable. A stochastic process is a collection of well-defined random variables $X_t, t \in T$ on the same probability space, where t denotes time. Thus, a stochastic process is a function that assigns elements of an index set to random variables in a collection. Normally, random variables are indexed by points in time that indicate the chance of a specific result. "States" are the values that the random variable assumes, and a "state-space" is a collection of states. One of the simplest

instances of a stochastic process is a person's commute to work. Let's pretend that the commute to work is made up of different routes, each of which represents a state in a stochastic process. When a person starts on one path and then switches to another, this is an example of states transitioning in a stochastic process. The individual's movement to the next route is determined by the present route and is unaffected by previous routes, much as the state of a stochastic process is determined by the current state and is unaffected by previous states. For each fixed t, $X_t(s)$ signifies a single random variable specified on S. $X_t(s)$ corresponds to a function defined on T for each fixed $s \in S$, which is referred to as a sample path or a stochastic realization of the process.

3.3.2 Markov Chain

Bhat (1984) proposed the notion of stochastic processes and included the following submissions in his definition:

A stochastic process is a set of random variables X_t defined on a probability space (Ω, L, P), with t signifying a parameter running over an appropriate index set T (subject to convenience, $X(t)$ can alternatively be expressed instead of X_t). The index t corresponds to discrete units of time with the index set $T = 0, 1, 2, ...$ in most cases. In other words, a stochastic process is a set of random variables $X(t), t \geq 0$, indexed by the time parameter t. However, the parameter space is the set of potential values of the indexing parameter t, which can be continuous or discrete. The set of alternative values is referred to as the state space, and the values assumed by the process are referred to as states. It is also possible for the state space to be discrete or continuous. The state space, or the range of possible values for the random variable X_t, as well as the time or parameter space, distinguish stochastic processes. When the indexing parameter is discrete, it is convenient to refer to the indexing time as n and the process as $X_n, n = 0, 1, ...$ (Bhat, 1984)

The Markov chain is a stochastic process with a countable number of values $X_n, n = 0, 1, 2, ...$ It is represented by the process's set of possible values, which excludes negative integers. Consider the random variable $X_n = i$, which denotes that the process is in state i at time n. Then, let $P(X = i)$ be is the probability of being in state i, and $P(X = j)$ is the probability of being in state j. The likelihood that the process will move to state j when it is in state i is given as $P(X_{n+1} = j \mid X_n = i) = P_{ij}$. Markov stochastic processes do not rely on previous behavior to guide the future; instead, they are impacted purely by the present. Moving states come with a few conditions:

1. $P_{ij} \geq 0$
2. $i, j \geq 0$
3. $p_{ij} = 1$
4. $i = 0, 1, 2, ...$

The application Markov chain processes is one of the most well-built of all stochastic processes. Because both time and state-space will employ the well-known discrete time Markov chain (DTMC), we will only undertake state classification in circumstances when we are dealing with state relationships. The **accessibility** of states to one another is the first necessary definition. If $P_{ij}^n > 0$, we say state j is accessible from state i, denoted $i \to j$. For example, $P_{ii}^0 = 1$, we can suppose that every state is accessible from itself. Another important term is **communicate**; two states i and j are said to communicate if they are accessible from one other and are denoted by $i \leftrightarrow j$. **Communication classes**, which are made up of members of the same class conversing with each other, can be formed from Markov chains. If a Markov chain has only one communication class, all states communicate with each other, it is said to be **irreducible**. Also, it is said to be recurrent if we leave a state and return to it with a probability of 1 in the future.

Nonetheless, if the number of times you return to the same state is less than one, we call it transient.

One of the simplest ways to tell if two states belong to the same class is if they are both recurrent or transient. A communication class is closed if one-step transitions from any state in the set of states to any state outside the set of states are difficult. When a process is in state i at time n and the next time step is $n+1$, it is called a transition probability. Transition probabilities are associated with a variety of state changes, but they will either remain in state i or shift to state j. Transition probabilities that are not time dependent are called stationary or homogeneous, while those that are time dependent are called non-stationary or non-homogeneous.

3.3.3 Probability Distribution

The stochastic process $X(t)$ is a basic random variable with a probability distribution that can be realized for any other random variable if the time parameter is set to t. The joint distribution of the basic random variables of the family $X(t), t \in T$ is required for complete information on the process. Let $(t_1, t_2, ..., t_n)$ be a discrete set of points within T, with $t_1 < t_2 < t_3 < \cdots < t_n$. At these moments, the joint distribution for the process $X(t)$ can now be defined as;

$$P\big[X(t_1) \leq x_1, X(t_2) \leq x_2, \cdots, X(t_n) \leq x_n\big]. \tag{3.1}$$

For a given value of the time parameter, there are conditional probability functions depending on some stochastic process information. Let t_0 and t_1 be two points in T where $t_0 \leq t_1$ is the case. Following that, the conditional transition distribution function is defined as follows:

$$F(x_0, x_1; t_0, t_1) = P\big[X(t_1) \leq x_1 \mid X(t_0) \leq x_0\big]. \tag{3.2}$$

The transition probabilities can be defined as;

$$P_{i,j}^{(m,n)} = P\big[X_n = j \mid X_m = i\big] \tag{3.3}$$

when the stochastic process contains discrete parameter space (m, n) and state space (i, j).

Definition 1. *If the transition distribution function given by 3.1 depends solely on the difference $t_1 - t_0$ instead of t_0 and t_1, a stochastic process $\{X(t), t \in T\}$ is said to be time homogenous. Then,*

$$F(x_0, x : t_0, t_0 + t) = F(X_0, x; 0, t) \tag{3.4}$$

for $t_0 \in T$. For the sake of convenience, 3.4 can be written as $F(x_0, x; t)$. The process's related expression $\{Xn, n = 0, 1, \cdots\}$ would then be $P_{ij}^{(n)}$. Let $F(x, t)$ be the cumulative distribution of the process $X(t)$ expressed as $F(x, t) = P[X(t) \leq x]$.

Also, let $f(x)$ be the probability density function of the process X at state 0 (that is, $X(0)$). Now, $F(x, t)$ can be determined from $F(x_0, x; t)$ using the expression;

$$F(x, t) = \int_{x_0 \in S} F(x_0, x, t) f(x_0) dx_0. \tag{3.5}$$

The state space is denoted by S. The equivalent relation in the discrete case is of the type,

$$P_j^n = \sum_{i \in S} P_i^0 P_{ij}^n. \tag{3.6}$$

3.3.4 Defining Markov Processes

We will discuss the basic definitions and properties of Markov chains (or Markov processes) with discrete time and spaces, whose state space may be finite or count-

ably infinite in this section (Bhat, 1984). In most real-life circumstances, stochastic processes are such that for a discrete collection of parameters $t_1, t_2, t_3, \cdots, t_n \in T$ there is some form of dependence between the random variables $X(t_1), X(t_2), \cdots, X(t_n)$.Clearly, as the dependency structure becomes more complex, the process analysis becomes more difficult. The first-order dependence is the most basic sort of stochastic process dependence. This is referred to as Markov dependency, and it is defined as follows: Consider a set of points that is finite (or countably infinite); $(t_0, t_1, t_2, t_3, ..., t_n, t), t_0 < t_1 < \cdots tn < t$ and $t, t_r \in T(r = 0, 1, 2, ..., n)$ where T is the parameter space of the process $X(t)$. The dependence demonstrated by the process $X(t), \in T$ is termed Markov-dependence if the conditional distribution of $X(t)$ for specified values of $X(t_1), X(t_2), \cdots, X(t_n)$ rely only on $X(t_n)$ This is the process's most recent known value.

Definition 2. *If a stochastic process $\{X(t), t \in T\}$ has a finite or countable state space and meets the Markov property described in 3.7, it is said to be a Markov chain; with*

$$P(X_t = i_t \mid X_{t-1} = i_{t-1}, \cdots, X_1 = i_1, X_0 = i_0) = P(X_i = i_t \mid X_{t-1} = i_{t-1}). \tag{3.7}$$

In other terms, a Markov process is a stochastic process that exhibits the properties indicated in 3.7. If the state of a Markov process is known for any given value of the time parameter t, that information is sufficient to forecast the process' behavior beyond that point. The following relationships can be drawn as a result of the property provided by 3.7:

$$F(x_0, x; t_0, t) = \int_{y \in S} F(y, x; \tau, t) dF(x_0, y; t_0, \tau), \tag{3.8}$$

where $t_0 < \tau < t$ and S is the state of the process $X(t)$. When the stochastic process has a discrete state space and a discrete parameter space, 3.7 and 3.8 can

also take the following forms, as shown in 3.9 and 3.10 for $n > n_1 > n_2 > \cdots > n_k$ where n and $n_1, n_2, \cdots, n_k$ belonging to the parameter space:

$$\begin{aligned} P(x_n = j \mid X_{n1} =& i_1, X_{n2} = i_2, \cdots, X_{nk} = i_k) \\ =& P(X_n = j \mid X_{n1} = i_1) \\ =& P_{i,j}^{(n_1,n)} \end{aligned} \tag{3.9}$$

Using this property in equation (3.9), for $m < r < n$ leads to the Chapman-Kolmogorov equation,

$$\begin{aligned} &\sum_{k \in S} P(x_n = j \mid X_r = k) P(X_r = k \mid X_n = i_1) \\ &= \sum_{k \in S} P_{ik}^{(m,r)} P_{kj}^{(r,n)} \end{aligned} \tag{3.10}$$

where S denotes the state space of the $X(t)$ process. The Chapman Kolmogorov equations for the process are 3.8 and 3.10, and these are the fundamental equations in the study of Markov processes. They allow for the creation of transition probability linkages between any points in T when the processes display the Markov-dependence trait. As shown in Table 3.1, Markov processes can be classified into four categories. Markov chains are discrete processes in which the parameter and state space (SS) are both discrete. For the purposes of this research, Markov Chains and Markov processes with discrete state spaces are the main topics of interest. For the purposes of this study, Markov chains with continuous state spaces are not considered.

Table 3.1: Types of Markov Processes

Parameter Space	**Discrete SS**	**Continuous SS**
Discrete	X(Markov Chain)	X
Continuous	X	X

3.3.5 Transition Probability

A Markov chain with a state space $S \subseteq \xi = 0, 1, 2, ...$ is defined as $X_n, n = 0, 1, 2, \cdots$. The state space S would be supplied by the set $1, 2, \cdots, m$, since a finite m-state chain is of interest. The nth-step transition probabilities will be denoted as, in general, only time-homogeneous Markov chains will be studied.

$$P_{ij}^{(n)} = (X_n = j \mid X_0 = 1) \tag{3.11}$$

It is worth noting that the word "nth-step" refers to the amount of time between observations. When $n = 1$, $P_{ij}^{(1)} = P_{ij}$ is true. It is easier to organize these transition probabilities from one state i to another state j in a matrix form below because of the dual subscript.

$$\begin{bmatrix} P_{00} & P_{01} & P_{02} & \cdots \\ P_{10} & P_{11} & P_{12} & \cdots \\ P_{20} & P_{21} & P_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

and

$$\begin{bmatrix} P_{00}^{n} & P_{01}^{n} & P_{02}^{n} & \cdots \\ P_{10}^{n} & P_{11}^{n} & P_{12}^{n} & \cdots \\ P_{20}^{n} & P_{21}^{n} & P_{22}^{n} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

where

$$\sum_{j \in S} P_{ij}^{(n)} = 1, n = 1, 2, \cdots, \quad i = 1, 2, \cdots$$

The nth-Step Transition Probability Matrix

Assume that P is the transition probability matrix of a Markov chain with $\{X_n, n = 0, 1, 2, \cdots\}$ as given above. Let $P_j^{(n)}$ be the probability that the process will be in state j following the n transition. The row vectors of the matrix $P^{(n)}$ are represented by the probability vector $P_j^{(n)}, j \in S$. 3.3.1 determines the n-step transition probabilities $P_{ij}^{(n)}$ and the cumulative probabilities $P_j^{(n)}, ij \in S$.

Theorem 3.3.1. *If P is the $\{X_n\}$ Markov chain's transition probability matrix, then*

$$P^{(n)} = P^n \tag{3.12}$$

and

$$P^{(n)} = P^0 P^n \tag{3.13}$$

Theorem 3.3.2. *Let P be the transition probability matrix of an irreducible, aperiodic n-state homogeneous Markov chain with finite time states; then*

$$\lim_{n\to\infty} P^n = \pi = \begin{bmatrix} \alpha \\ \alpha \\ \vdots \\ \alpha \end{bmatrix}$$

whwere $\alpha = (\pi_1, \pi_2, \cdots, \pi_m)$ with $0 < \pi < 1$ and

$$\sum_{i=1}^{m} \pi_i = 1$$

and hence;

i (i) $P(t) = [P_1(t), P_2(t), ..., P_m(t)]$

ii $\exists$ *constants* c *and* r $(c > 0, 0 < r < 1)$ *such that* $|P_{ij}^n - \pi_j| \leq cr^n$ $\forall$, $j = 1, 2, ..., m$.

iii $P\pi = \pi P = \pi$

Note that the convergence property in (ii) is known as geometric ergodicity, and a chain possessing this attribute is said to be strongly ergodic.

Theorem 3.3.3. *Any finite irreducible time-homogeneous Markov chain,* $\{X_m \mid m \in T\}$ *with transition probability* P *, has a stationary distribution* V *given by* $VP = V$ *, where* V *is a (row) probability vector which can be found using the equation and the constraint* $VI = 1$ *with* I*, being a column vector as the same dimension as the vector* V *and all components equal to unity. In other words, a finite irreducible-aperiodic Markov chain* $\{X_m \mid \in T\}$ *where a regular transition matrix* P *has a limiting distribution* α*, given by* $\alpha P = \alpha$ *where* $\alpha = (\pi_1, \pi_2, \cdots, \pi_n)$ *assuming* n *states with*

$$\pi_i = \lim_{n \to \infty} P(X_n = i) \tag{3.14}$$

Using the equation given by $\alpha P = \alpha$ together with $\alpha I = 1$ (constraint), α can be found (Twumasi, 2018)

3.3.6 S-I-D Modelling by Markov Chain Approach

With regards to this study, the Markov chain model is used as an alternate approach to studying the disease dynamics of COVID 19. The S-I-D represents Susceptible, Infected, and Death, respectively. Below are the meanings of basic terms of disease or infection dynamics as suggested.

Susceptible (S): it is comprised of individuals who are prone to the disease or individuals who are not infected with the disease.

Infectious state (I): it is comprised of infected individuals and carriers of the disease, or it is comprised of infected individuals and carriers of the disease..

Death state (D): it is comprised of individuals who died from the COVID 19 disease in the course of the study period (March 2020 to October 2021).

Consider three discrete states: susceptible (state 0), infected (state 1), and death (state 2). If $\{X(t), t \in T\}$ denotes the number of people in each stage of the underlying disease (COVID-19) at any given time t, then $\{X(t), t \in T\}$ is clearly a stochastic process with states $0, 1$, and 2. As a result, the statistical model for the first-order time-homogeneous Markov dependency is:

$$P(X_n = i_n \mid X_{n-1} = i_{n-1}, \cdots, X_1 = i_1, X_0 = i_0)$$
$$P(X_n = i_n \mid X_{n-1} = i_{n-1}). \tag{3.15}$$

Then, for $i, j = 0, 1, 2$, the transition probability P_{ij} is given in matrix form as;

$$\begin{pmatrix} P_{00} & P_{01} & P_{02} \\ P_{10} & P_{11} & P_{12} \\ 0 & 0 & 1 \end{pmatrix}$$

where

$$\sum_{j=0}^{2} P_{ij} = 1, i = 0, 1, 2.$$

The Chain's Specifications (Probabilities of transition):

P_{00}: Probability of remaining in a susceptible state.

P_{01}: transition Probability from susceptible state to infectious state.

P_{02}: transition probability from susceptible to a dead state.

P_{10}: transition probability from infectious state to susceptible state

P_{11}: Probability of remaining in an infectious state.

P_{12}: transition probability from an infectious state to a dead state.

P_{22}: Probability of remaining in a dead state.

Remarks

In the literature, the parameter P_{01} is commonly referred to as discrete time force of infection. In addition, the elements P_{02} and P_{12} represent mortality for uninfected and infected people, respectively, whereas P_{10} represents the chance of recovery or defection (Cohen, 1973). Because there is no chance of being vulnerable or sick in the removed condition, it is an absorbing state. The yearly time step unit is used to provide a smooth transition from one stage to the next.

Model's Assumptions

- An individual's current state is solely determined by the state of the individuals at the preceding time step.
- In the dead state, an individual cannot be susceptible or infected.
- Transitioning probabilities are time-independent and remain constant over an observed infection period (such that the process is assumed to be time-homogeneous).
- Transitions or relapses between diseases, verified co-infections, or other medical problems were not considered or did not match the study's inclusion requirements.
- The Markov model assumed that no individual was immune to the COVID-19 infection, and thus, immune states were not captured.

3.3.7 Estimating Transition Probabilities

The transition probabilities for each disease, as well as their related standard errors, were estimated using maximum likelihood principle. Table 3.2 illustrates the number of people in each cohort during the study period in any state (S-I-D).

Table 3.2: Number of Individuals at Each State

Groups	Susceptible	Infected	Dead
Susceptible Individuals	X_{00}	X_{01}	X_{02}
Infected Individuals	X_{10}	X_{11}	X_{12}

The entries in Table 3.2 are as follows:

X_{00}:The number of individuals who stayed susceptible at the end of the study.

X_{01}: Number of susceptible individuals who became infected at the end of study period

X_{02}: Number of susceptible individuals who died at the end of the study period

X_{10}: Number of infected individuals who recovered at the end of the study period

X_{11}: Number of infected individuals who remained infected at the end of the study period

X_{12}: Number of infected individuals who died at the end of the study period

3.3.8 Distribution of transitions

A binomial distribution governs the number of individuals in each transition state. This is because, at the end of the study period, each state has a constant number of subjects, and each individual's disease outcome is independent of the other within each state. Also, with a constant probability P_{ij}, each person is subjected to two mutually exclusive outcomes (remaining in or leaving that state). As a result, the transition events are unrelated to one another (as defined by the Markov principle). The transition probability's likelihood, P_{ij}, follows a binomial distribution as follows:

$$L(P_{ij} \mid N, x) = \binom{N_i}{x_{ij}} P_{ij}^{x_{ij}} (1 - P_{ij})^{N_i - x_{ij}} \tag{3.16}$$

where N_{ij} is the number of observed transition that starts from state i to j and

$$\sum_j P_j = 1$$

3.3.9 Maximum Likelihood estimation

Maximum likelihood estimation (MLE) is a method for estimating the parameters of a statistical model given observations by determining the parameter values that maximize the Likelihood of observing the parameters. The transition probabilities' maximum likelihood estimate is as follows:

$$\hat{P}_{ij} = \frac{x_{ij}}{\sum_j x_{ij}} = \frac{x_{ij}}{N_j} \tag{3.17}$$

for $i, j = 0, 1$ with standard errors from the sampling distribution of the ML estimate given as;

$$s.e(P_{ij}) = \sqrt{\frac{\hat{p}_{ij}(1-\hat{p}_{ij})}{N_i}}$$

Anderson & Goodman (1957) shown that the estimator in (3.17) is a maximum-likelihood estimator of the transition probabilities, and that it is also consistent. According to some researchers, the MLE assumes a uniform prior distribution of the underlying parameters. Given a uniform prior distribution on the parameters, the ML estimator is known to correspond with the most probable Bayesian estimator.

Remarks

The log-likelihood function $l(P)$ based on the transition probabilities is given as

$$l(P) = \sum_{i=0}^{2}\sum_{j=0}^{2} n_{ij}\log(\hat{P}_{ij}),$$

and the Akaike information criterion (AIC) is given as

$$\text{AIC} = -2l(P) + 2k = -2\left[\sum_{i=0}^{2}\sum_{j=0}^{2} n_{ij}\log(\hat{P}_{ij})\right] + 2k$$

where k is the number of parameters estimated (in this study, $k = 6$). The proposed AIC is used to compare between the two fitted Markov Models at the two infection periods (2020 and 2021, given their estimated transition probabilities).

3.3.10 Estimating P^n Transition matrix

The notion of eigenvalues and eigenvectors, as established by Bhat (1984), is used to estimate the nth step transition probability matrices for each disease. If P is a $m \times m$ matrix, its eigenvalues are the numbers λ for which the characteristic equation is true.

$$PX = \lambda X \tag{3.18}$$

has a solution for $X \neq 0$. Column vector X is called the right eigenvector of the eigenvalue λ. Similarly, if $Y \neq 0$ exists, we can build a row vector Y that is defined as a left eigenvector belonging to the eigenvalue λ.

$$YP = \lambda Y \tag{3.19}$$

As a result, it's worth noting that the eigenvalues are found by solving the characteristic equation.

$$|\lambda I - P| = 0 \tag{3.20}$$

m linearly independent right eigenvectors and m linearly independent left eigenvectors pertaining to these unique eigenvalues can be produced when the m eigenvalues of matrix P are all distinct. Let Q be a non-singular matrix $(X_1, X_2, \cdots, X_n)$, with X_i representing the right eigenvalue $\lambda_i (i = 1, 2, \cdots, m)$. Assume Λ is a di-

agonal matrix with $\lambda_i (i = 1, 2, ..., m)$ eigenvalues. Then if

$$PX_j = \lambda_j X_j \tag{3.21}$$

gives

$$PQ = Q\Lambda \tag{3.22}$$

and

$$Q^{-1}PQ = \Lambda \tag{3.23}$$

where Λ is the diagonal matrix of the form,

$$\Lambda = \begin{bmatrix} \lambda_1 & \cdots & \cdots \\ \vdots & \lambda_2 & 0 \\ \vdots & 0 & \ddots \lambda_m \end{bmatrix}$$

The matrix Q^{-1} in 3.23 is a non-singular matrix made up of the m linearly independent left eigenvectors, $Y_i(i, 2, \cdots, m)$. The matrix P is said to be diagonalized if 3.23 holds. From this relationship, iteratively;

$$\begin{aligned} P &= Q^{-1}\Lambda Q \\ P^2 &= (Q^{-1}\Lambda Q)(Q^{-1}\Lambda Q) = Q\Lambda^2 Q^{-1} \\ &\vdots \end{aligned} \tag{3.24}$$

$$P^n = Q\Lambda^n Q^{-1} \tag{3.25}$$

where

$$\Lambda^n = \begin{bmatrix} \lambda_1^n & \cdots & \cdots \\ \vdots & \lambda_2^n & 0 \\ \vdots & 0 & \ddots \lambda_m^n \end{bmatrix}$$

Lemma 1 (Estimation of the P^n Transition Matrix). *Suppose that P is a $k \times k$ transition probability matrix of a discrete-time Markov chain model with entries*

$\{P_{ij}\}$ *for* $i, j = 0, 1, \cdots, k-1$ *such that*

$$\sum_{j=0}^{k-1} P_{ij} = 1 \quad for \quad i = 0, 1, \cdots, k-1. \tag{3.26}$$

Then, the nth step transition probability matrix (P^n) *is given as*

$$P^n = Q\lambda^n Q^{-1} \tag{3.27}$$

where $Q = \{x_{ij}\}$ *is a* $k \times k$ *non-singular matrix of eigenvectors corresponding to the* j*th eigenvalue* $(\lambda_j, j = 0, 1, \cdots, k-1)$ *of the transition matrix* P*, with the matrix*

$$\lambda^n = diag(\lambda_0^n, \lambda_1^n, \cdots, \lambda_{k-1}^n) \tag{3.28}$$

Corollary 1. *Given that the nth-step transition probability matrix is given as* $P^n = Q\lambda^n Q^{-1}$ *in accordance with Lemma 1, where* Q^{-1} *is the inverse of matrix* Q *and* $(\lambda_j, j = 0, 1, \cdots, k-1)$*. Suppose* Q *is a matrix of eigenvectors with entries* $\{x_{ij}\}$ *and its inverse matrix* (Q^{-1}) *has entries* $\{v_{ij}\}$ *for* $i, j = 0, 1, 2, \cdots, k-1$*, then, the* i*th row entry of the* P^n *transition matrix is given as;*

$$a_{ij} = \sum_{r=0}^{k-1} x_{ir} v_{rj} \lambda_j^n \quad for \quad i, j = 0, 1, 2, \cdots, k-1 \tag{3.29}$$

Remarks

The decomposition approach, which requires eigenvalues and their corresponding eigenvectors, is used to estimate the $P_{ij}^n, i, j = 0, 1, 2$ transition probability matrix. As a result, it may be calculated using the decomposition from 3.25 (by invoking Lemma 1 and Corollary 1); where Q is a 3×3 non-singular matrix and $X_j, (j = 0, 1, 2)$ are the right eigenvectors corresponding to the eigenvalues $\lambda j (j = 0, 1, 2)$.

As a result, $PX_j = \lambda_j X_j$ and

$$\Lambda^n = \begin{bmatrix} \lambda_0^n & 0 & 0 \\ 0 & \lambda_1^n & 0 \\ 0 & 0 & \lambda_2^n. \end{bmatrix}$$

3.4 Estimating Relevant Disease Metrics from COVID-19 S-I-D Markov Model

Disease metrics are epidemiological variables that are used to understand or predict disease pathogenicity and spread in both susceptible and infected individuals. As proposed by Zipkin et al. (2010) and also implemented by Twumasi (2018), the discrete-time Markov chain model is also employed in the current study to estimate these key disease metrics (defined below) for describing COVID-19 transmission in Ghana at the two observed infection periods i.e., 2020 and 2021.

Probability of First Infection and Recovery

For each of the two study population, the probability of first infection and probability first recovery would be estimated for COVID-19. Before these disease measures are estimated we state a Lemma as given in (Twumasi, 2018)

Lemma 2. *Let X_n be a Markov chain with state space $S = 0, 1, 2, \cdots, m-1$, then the probability of first transitioning from state i to state j in m-steps is given as*

$$f_{ij}^{(m)} = P(X_m = j, X_r \neq j, r = 1, 2, \cdots, m-1 \mid X_0 = i) \tag{3.30}$$

Following the line of reasoning as stated in Lemma 2, between the $m-1$ and m time steps, the probability of a susceptible individual becoming infected for the

first time would be expressed as:

$$\begin{aligned} f_{01}^{(m)} &= P(X_{m+m} = 1, X_{n+m-1} = 0, \cdots, X_{n+1} = 0 \mid X_n = 0) \\ &= P_{00}^{m-1} P_{01}. \end{aligned} \tag{3.31}$$

Estimating the probability that an infected individual will recover first between the $m-1$ and m time steps follows the same logic:

$$\begin{aligned} f_{10}^{(m)} &= P(X_{m+m} = 0, X_{n+m-1} = 1, \cdots, X_{n+1} = 1 \mid X_n = 1) \\ &= P_{11}^{m-1} P_{10}. \end{aligned} \tag{3.32}$$

The cumulative probabilities that a susceptible individual had become infected by a certain time step, m can be calculated using the cumulative sum of the first transition probabilities defined in equation (3.31) as:

$$T_{f01}^{(m)} = \sum_{t=1}^{m} f_{01}^{(t)} = \sum_{t=1}^{m} P_{00}^{t-1} P_{01} \tag{3.33}$$

Overall Probability of Infection and Recovery

As indicated by Zipkin et al. (2010), as m increases, the probability of first infection (or recovery) turns towards zero, implying that the overall probability to become infected (or recover) nears a limit value of 0% or 100%. This allows for the estimation of the overall probability of a susceptible individual becoming infected with COVID-19 (or an individual recovering from infection of COVID-19) for both study population, as well as the rate at which the process happens. The overall probability that an individual transitions from state i to state j has a simple closed form solution in our three-state model and could be explicitly expressed as;

$$P\{i \to j\} = \frac{p_{ij}}{1 - p_{ii}} \tag{3.34}$$

Expected Time to First Infection and Recovery

From equation (3.34), the estimated time to first infection, that is the average time for a suspectible individual to become infected and the estimated time to recovery which is the average time taken for an infected individual to recover. The expected time for an indvidual in state i to first enter state j, at time n represented as $\mathrm{E}[\tau_{ij}^{(1)}]$ could be calculated as;

$$\mathrm{E}[\tau_{ij}^{(1)}] = \frac{\sum m f_{ij}^{(m)}}{P\{i \to j\}}. \tag{3.35}$$

where the numerator is the sum of the first passage time, the denominator is equation (3.34), and the expected time which explains the expected time an individual would leave to an infection state or recovery state could be written in a closed form as:

$$\mathrm{E}[\tau_{ij}^{(1)}] = \frac{1}{1 - p_{ii}}. \tag{3.36}$$

Life Expectancies for Healthy and Infected Individuals

Markov chain models are also effective for predicting the life expectancies of infected and vulnerable or healthy individuals. We define life expectancy in the framework of Markov chains as the predicted time to death for an individual from the present time period n. We find the life expectancies of the healthy individual and infected individuals as proposed by Zipkin et al. (2010). The life expectancy

for individuals in the susceptible (0) and infected (1) is expressed as:

$$W = (I - Q)^{-1} \times \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{3.37}$$

where $W_i(i = 0, 1)$ statistically means an estimated time of absorption into the dead (2) state, I is a 2×2 identity matrix and Q which is the matrix of transition probabilities for the transient states, infection and recovery is defined as:

$$\begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} \tag{3.38}$$

and Q^{-1} means the inverse of Q It is clear that, W is the total number of time periods expected in each transient state given the previous state.

3.5 Time Series Modelling of COVID-19

A time series is defined as a series of observations, X at corresponding time, t. The parameter t can be discrete or continuous (Box et al., 2015). To summarize, the process $X = X_t, t \in T$ is stochastic, where T denotes the time space. The goal of time series analysis is to use historical observations (X_t) at a given time t to forecast future observations at time $t + k$, where k is defined as the lead time. Forecasting is the statistical process of predicting future events based on past values (Takele, 2020). There are various classical time-series models, but the more robust model is the ARIMA model (Takele, 2020; Twumasi & Twumasi, 2021). The ARIMA model is considered a baseline model for comparing the five selected machine learning time-series models via a rolling-origin evaluation technique. All models were fitted in R statistical software version 3.6.3 (R Core Team, 2022).

3.6 The ARIMA Model

Given a time series process X_t at time t, the generalized homogeneous non-stationary Autoregressive Integrated Moving Average (ARIMA) (p, d, q) model can be written as follows:

$$\Phi(L)x_t = \phi(L)\Delta^d x_t = \theta_0 + \theta(L)\xi_t, \tag{3.39}$$

where $\phi(L)$ is the pth order AR characteristics polynomial, $\theta(L)$ is the qth order MA characteristics polynomial. d denotes the degree or order of differencing that is required until trends or fluctuations are removed – that is until stationarity is achieved (Shumway et al., 2000). And $\Delta = 1 - L$ is the difference operator, and ξ_t is the white noise which which is Guassian. i.e $\xi_t \sim N(0, \sigma^2)$.

3.7 Machine Learning Time Series Model

According to Martínez et al. (2019), even though the ARIMA model has been hitherto used to forecast time series, in the last ten years, the use of machine learning models has been adopted due to their non-parametric nature. In this section, the theoretical frameworks of ML time series models are laid out.

3.7.1 The K-Nearest Neighbour regression

The K-Nearest Neighbour (KNN) regression model is a structural analogy-based learning model that collects the n features of the training data. Every training point belongs to the n-dimensional space, and given a new point, the KNN finds its k nearest point in the n-dimensional space using similarity metrics (see equation (3.41) to predict the values of the unknown target. If we have n training

points $x_i(n = 1, 2, \cdots, n)$ with a corresponding target vector $g_i(i = 1, 2, \cdots, m)$ closest to the new point $p_i(i = 1, 2, \cdots, n)$, the new target $(\hat{y})$ is predicted by majority voting (see equation (3.40)) of the k neighbours or similar points. That is:

$$\hat{y} = \sum_{i=1}^{k} \frac{g_i}{k}, \tag{3.40}$$

and the Euclidean distance, d used as the similarity metric is (Hastie et al., 2001; Martínez et al., 2019):

$$d_i(x, p) = \|(x_i - p_i)\| = \sqrt{\sum_{i=1}^{k} (x_i - p_i)^2}. \tag{3.41}$$

It should be noted, as Ahmed et al. (2010) suggests, that because k is an important hyperparameter of the KNN, care should be taken when selecting it. And when the value of k increases, variance decreases, leading to a suitable fit, but at the expense of an increased bias.

3.7.2 Generalized Regression Neural Network

Generalized Regression Neural Network (GRNN) Model pointed out by Ahmed et al. (2010) was proposed by Nadaraya (1964) and Watson (1964) and hence known as Nadaraya–Watson Kernel Regression, but identified in the field of machine learning as GRNN. Just like most machine learning methods, the GRNN is a non-parametric model where the output of the target of the training points in topological vicinity of the given point x is averaged in order to predict the output of the given point x . A defined kernel function is employed in averaging the outputs of the target of the training data based on their distance from x, hence the prediction made is simply a weighted sum of the target outputs. Now, assuming we have training data points $x_i(i = 1, 2, \cdots, n)$ and an associated target outputs

$y_i (i = 1, 2, \cdots, n)$, then the estimated output $(\hat{y})$ of the given data input x is,

$$\hat{y} = \frac{\sum_{i=1}^{n} y_i \text{kdf}(x, x_i)}{\sum_{i=1}^{n} \text{kdf}(x, x_i)} \tag{3.42}$$

where $\text{kdf}(x, x_i) = e^{\frac{-d_i}{2\sigma^2}}$ which is a similarity measure is the radial basis function(RBF), particularly the Guassian kernel with

$$d_i = \|x - x_i\|^2 = (x - x_i)^T (x - x_i). \tag{3.43}$$

That is d_i is the square of the Euclidean distance. The predicted value, $\hat{y}$, can be thought of as the exponentially weighted average of all the target outputs, y_i, as determined by the Euclidean distance (Specht et al., 1991; Rooki, 2016). σ is called the smoothing parameter. And, as Specht et al. (1991) pointed out, as σ approaches infinity, the $\hat{y}$ approaches the observed values, y_i, and as σ approaches zero, the $\hat{y}$ becomes y_i, which corresponds to the observation closest to x. If σ is between zero and infinity, all the observed values are considered yet the closest to the x assumed much weight. One of the key advantages of the GRNN is that its algorithm can be used for both linear and non-linear data, and also that the GRNN is a one-pass learning which easily learns from examples to predict (Specht et al., 1991; Twumasi & Twumasi, 2021).

3.7.3 Non-Seasonal Neural Network Auto-Regression Model

The Neural Network Auto-Regression (NNAR) model is a feed-forward (having no loops or cycles) Artificial Neural Network (ANN). The NNAR has three layers; the input layer, the hidden layer, and the output layer. In this study, a single hidden layer with different nodes k is adopted. The lags p are used as inputs in an autoregressive neural network model (Faraway & Chatfield, 1998; Hyndman & Athanasopoulos, 2018). The NNAR is structurally designed in such a way that, there is a connection between each input and the nodes and then a link

between the nodes and the output. That is, to obtain the output, first a weighted linear sum (as in equation (3.44)) is computed for each input where the weight, w_{ik} measures the link between the input y_i and the kth node. The weighted linear sum is then transmuted through an activation function in which the logistic sigmoid function, ψ_h is normally used. This transformation gives an input for the next layer. The use of the activation function tends to reduce the negative impact of outliers in the data (Hyndman & Athanasopoulos, 2018). As Faraway & Chatfield (1998) suggests, to predict the output, a transformation for the z_ks and the constant term can be performed. However, the logistic sigmoid function is normally not used for the output (Moshiri & Cameron, 2000; Faraway & Chatfield, 1998), unless the data is appropriately scaled to fall between 0 and 1, but rather the identity activation function could be used at the output layer. The layout of the NNAR model is presented in these three steps (i.e equations (3.44), (3.45), and (3.46)): First the weighted linear sum (z_k) connecting the inputs with the hidden nodes;

$$z_k = \left(w_{ch} + \sum_{i=1}^{p} w_{ih} y_{t-i}\right), \tag{3.44}$$

which is then transformed using the sigmoid function, ψ_h

$$\psi_h(z_k) = \frac{1}{1 + e^{-z_k}} \tag{3.45}$$

and with ψ_0 being the activation function that links the nodes to the output, the general formula for NNAR (p, k) for predicting the values of y_t using historical observations or inputs; $y_{t-1}, y_{t-2}, \cdots, y_{t-p}$ is:

$$\hat{y}_t = \phi_0\left[w_{c0} + \sum_{h=1}^{k} w_{ho}\psi_h\left(w_{ch} + \sum_{i=1}^{p} w_{ih} y_{t-i}\right)\right] \tag{3.46}$$

Where, w_{ch} represents the weight of the link between the constant input and the hidden nodes, w_{co} represents the weight of the link between the constant input and the output. Also w_{ih} is the weight of the link between the input and the

hidden nodes while w_{ho} represents the weight of the link between the hidden nodes and the output.

3.7.4 Multilayer Perceptron Model

The Multilayer Perception (MLP) is also part of the family of feedforward artificial neural network. The most common among the neural networks, MLP was developed initially to deal with sophisticated classification problems but was later extended to regression problems and time series forecasting due to its powerful feature of approximating any measurable function irrespective of the activation function. (Hornik et al., 1989; Bishop et al., 1995). Just like its counterpart in NNAR, the MLP consists of three phases; input, hidden, and output layers or more. Hence the predicted output can be presented as:

$$\hat{y}_t = \phi_0\Big[w_{c0} + \sum_{h=1}^{k} w_{ho}\psi_h\Big(w_{ch} + \sum_{i=1}^{n} w_{ih}y_t\Big)\Big] \tag{3.47}$$

for inputs, y_i with n entries, that is $i = 1, 2, \cdots, n$. All other notation preserve their explanation as in 3.46 and the logistic sigmoid function as still used as described in 3.45. It could be observed that, the difference between the MLP and the NNAR, is that is the autoregression aspect of the NNAR which is absent in the MLP. One shortcoming of the MLP is that it can take a lot of time to run ((Ahmed et al., 2010; Twumasi & Twumasi, 2021) which could be the cost for its efficiency (Ahmed et al., 2010)

3.7.5 Extreme Learning Machine

Extreme Learning Machine (ELM) was developed through inspiration from the network of neurons in biology. The ELM, unlike other neural networks, does not require the tuning of the weight of the hidden layer (Atiquzzaman & Kandasamy,

2016). And the ELM does not use the gradient approach in training the network but instead uses the Moore-Penrose inverse method (G.-B. Huang et al., 2011; Singh & Balasundaram, 2007). Let $x_i(i = 1, 2, \cdots, n)$ represent the input vector and $y_i(i = 1, 2, \cdots, k)$ represent the associated output vector. Then, the ELM could be expressed explicitly as:

$$\hat{y}_j = \sum_{i=1}^{H} \alpha_i F(\gamma_i, b_i, x) = \sum_{i=1}^{H} \alpha_i F(\gamma_i x_j + b_i), \qquad \forall j = 1, 2, \cdots, N \tag{3.48}$$

where N is the number of entries or input data values while H is the number of hidden units. γ_i is the vector of the weights that link the input layer to the hidden layer and b_i is bias of the hidden layer, both of which are randomly created. The α_i are the values of the weight linking the hidden units and the output layer. $F(.)$ is the squashing function used to transmute the additive hidden units into outputs. The ELM is noted for learning faster than the traditional neural networks, does not use iteration to learn, increase in performance, automatic analytic determination of the parameters in the network, among others. For other advantages of the ELM, see works by Atiquzzaman & Kandasamy (2016).

3.8 Preprocessing Methods and Evaluation Techniques

The daily time-series data on infected COVID-19 cases was cleaned (by checking and correcting for outliers and missing values) using the *tsclean()* function in R. In this study, the Webel and Ollech (WO) overall seasonality test (which combines two seasonality tests: QS-test and kwman-test) was used to determine whether or not the time series data had seasonality. The WO-test concluded that there is no seasonality in the daily time-series data (p-value = 0.832). Regarding the test of non-stationarity in the daily series, the Augmented Dickey-Fuller (ADF) unit root

test was employed. It was revealed that the (corrected) daily cases of COVID-19 in Ghana over the study period were stationary (p-value ≤ 0.01). The corrected time-series data was then split based on the out-of-sample rolling approach. The separation of the historical data series into fit periods (training data) and test periods (testing data) is the first step in an out-of-sample evaluation strategy (Tashman, 2000).

Assume h is the maximum forecasting horizon, chosen in this study to be 120 days, and m is the minimum forecasting horizon, chosen to be 90 days in this study. And if T is the total length of the observed series, then the training data for fitting a time series model is generated with increasing lengths of $T-h, T-h-1, \cdots, T-m$ for forecasting, while the testing data for cross-validation is generated with decreasing lead times of $h, h-1, \cdots, m$. And the estimated number of various sets of forecasts or the number of times an individual model is re-calibrated under this rolling-origin strategy is $N = h - m + 1$ (in this case, 120-90+1=31 forecasting equations). As a result, for this rolling-origin method with a maximum forecasting horizon of h, the total number of the forecasts is given by $\frac{N(h+m)}{2}$ (Twumasi & Twumasi, 2021).

For each lead time, the rolling out-of-sample evaluation generates forecast. As a result, we can evaluate an individual time series' forecasting performance at each lead time considering some summary statistics of the forecast errors. Performance measures in terms of Median Root Mean Squared Error (MdRMSE) as in equation (3.50) was computed from Root Mean Squared Error (RMSE) as in equation (3.49) to measure the accuracy of the competing time series models. The distributions of the forecast errors for each model are also compared using the Kruskal-Wallis test of difference in forecast errors and the Bonferroni Dunn's test of multiple comparisons at a 5% significant level.

$$\text{RMSE} = \sum_{i=1}^{N_t} \frac{(\hat{y}_i - y_i)^2}{N_t} \tag{3.49}$$

and the

$$\text{MdRMSE} = \text{Median}\{\text{RMSE}_2, \text{RMSE}_1, \cdots, \text{RMSE}_{N_t}\} \tag{3.50}$$

$\hat{y}_t$ and y_t denote the forecasted values and the observed values at time t respectively. N_t is the interval of the fit period at lead times t. The lower the MdRMSE, the better the fit. The use of MdRMSE is suitable due to its robustness to potential outliers in the error distribution across the fitted time-series models (Hyndman & Koehler, 2006).

Chapter 4

Analysis and Discussion

This chapter presents the results obtained from the fitted discrete-time Susceptible-Infected-Disease (S-I-D) Markov model and the time-series models (the five ML algorithms against the baseline ARIMA model). First of all, the transition counts among the suspectible, infected, and death states for each study period are provided. After that, the transition probabilities for the respective years are also presented based on which various classes of states are discussed. Furthermore, other disease metrics are computed and explained in relation to COVID-19 infection. Finally, the time series models were discussed.

4.1 S-I-D Discrete-time Markov Chain Modelling

To estimate the disease metrics of COVID-19 transmission in Ghana based on the aggregated COVID-19 data at the two infection periods (2020 and 2021), and then predict COVID-19 cases using the estimated generalized P^n transition matrix, the three-state discrete-time S-I-D Markov model (defined in section 3.3.6) was adopted.

4.1.1 Transition Counts

The transition counts, which tell the number of people in a particular state, are shown for the years 2020 and 2021 in Table 4.1 and Table 4.2, respectively.

Table 4.1: Number of Persons at each State of COVID-19 in 2020

States	Susceptible	Infected	Death	Total
Susceptible	30813769	54711	204460	31072940
Infected	53594	1117	335	55046

Table 4.2: Number of Persons at each State of COVID-19 in 2021

States	Susceptible	Infected	Death	Total
Susceptible	31402777	129440	199912	31732129
Infected	125839	3601	1167	130607

4.1.2 Estimation of Transition Probabilities

The estimated transition probabilities as obtained from the MLE method from 3.17 with their standard errors and their confidence intervals are presented in Table 4.3 and Table 4.4 for COVID-19 in infection periods 2020 and 2021, respectively.

Table 4.3: Estimates of the Transition Probabilities for COVID-19 2020 data

Parameters	Estimates ($\hat{p}_{ij}$)	Standard Error	99%Confidence Interval
P_{00}	0.992	1.63151×10^{-5}	[0.99162 , 0.99170]
P_{01}	0.002	7.52094×10^{-5}	[0.00174 , 0.00178]
P_{02}	0.007	1.45040×10^{-5}	[0.00654 , 0.00662]
P_{10}	0.974	6.83050×10^{-4}	[0.97186 , 0.97538]
P_{11}	0.020	$6.00965 \times 10^{-}4$	[0.01874 , 0.02184]
P_{12}	0.006	3.31491×10^{-4}	[0.00523 , 0.00694]

Table 4.4: Estimates of the Transition Probabilities for COVID-19 2021 data

Parameters	**Estimates** $(\hat{p}_{ij})$	**Standard Error**	**99%Confidence Interval**
P_{00}	0.990	1.79914×10^{-5}	[0.98957 , 0.98967]
P_{01}	0.004	1.13148×10^{-5}	[0.00405 , 0.004101]
P_{02}	0.006	1.40459×10^{-5}	[0.00626 , 0.00634]
P_{10}	0.963	5.18951×10^{-4}	[0.96216 , 0.96483]
P_{11}	0.028	$4.53079 \times 10^{-}4$	[0.02640 , 0.02874]
P_{12}	0.009	2.60387×10^{-4}	[0.00826 , 0.00961]

The estimated transition probability matrices at the two infection periods (2020 and 2021) are denoted by $\hat{\mathbf{P}}_{2020}$, and $\hat{\mathbf{P}}_{2021}$, respectively. From the estimated transition probabilities, it can be observed that 4 out of every 1000 susceptible individuals tested positive for COVID-19 in 2021 ($\hat{P}_{01} = 0.004$), with 9 out of every 1000 infected individuals dying from COVID-19 in the same year ($\hat{P}_{12} = 0.009$). These rates were relatively higher in 2021 than in 2020 ($\hat{P}_{01} = 0.002$), with the estimated recovery rate ($\hat{P}_{10}$) comparatively lower in 2021. The transition probability matrices at the two infection periods are thus given as

$$\hat{\mathbf{P}}_{2020} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{pmatrix} 0.992 & 0.002 & 0.007 \\ 0.974 & 0.020 & 0.006 \\ 0 & 0 & 1 \end{pmatrix} \end{array} \quad \text{and} \quad \hat{\mathbf{P}}_{2021} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{pmatrix} 0.990 & 0.004 & 0.006 \\ 0.963 & 0.028 & 0.009 \\ 0 & 0 & 1 \end{pmatrix} \end{array}$$

4.1.3 Classification of model states using directed multi-graph

The directed multi-graphs in Figure 4.1 are diagrammatic representation of the transition probabilities. The vertices of the graphs indicate the states, while the directed arcs represent the non-zero probabilities of one-step transitions. Figure 4.1 also suggests that the susceptible (S) state and infected (I) state are transient class (since states S and I communicate with the same equivalence class, but upon leaving their class, there is a zero probability of the process returning to this class);

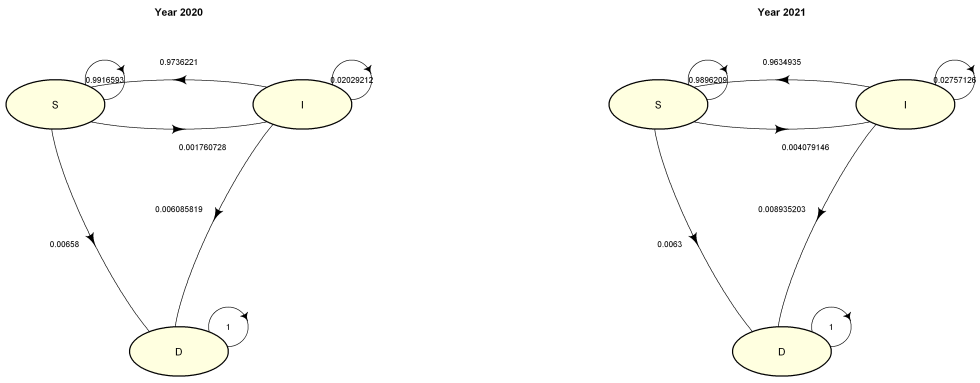


Figure 4.1: Directed multigraph at the two infection periods (2020 and 2021, respectively)

whereas the dead state (D) is an absorbing state (such that an individual cannot be susceptible or infected after death). Also, Figure 4.1 indicates that the Markov chains at the two infection periods are not irreducible since the three states do not mutually communicate or belong to the same equivalence class (though the process is aperiodic but not ergodic). Hence, there is no steady-state distribution but the n-th transition matrix exist at both infection periods (for the infection periods 2020 and 2021, respectively).

4.2 Estimating other disease metrics for COVID-19 in Ghana During Year 2020 and 2021

In this section, relevant disease metrics such as the probability of first infection and recovery, the overall probability of infection, expected time to infection and recovery, and the life expectancy for infected and susceptible individuals are estimated for COVID-19 transmission with the entire Ghana population during the infection periods 2020 and 2021, respectively.

4.2.1 Probability of First Infection and First Recovery

The probability of a susceptible individual first becoming infected and an infected individual first recovering were estimated for COVID-19 during the year 2020 and 2021 over 50 time steps. These are illustrated in Figure 4.2.

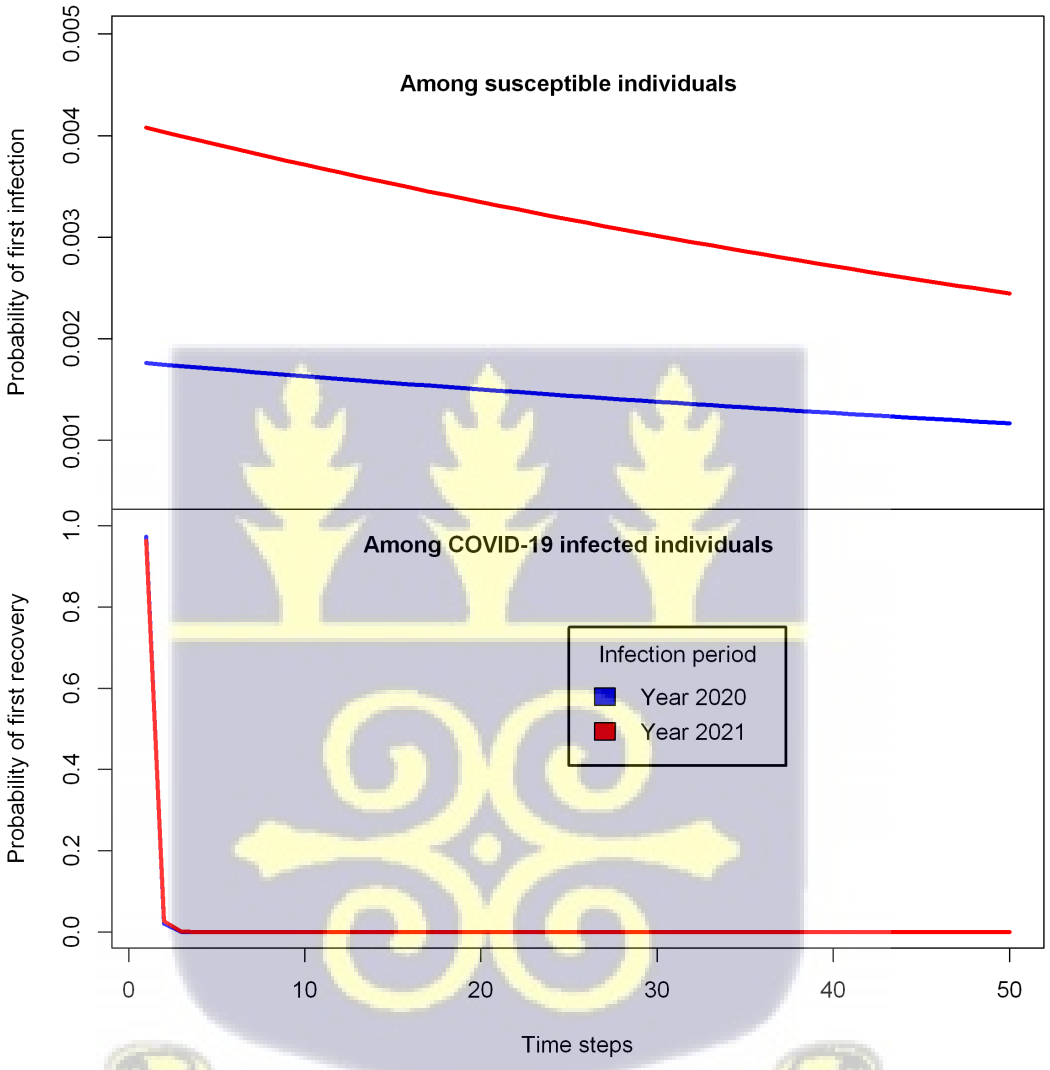


Figure 4.2: Probability of First Infection and First Recovery

Interpretation of Figure 4.2

Figure 4.2 shows the probability of first infection (top of the panel) and the probability of first recovery (bottom of the panel) for COVID-19 for 2020 and 2021. It can be seen that the likelihood of first infection among susceptible individuals in 2020 was comparatively lower in Ghana (since the disease was declared a pandemic by WHO in 2020). The stringent preventive measures and public health interventions (such as compulsory use of facemasks, lock-down rules, social distancing, personal hygiene measures, tight policies for new arrivals at the Ghana Kotoka International airport, and the closing of borders) by the government of Ghana in 2020 compared to 2021, according to Kenu et al. (2020), may explain these variations .

Additionally, more deadly variants of the coronavirus developed in 2021 may explain the high rate of the first infection in 2021 compared to 2020. Generally, the probability of the first infection of the COVID-19 infection decreases for both infection periods (2020 and 2021) as the number of years or time steps increases (hopefully due to several factors, including but not limited to COVID-19 induced immunity and more informed decisions over time). The probability of the first infection of COVID-19 in Ghana (at both infection periods) was below 0.005 (i.e., < 5 out of every 1000 susceptible individuals had a chance of contracting the coronavirus in Ghana). Nonetheless, the probability of first recovery (estimated from equations (3.31) and (3.32) was consistent across both infection periods, with a sharp decrease towards 0 (as the time steps increased).

4.2.2 Probability of Infection at each observed period

The cumulative probability of infection (given by Figure 4.3), which quantifies the probability of COVID-19 infection within the population is estimated

as the cumulative sum of the probability of first infection (as defined in equation (refeq277)).

Interpretation of Figure 4.3

Figure 4.3 confirms that the national COVID-19 transmission in Ghana was indeed higher in the 2021 (based on the 2021 aggregated data) compared to the year 2020 (notwithstanding that the first two confirmed cases were reported on March 12, 2020; see work by Kenu et al. (2020)). Enhanced surveillance in the form of active case search and contact tracing procedures were launched during the lockdown in 2020 to find, isolate, and treat all confirmed cases early. With funding from the Ghana Field Epidemiology and Laboratory Training Program (FELTP), the Ghana Health Service mapped out all existing cases in 2020, conducted a risk assessment, and sampled members of households within a 1–2 km radius of the cluster of cases based on the setting (Kenu et al., 2020).

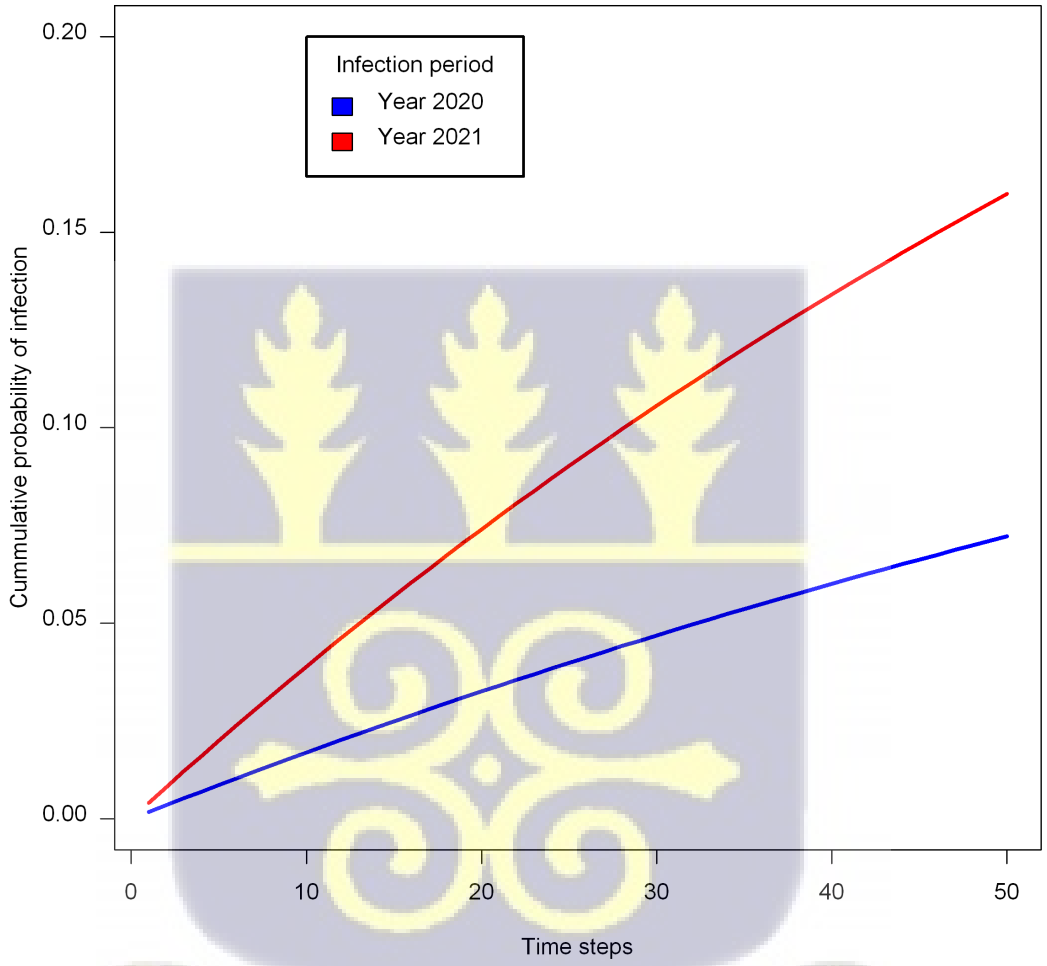


Figure 4.3: Cummulative Probability of Infection

4.2.3 Other estimated disease metrics of COVID-19

The other relevant disease metrics (such as the overall probability of infection and recovery, expected time to infection and recovery, and life expectancies based on the national aggregated data at the two infection periods (under investigation) were also estimated (presented in Table 4.5).

Overall Probability of Infection and Recovery

Table 4.5 also confirms that the overall probability of COVID-19 transmission in Ghana was approximately 18.2% higher in 2021 compared to 2020. This finding agrees with the first infection probability and the cumulative probability of infection. Moreover, it was revealed that the overall probability of recovery in 2020 (99.4%) and 2021 (99.1%) was not significantly different (with only a 0.3% difference). These overall recovery probabilities (obtained for 2020 and 2021) imply that almost 99% of COVID-19 infected individuals in Ghana are highly likely to recover from the infection. According to preliminary projections by scientists and researchers, the overall COVID-19 recovery rate will be between 97 and 99.75 percent (Nazario, 2021). Nazario (2021) discoveries also validate the fitted S-I-D Markov model since the estimates of the overall recovery probability for the two infection periods (2020 and 2021) lie within the published range (97%-99.75%). A major reason for this finding in Ghana could be attributed to an adaptive immune response from the previous history of other outbreaks in Ghana, the favourable climatic conditions against this airborne infection, and the public health measures put in place (amidst the economic crisis the pandemic has had on every country, especially the developing and under-developed nations).

Expected Time to Infection and Recovery

In this study, the expected time to infection within the study population is the average infection duration of the COVID-19 outbreak. Thus, the high value of the expected time to infection indicates that it will take a very long time, almost 120 years (considering the aggregated data in 2020) or at least 96 years (considering the aggregated data in 2021), before the COVID-19 outbreak will die out or extinct from the population (see Table 4.5). This confirms findings from researchers who believe that COVID-19 is here to stay like the regular flu, and thus will develop into an endemic disease irrespective of healthcare interventions to control its transmission or protect susceptible individuals (Phillips, 2021). COVID-19 will be in circulation globally in some areas in the future years. The relatively low average infection duration in 2021 compared to 2020 (although the overall probability of infection was higher in 2021) suggests a possibility of either acquired immune response after a high infection rate within the population or other public health interventions to control the infection is having a positive effect. The latter possible reason may also suggest that some of the infected individuals in the previous year are now immune to the coronavirus. In other words, the lower expected time to infection with the 2021 study population could be due to the herd vaccination policy adopted by the country, the construction of more quarantine centres, and the free distribution of personal protective equipment (PPE) in the country (in 2021). Alternatively, it could also be because other individuals who were infected in the previous population (i.e., in 2020 but remained alive) have acquired immunity to the virus.

Life Expectancy for Healthy and Infected Individuals

In addition, Table 4.5 indicates that the average time for the individuals in the 2021 study population to survive with or without the infection is higher than

that of the individuals in the 2020 infection period (based on the estimated life expectancy for the entire healthy or infected population). The life expectancy estimates for the population are high even though the COVID-19 infection is expected to take a long time to extinct from the population. This also confirms that with a high overall recovery rate, a low infection rate, and a more extended infection period, there is a possibility of herd immunity or adaptive immunocompetence within the study population (with the help of ongoing routine vaccination programs such as the booster jabs). It thus explains why the life expectancy of infected individuals are slightly greater than susceptible or healthy individuals (irrespective of the infection period).

Table 4.5: Other Relevant Diseases Metrics Estimates

Metrics	**Estimates**	
	2020	**2021**
Overall probability of infection	0.211	0.393
Overall probability of recovery	0.994	0.991
Expected time to infection (years)	119.89	96.35
Expected time to recovery (years)	1.02	1.03
Life expectancy for healthy individuals(years)	151.99	158.45
Life expectancy for infected individuals(years)	152.07	158.03

4.2.4 Comparison of the Two Markov Models

The Akaike Information Criterion (AIC) described in section 3.3.9 was adopted for comparing the two fitted Markov models at the two infection periods (based on the maximum likelihood estimate of the transition probability matrix, respectively). Here, the motivation was to determine which aggregated data (up to 2020 or 2021) produced more reliable parameter estimates of the underlying disease metrics. The results of the AIC, as seen in Table 4.6, show that the Markov model fitted the 2020 COVID-19 aggregated data better than the 2021 COVID-19 data since it yielded the least information loss, comparatively (i.e., smaller AIC). However, modelling these two aggregated datasets helped us discover new findings about the COVID-19 disease outbreak in Ghana from 2020 to 2021 (by employing the

discrete-time Markov model).

Table 4.6: Results from model comparison

Markov Model	AIC
2020	3279433
2021	4151838

4.3 The Estimated P^n Transition Probability Matrix

The *nth* step transition probability matrix (defined by equation (3.25)) was also estimate for $n \geq 1$ at the two infection periods based on the proposed Lemma 1 and Corollary 1. However, the limiting distribution of the Markov process does not exist since the process is aperiodic but not ergodic (and thus, not an irreducible chain). Sections 4.4.1 and 4.4.2 present the estimated P^n transition matrix for COVID-19 at the two infection periods using equation (3.29).

4.3.1 P^n Transition Matrix for COVID-19 for the 2020 study population

Given the estimated transition matrix for COVID-19 transmission at the 2020 infection period, such that

$$\hat{\mathbf{P}}_{2020} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{pmatrix} 0.992 & 0.002 & 0.007 \\ 0.974 & 0.020 & 0.006 \\ 0 & 0 & 1 \end{pmatrix} \end{array},$$

the eigenvalues of the transition matrix $\hat{\mathbf{P}}_{2020}$ are $\lambda_0 = 1$, $\lambda_1 = 0.993$, and $\lambda_2 = 0.019$; with corresponding matrix of eigenvectors given as

$$\mathbf{Q} = \begin{pmatrix} 0.5773503 & 0.7069276 & -0.001809345 \\ 0.5773503 & 0.7072859 & 0.999998363 \\ 0.5773503 & 0 & 0 \end{pmatrix}$$

and its inverse,

$$\mathbf{Q^{-1}} = \begin{pmatrix} 0 & 0 & 1.7320507 \\ 1.4120159 & 0.0025548 & -1.4145707 \\ -0.9987006 & 0.9981946 & 0.0005059 \end{pmatrix}.$$

From equation (3.29) under Corollary 1, the resulting estimated transition matrix for the COVID-19 transmission in the year 2020 in Ghana was

$$\hat{P}^n_{2020} = \begin{pmatrix} 0.998(0.993)^n + 0.002(0.019)^n & 0.002(0.993)^n - 0.002(0.019)^n & 1 - 0.993^n \\ 0.999(0.993)^n - 0.999(0.019)^n & 0.002(0.993)^n + 0.998(0.019)^n & 1 - 1.001(0.993)^n + 0.001(0.019)^n \\ 0 & 0 & 1 \end{pmatrix}.$$

To test the accuracy of the estimated P^n transition probability matrix, setting $n = 1$ gives its transition matrix (in 3 d.p) approximately at first time step (for instance).

4.3.2 P^n Transition Matrix for COVID-19 for the 2021 study population

Similarly, given the estimated transition probability matrix for COVID-19 at the 2021 infection period, such that

$$\hat{\mathbf{P}}_{2021} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{pmatrix} 0.990 & 0.004 & 0.006 \\ 0.963 & 0.028 & 0.009 \\ 0 & 0 & 1 \end{pmatrix} \end{array},$$

the eigenvalues the transition matrix $\hat{\mathbf{P}}_{2021}$ are $\lambda_0 = 1$, $\lambda_1 = 0.994$, and $\lambda_2 = 0.024$; with corresponding matrix of eigenvectors given as

$$\mathbf{Q} = \begin{pmatrix} 0.5773503 & 0.7080677 & -0.004222167 \\ 0.5773503 & 0.7061445 & 0.999991087 \\ 0.5773503 & 0 & 0 \end{pmatrix},$$

and its inverse,

$$\mathbf{Q^{-1}} = \begin{pmatrix} 0 & 0 & 1.7320507 \\ 1.4063725 & 0.0059380 & -1.4123105 \\ -0.9931110 & 0.9958158 & -0.0027048 \end{pmatrix}.$$

Also, from equation (3.29) under Corollary 1, the resulting estimated transition matrix for the COVID-19 transmission in the year 2021 in Ghana was

$$\hat{P}^n_{2021} = \begin{pmatrix} 0.996(0.994)^n + 0.004(0.024)^n & 0.004(0.994)^n - 0.004(0.024)^n & 1 - 1(0.994)^n \\ 0.993(0.994)^n - 0.993(0.024)^n & 0.004(0.994)^n + 0.996(0.024)^n & 1 - 0.997(0.994)^n - 0.003(0.024)^n \\ 0 & 0 & 1 \end{pmatrix}.$$

Also, setting $n = 1$ gives its transition matrix (in 3 d.p) approximately at first time step.

4.4 Predictions based on the estimated $\hat{P}^n$ Transition Matrix

The nth step transition matrix estimated for the two infection periods for COVID-19 transmission dynamics in Ghana was further used to predict the proportion of infected individuals who will become susceptible, remain infected or die from COVID-19 infection over time (shown in Figure 4.4).

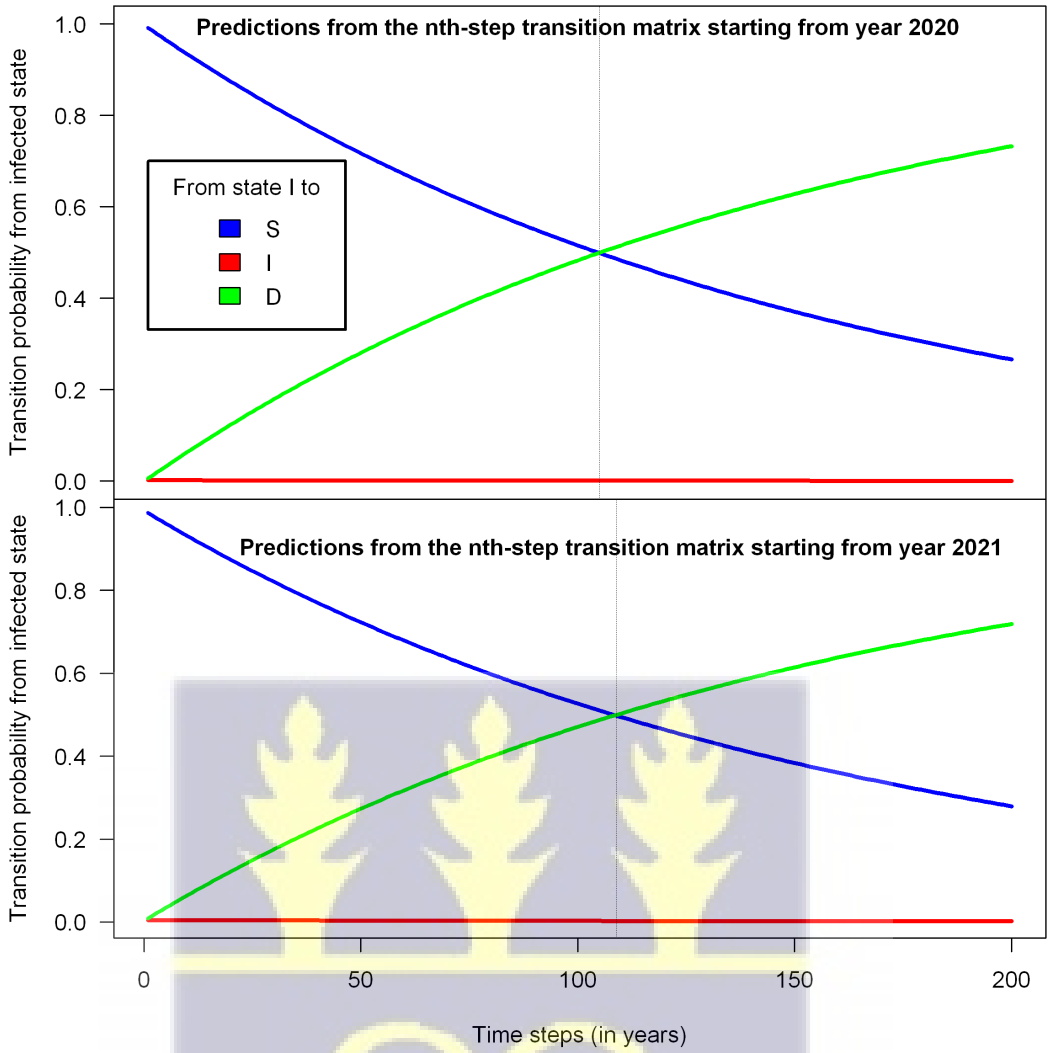


Figure 4.4: Transition Probability from Infected State over time for $1 \leq n \leq 200$

Interpretation of Figure 4.4

From Figure 4.4, it can be observed that the predicted probability of recovery amongst infected individuals decreases as the time step increases for both the 2020 and 2021 study populations. However, a reverse pattern was found for the rate of induced COVID-19 related deaths over time (as the recovery rate decreases); whereas the chance of remaining infected within the population (for both infection periods) was very minimal over time (towards zero). Hence, the lower the recovery rate becomes, the higher the chance of mortality among infected individuals, with a significantly smaller number of infected individuals constantly

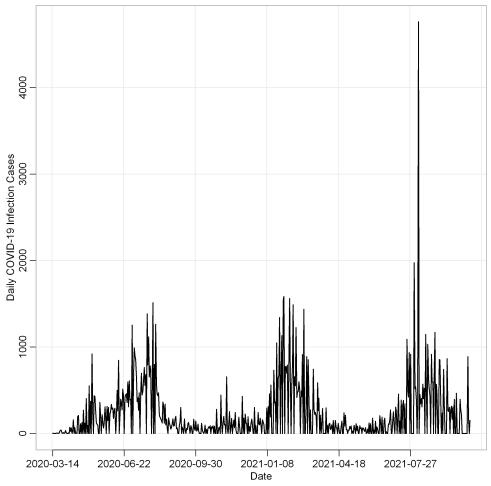
remaining infected at any given time point. As asserted by Phillips (2021), when the COVID-19 infection rate remains invariant over a long period, the coronavirus infection will yield endemicity (which could account for the quadratic growth in the number of COVID-19 related deaths in Ghana over time).

4.5 Time series Forecasting Using Machine Learning Algorithms and ARIMA

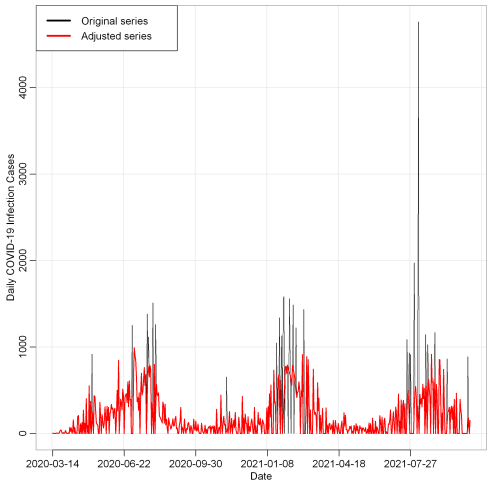
The K-Nearest Neighbor Regression (KNN), Neural Network Auto-Regressive (NNAR), Generalized Regression Neural Network (GRNN), Multi-Layer Perceptron (MLP) neural networks, and Extreme Learning Machines (ELM) were used to predict the daily cases of COVID-19 infection (with a 120-day one-step-ahead forecast via the rolling-origin evaluation strategy). The traditional ARIMA model was considered a baseline model for comparison.

4.6 Preprocessing and Preliminary Analysis of Time Series Data

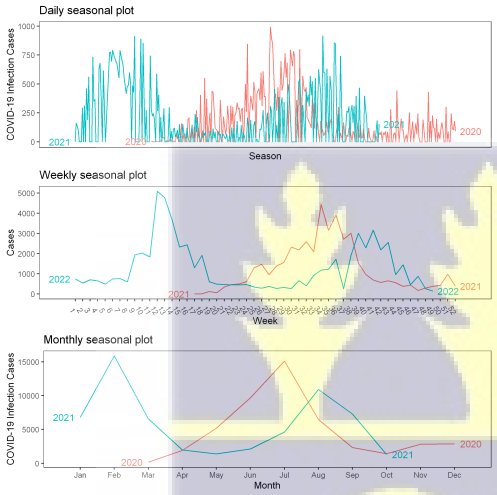
The daily time-series data (from March 14, 2020, to October 18, 2021) was preprocessed and corrected or adjusted for outliers and any potential missing values (see Figures 4.5 a and 4.5 b). The Webel-Ollech (WO) test identified only a weekly seasonality (p-value=0.001) in the adjusted daily series. Nonetheless, there was no daily (p-value=0.832) or monthly (p-value=0.755) seasonality (see Figure 4.5 c). This is also confirmed by the ACF and PACF plots (Figure 4.5 d). However, the current study focused on the daily prediction of COVID-19 cases in Ghana. The Augmented Dickey-Fuller (ADF) stationarity test concluded that the daily adjusted series is stationary (p-value=0.01) at lags ≥ 0.



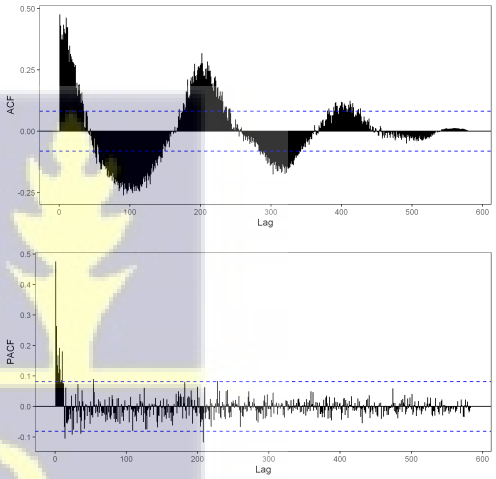
(a) Time Series Plot of the COVID 19 Cases



(b) Original and Adjusted Time Series Plots of the daily COVID-19 Cases



(c) Various Seasonal Plots of the Adjusted Time Series



(d) ACF and PACF plots

Figure 4.5: Adjusted and Original Time Series Plots for COVID-19 Infection Cases, Seasonal Plots, and ACF and PACF plots

4.7 Fitting the Time Series Model

The time series models were fitted based on the daily adjusted series via the out-of-sample evaluation strategy (with maximal and minimal forecasting horizons of 120 days and 90 days, respectively). The first forecast origin was June 20, 2021, and the last was July 20, 2021 (equivalent to 31 lead times). In addition, the underlying time series models were recalibrated (at the different forecast origins) and model parameters re-estimated. Hence, for each fitted time-series model,

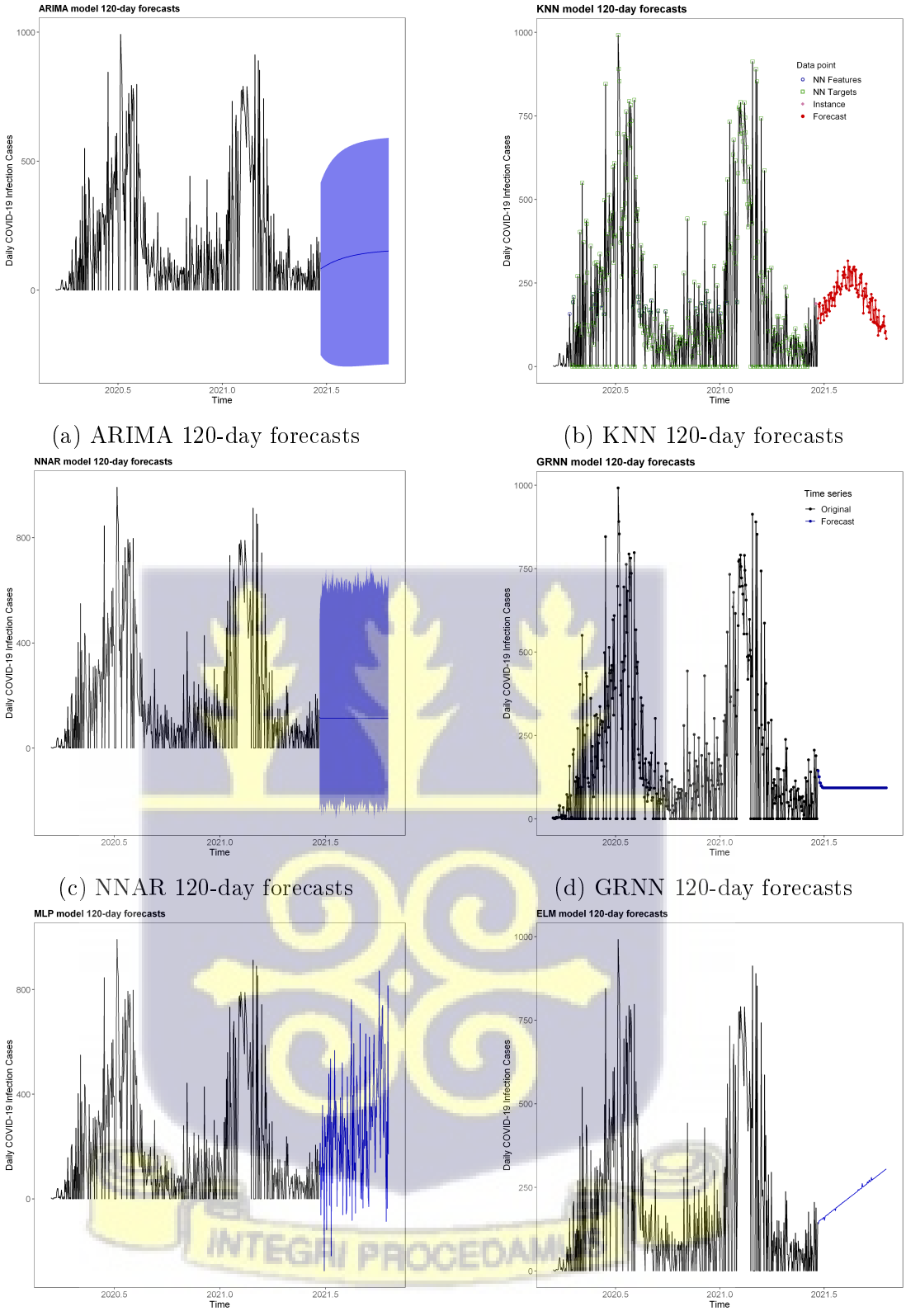
there was an estimated error distribution with a size of 31.

4.7.1 Results of the Machine Learning Time Series Models and the baseline ARIMA model

An automatic baseline ARIMA model and five different machine learning (ML) algorithms were used to predict the daily COVID-19 cases in Ghana with 120-day forecasts via the rolling-origin evaluation strategy. The fitted automatic ARIMA model found the ARIMA (1,0,1) model with a non-zero mean to be the best fit among different classes of non-seasonal ARIMA models. Table 4.7 summarises the fitted ARIMA model parameter estimates (where σ^2 is the variance of the error term in the ARIMA model, and AICc is the corrected AIC value). Like the ARIMA model, the five machine learning algorithms (KNN, GRNN, NNAR, ELM, and MLP) were adaptively fitted via the rolling-origin evaluation strategy (with their forecasting errors evaluated at each lead time as well as estimating their respective computational times). Figure 4.6 compares the six fitted time-series models at the maximal 120-day forecast horizon.

Table 4.7: The Estimates of the ARIMA Model Through MLE

	AR(1)	MA(1)	Mean
Coefficients	0.9802	-0.8119	158.6319
Std.Errors	0.0102	0.0291	69.0865
$\sigma^2 = 29207$	AIC=6093.83	AICc=6093.91	LRT=-3042.91



(a) ARIMA 120-day forecasts

(b) KNN 120-day forecasts

(c) NNAR 120-day forecasts

(d) GRNN 120-day forecasts

(e) MLP 120-day forecasts

(f) ELM 120-day forecasts

Figure 4.6: Comparative plot of the 120-day COVID-19 infection cases from the fitted time-series models

4.8 Model Evaluation and Comparison

The five machine learning time series models were compared to the baseline ARIMA model by exploring the trade-off between computational speed (measured by the computer CPU running time in seconds) and accuracy (across all the 31 lead times or forecast origins as shown in Figure 4.7). A pooled error statistics (quantified by the median RMSE from equation (3.50)) was estimated from their respective error distributions (and presented in Table 4.8), whereas the error distributions between the fitted time-series models were further compared using Kruskal-Wallis rank sum test ($\chi^2 = 144.23$, df=5, p-value=0.000) and its post-hoc Bonferroni-Dunn's test (with p-values presented in Table 4.9).

Table 4.8: Comparing pooled error statistic and computational times

Model	CPU time (in secs)	MdRMSE
ARIMA	4.16	249.4662
KNN	27.77	219.7099
NNAR	5105.73	261.4977
GRNN	8586.38	277.7282
MLP	19123.84	442.2727
ELM	432.51	292.5679

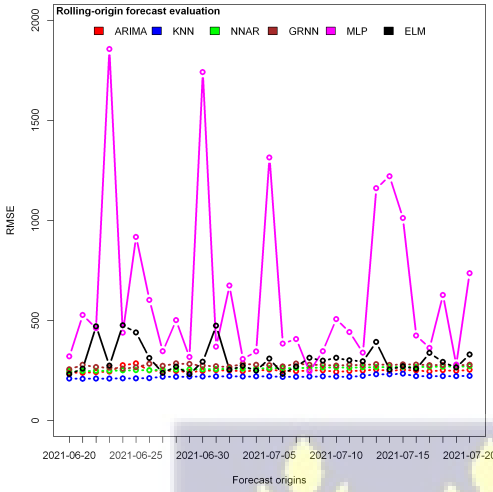
The KNN outperformed the other four ML algorithms (NNAR, GRN, MLP, and ELM) and the baseline ARIMA model in terms of the pooled error statistic measure, with the least computational time among the ML models. Hence, the KNN algorithm is the most cost-effective model among the ML models in predicting the daily cases of COVID-19 in Ghana; whereas the ARIMA model performed better than the other ML algorithms in terms of both the pooled error statistic and its computational time. Thus, the best ML algorithm was KNN followed by NNAR, GRNN, ELM and lastly, MLP; whereas, the least-performing machine learning model was computationally expensive to fit based on the 120-day rolling origin evaluation strategy. Figure 4.7 (a) describes the prediction errors the error distributions of the fitted time-series models across the various forecast origins; whereas Figure 4.7 (b) compares the error distribution between the

predictive models (with MLP significantly showing the greatest error variability or high variance). Figure 4.7 (c) shows that the three most computationally expensive ML algorithms for the daily COVID-19 infection cases (among the machine learning models) are MLP>GRNN>NNAR (with the computational time of ARIMA<KNN). Table 4.9 suggests with the exception of GRNN-ELM insignificant error difference (p-value=1), there as a significant difference between any other pair of the fitted time-series models.

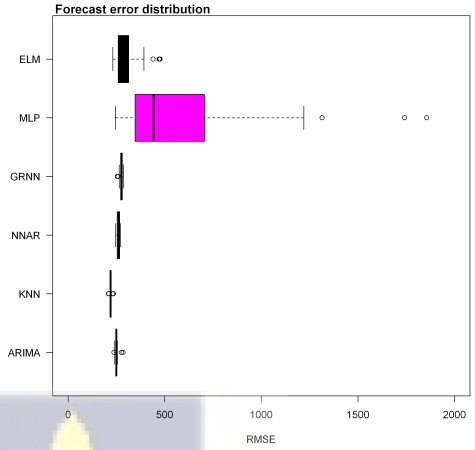
Table 4.9: P-values of the Bonferroni Dunn's Multiple Comparison of Forecast Errors

Model(m,n)	**ARIMA**	**ELM**	**GRNN**	**KNN**	**MLP**
ELM	0.000				
GRNN	0.000	1.000			
KNN	0.000	0.000	0.000		
MLP	0.000	0.000	0.000	0.000	
NNAR	0.000	0.011	0.000	0.000	0.000

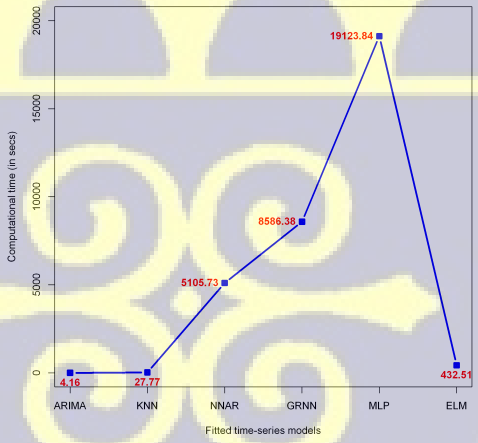
*$p\text{-}value < 0.05$ in ijth position implies forecasting error between models i and j are significantly different at 5% alpha level.



(a) Prediction error comparison plot



(b) Comparison of the forecast error distributions



(c) Comparative plot of computational times

Figure 4.7: Model comparison based on the Forecast Error Distributions, and Computational Time

Chapter 5

Summary, Conclusions, and Recommendations

This chapter summarizes the key findings obtained from the analysis of the COVID-19 transmission in Ghana from 2020 to 2021, highlights some theoretical considerations for further research, and concludes with recommendations to policymakers and individuals.

5.1 Summary of Major Findings

The current study mainly adopted the discrete-time S-I-D Markov model and a few selected state-of-the-art machine learning algorithms to investigate COVID-19 transmission in Ghana from 2020 to 2021. It was identified from the study that through the discrete-time S-I-D Markov model, the probability of first infection among susceptible individuals in the 2020 infection period was comparatively lower than in the 2021 infection period. This finding agrees with other previously published works on COVID-19 transmission during 2021, when new variants emerged. for Disease Prevention & Control (2021) opined that the possible mutation of the virus might lead to a potential reduction in protection against infection. They are of the view that a decrease in serum polyclonal antibodies' neutralizing activity against variant viruses could imply a reduced ability to guard against re-infection. Moreover, as noted by Bakhshandeh et al. (2021), SARS-CoV-2 has a relatively high rate of dynamic mutation compared to other RNA viruses. Thus, these mutations can undoubtedly drive viral genetic variability, resulting in a virus with lower natural pathogenic fitness or, on the other hand,

facilitating rapid antigenic shifting in order to escape host immunity and also creating a drug-resistant virus, thereby converting it to a more infectious or deadly virus due to genotype-phenotype relationship.

The estimated probability of first recovery among infected individuals during both the 2020 and 2021 infection periods was consistent with a decreasing recovery rate over time towards zero. The decrease in the first recovery probability could be due to the fact already highlighted about how dangerous the latter variants of the virus are compared to the former. Thus, it is not surprising that the current study found the overall infection rate (39.3%) during the 2021 infection period to be higher than that of the 2020 infection rate (21.1%). Interestingly, the 2020 overall recovery rate was slightly lower (99.4%) than that of the 2021 (99.1%), however these two recovery rates lie within the global COVID-19 recovery rate indicated by Nazario (2021). It was found that the COVID-19 infection will take approximately 120 years to end if the 2020 study population is considered, but when the 2021 study population was used, the average infection duration was over 96 years (possibly due to the enormous public-health interventions at the local and global levels). This agrees with the findings of Phillips (2021), which suggests COVID-19 will likely exist globally for some years to come. It was also estimated that the mean time to recover fully from the infection in Ghana (at both infection periods) was at least a year. In addition, the Markov model discovered that estimated life expectancy (given the aggregated data at the two infection periods) was high even though it was predicted that the COVID-19 infection is likely to take a longer time to die out from the study populations. Hence, it was confirmed that infected individuals had a relatively better life expectancy than susceptible individuals, possibly because of adaptive immunocompetence after more extended infection periods. Finally, a mathematical expression of the general form of the nth step transition probability matrix at each infection was proposed (via a proposed Lemma and a Corollary), and further prediction of the proportion of infected persons who recover, remain infected or die from the

infection were explored.

Furthermore, a few state-of-the-art ML algorithms (KNN, GRNN, NNAR, ELM, and MLP) were compared to a baseline ARIMA via a rolling-origin evaluation strategy, proposed initially by Twumasi & Twumasi (2021), to forecast the number of COVID-19 infection cases in Ghana over the study period (from 2020 to 2021). Traditional time-series approaches, such as ARIMA and exponential smoothing models, have consistently outperformed complex machine learning algorithms in previous research (Makridakis et al., 2018). Cerqueira et al. (2019) found that the prediction inadequacy of ML models in such circumstances is mainly attributable to potentially small training sample sizes in their studies. Although appropriate hyperparameter adjustment can potentially boost predictive performance, ML algorithms perform better with large training samples. The current study found that the KNN algorithm was the most accurate model to predict the daily COVID-19 infection cases, with the baseline ARIMA model (the fastest model to fit) outperforming the other ML models. Among the ML algorithms, KNN was relatively cost-effective in that it achieved the least prediction error and computational time. However, the Multi-layer Perceptron algorithm was over 688 times more computationally expensive to fit (compared to KNN) via the rolling-origin evaluation strategy. MLP, relative to the other ML algorithms (KNN, NNAR, GRNN, and ELM), significantly showed the highest variability in its error distribution.

5.2 Conclusions

According to the S-I-D Markov model, COVID-19 transmission in Ghana was more prevalent in the 2021 cohort or study population than in the 2020 cohorts (with a lower chance of first infection), and the likelihood of recovery was relatively higher in the 2020 study population than in the 2021 study population. It

was estimated that there would be a prolonged COVID-19 transmission in Ghana for at least 150 years before infection could die out in Ghana. It was confirmed that with a high overall recovery rate, a low infection rate, and a longer period of infection, there is a possibility of herd immunity at the country level and/or adaptive immune response at the individual level (as evident in the 2021 infection period despite the relatively high overall rate of infection). Moreover, the estimated nth step transition matrix showed that the lower the recovery rate, the higher the chance of mortality among the infected population, with a significantly smaller number of infected cases constantly remaining infected in the long run. Finally, the K-Nearest Neighbour (KNN) regression was found to be the most cost-effective machine learning algorithm to predict the daily cases of COVID-19 in Ghana via the rolling-origin evaluation strategy; with the Multilayer Perceptron (MLP) considered as the most under-performing model after exploring the trade-off between computation speed and accuracy.

5.3 Recommendations

Based on the study's findings and discussions, the following recommendations are made:

1. As a result, the Markov chain model is suggested as a viable and robust technique for calculating other essential epidemiological quantities of infectious diseases, as well as generalising transition probabilities for future predictions. However, the study was limited in that impact demographic factors or covariates (such as age, gender, etc.) were not incorporated into the studies.

2. Future studies can also expand the COVID-19 disease outcomes in the S-I-D Markov model to include other relevant states (such as the exposed and

vaccinated states) that can also drive or influence the infection transmission within a given population.

3. Moreover, efficient preventative measures and strict adherence to laws such as preventing public gatherings, travel limitations, personal protection measures, and social distance may help reduce the infection rates in Ghana.

4. The policy of herd immunity should be implemented effectively, and there should be a public orientation on vaccines to reduce hesitancy towards vaccination.

References

Ahmed, N. K., Atiya, A. F., Gayar, N. E., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, *29*(5-6), 594–621.

Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.

Anderson, T. W., & Goodman, L. A. (1957). Statistical inference about markov chains. *The annals of mathematical statistics*, 89–110.

Anyoriga, D. (2020). *Coronavirus: Norwegian embassy in ghana shuts down after staff member tested positive*. Retrieved from `https://citinewsroom.com/2020/03`

Araújo, R. d. A., Oliveira, A. L., & Meira, S. (2017). On the problem of forecasting air pollutant concentration with morphological models. *Neurocomputing*, *265*, 91–104.

Assiri, A., Al-Tawfiq, J. A., Al-Rabeeah, A. A., Al-Rabiah, F. A., Al-Hajjar, S., Al-Barrak, A., ... others (2013). Epidemiological, demographic, and clinical characteristics of 47 cases of middle east respiratory syndrome coronavirus disease from saudi arabia: a descriptive study. *The Lancet infectious diseases*, *13*(9), 752–761.

Atiquzzaman, M., & Kandasamy, J. (2016). Prediction of hydrological time-series using extreme learning machine. *Journal of Hydroinformatics*, *18*(2), 345–353.

Bai, J., Del Campo, C., & Keller, L. R. (2018). Markov chain models in practice: a review of low cost software options. *Investigación Operacional*, *38*(1), 56–62.

Bakhshandeh, B., Jahanafrooz, Z., Abbasi, A., Goli, M. B., Sadeghi, M., Mottaqi, M. S., & Zamani, M. (2021). Mutations in sars-cov-2; consequences in structure, function, and pathogenicity of the virus. *Microbial Pathogenesis*, *154*, 104831.

Bassetti, M., Vena, A., & Giacobbe, D. R. (2020). The novel chinese coronavirus (2019-ncov) infections: Challenges for fighting the storm. *European journal of clinical investigation*, *50*(3).

Beck, J. R., & Pauker, S. G. (1983). The markov process in medical prognosis. *Medical decision making*, *3*(4), 419–458.

Bhat, U. N. (1984). *Elements of applied stochastic process* (2nd ed.). Dallas, Taxas: John Wiley & Sons.

Bishop, C. M., et al. (1995). *Neural networks for pattern recognition*. Oxford university press.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Carlos, W. G., Dela Cruz, C. S., Cao, B., Pasnick, S., & Jamil, S. (2020). Novel wuhan (2019-ncov) coronavirus. *Am J Respir Crit Care Med*, P7–P8.

Catak, M., & Duran, N. (2020). Nonlinear markov chain modelling of the novel coronavirus (covid-19) pandemic. *medRxiv*.

Cauchemez, S., Carrat, F., Viboud, C., Valleron, A., & Boelle, P. (2004). A bayesian mcmc approach to study transmission of influenza: application to household longitudinal data. *Statistics in medicine*, *23*(22), 3469–3487.

Cerqueira, V., Torgo, L., & Soares, C. (2019). Machine learning vs statistical methods for time series forecasting: Size matters. *arXiv preprint arXiv:1909.13316*.

Chen, H., Guo, J., Wang, C., Luo, F., Yu, X., Zhang, W., ... others (2020). Clinical characteristics and intrauterine vertical transmission potential of covid-19 infection in nine pregnant women: a retrospective review of medical records. *The lancet*, *395*(10226), 809–815.

Chen, J., Fu, M. C., Zhang, W., & Zheng, J. (2020). Supporting real-time covid-19 medical management decisions: The transition matrix model approach. *arXiv preprint arXiv:2007.01201*.

CNR. (2021). *Koforidua technical university lauds citi tv, zoomlion for fumigation exercise.* Retrieved from `https://citinewsroom.com/2020/04/koforidua-technical-university-lauds-citi-tv-zoomlion-for-fumigation-exercise/`

Cohen, J. E. (1973). Selective host mortality in a catalytic model applied to schistosomiasis. *The American Naturalist*, *107*(954), 199–212.

Dechsiri, C. (2004). Particle transport in fluidized beds: experiments and stochastic models.

Dehghan Shabani, Z., & Shahnazi, R. (2020). Spatial distribution dynamics and prediction of covid-19 in asian countries: Spatial markov chain approach. *Regional Science Policy & Practice*, *12*(6), 1005–1025.

Demongeot, J., Oshinubi, K., Rachdi, M., Seligmann, H., Thuderoz, F., & Waku, J. (2021). Estimation of daily reproduction numbers during the covid-19 outbreak. *Computation*, *9*(10), 109.

Ducan, J. (2014). *Two cases of corona virus confirmed in ghana.* Retrieved from `https://citinewsroom.com/2020/03`

Dudek, G. (2016). Neural networks for pattern-based short-term load forecasting: A comparative study. *Neurocomputing*, *205*, 64–74.

Du Toit, A. (2020). Outbreak of a novel coronavirus. *Nature Reviews Microbiology*, *18*(3), 123–123.

Dzisi, E. K. J., & Dei, O. A. (2020). Adherence to social distancing and wearing of masks within public transportation during the covid 19 pandemic. *Transportation Research Interdisciplinary Perspectives*, *7*, 100191.

Faraway, J., & Chatfield, C. (1998). Time series forecasting with neural networks: a comparative study using the air line data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *47*(2), 231–250.

for Disease Prevention, E. C., & Control. (2021). *Risk of sars-cov-2 transmission from newly-infected individuals with documented previous infection or vaccination*. ECDC Stockholm, Sweden.

Ghosh, I., & Chakraborty, T. (2021). An integrated deterministic–stochastic approach for forecasting the long-term trajectories of covid-19. *International Journal of Modeling, Simulation, and Scientific Computing*, *12*(03), 2141001.

GhraphicOnline. (2020). *Un ghana staff tests positive for coronavirus, office shutdown*. Retrieved from `https://graphic.com.gh/news/general-news`

GHS. (2020). *Covid-19 updates-ghana*. Retrieved from `ghanahealthservice.org`

Giuliani, D., Dickson, M. M., Espa, G., & Santi, F. (2020). Modelling and predicting the spatio-temporal spread of covid-19 in italy. *BMC infectious diseases*, *20*(1), 1–10.

Gupta, R., & Pal, S. K. (2020). Trend analysis and forecasting of covid-19 outbreak in india. *MedRxiv*.

Haque, A., & Pant, A. B. (2020). Efforts at covid-19 vaccine development: challenges and successes. *Vaccines*, *8*(4), 739.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). Data mining, inference, and prediction. *The elements of statistical learning Springer Series in Statistics. Springer-Verlag, New York*.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). Springer.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, *2*(5), 359–366.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., ... others (2020). Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *The lancet*, *395*(10223), 497–506.

Huang, G.-B., Zhou, H., Ding, X., & Zhang, R. (2011). Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *42*(2), 513–529.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, *22*(4), 679–688.

Kadhem, S. K., & Kadhim, S. A. (2022). Absorbing markov chains for analyzing covid-19 infections. *Asian-European Journal of Mathematics*, *15*(01), 2250008.

Kenu, E., Frimpong, J., & Koram, K. (2020). Responding to the covid-19 pandemic in ghana. *Ghana medical journal*, *54*(2), 72–73.

Khoshnaw, S. H., Salih, R. H., & Sulaimany, S. (2020). Mathematical modelling for coronavirus disease (covid-19) in predicting future behaviours and sensitivity analysis. *Mathematical Modelling of Natural Phenomena*, *15*, 33.

Khrapov, P., & Loginova, A. (2020). Mathematical modelling of the dynamics of the coronavirus covid-19 epidemic development in china. *International Journal of Open Information Technologies*, *8*(4), 13–16.

Kim, K.-j. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, *55*(1-2), 307–319.

Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., ... others (2020). Early dynamics of transmission and control of covid-19: a mathematical modelling study. *The lancet infectious diseases*, *20*(5), 553–558.

Lee, N., Hui, D., Wu, A., Chan, P., Cameron, P., Joynt, G. M., ... others (2003). A major outbreak of severe acute respiratory syndrome in hong kong. *New England Journal of Medicine*, *348*(20), 1986–1994.

Le Gallo, J. (2004). Space-time analysis of gdp disparities among european regions: A markov chains approach. *International Regional Science Review*, *27*(2), 138–163.

Leon-Garcia, A. (1994). *Probability and random processes for electrical engineering*. Pearson Education India.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., ... others (2020). Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. *New England journal of medicine*.

Liu, J., Liao, X., Qian, S., Yuan, J., Wang, F., Liu, Y., ... Zhang, Z. (2020). Community transmission of severe acute respiratory syndrome coronavirus 2, shenzhen, china, 2020. *Emerging infectious diseases*, *26*(6), 1320.

Lu, H. (2020). Drug treatment options for the 2019-new coronavirus (2019-ncov). *Bioscience trends*, *14*(1), 69–71.

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., ... others (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The lancet*, *395*(10224), 565–574.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PloS one*, *13*(3), e0194889.

Marfak, A., Youlyouz-Marfak, I., El Achhab, Y., Saad, E., Nejjari, C., Hilali, A., & Turman Jr, J. (2020). Improved ridit statistic approach provides more intuitive and informative interpretation of eq-5d data. *Health and quality of life outcomes*, *18*(1), 1–14.

Martínez, F., Frías, M. P., Charte, F., & Rivera, A. J. (2019). Time series forecasting with knn in r: the tsfknn package. *R J.*, *11*(2), 229.

Moshiri, S., & Cameron, N. (2000). Neural network versus econometric models in forecasting inflation. *Journal of forecasting*, *19*(3), 201–217.

Musa, S. S., Zhao, S., Wang, M. H., Habib, A. G., Mustapha, U. T., & He, D. (2020). Estimation of exponential growth rate and basic reproduction number of the coronavirus disease 2019 (covid-19) in africa. *Infectious diseases of poverty*, *9*(1), 1–6.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications*, *9*(1), 141–142.

Nazario, B. (2021). *Coronavirus and covid-19: What you should know.* Retrieved from `https://www.webmd.com/lung/coronavirus` (Last accessed 1 January 2022)

Nishiura, H., Jung, S.-m., Linton, N. M., Kinoshita, R., Yang, Y., Hayashi, K., ... Akhmetzhanov, A. R. (2020). *The extent of transmission of novel coronavirus in wuhan, china, 2020* (Vol. 9) (No. 2). Multidisciplinary Digital Publishing Institute.

Ofori-Atta, K. (2020). Statement to parliament on economic impact of the covid-19 pandemic on the economy of ghana. *Accra: Ministry of Finance*.

Ong, S. W. X., Tan, Y. K., Chia, P. Y., Lee, T. H., Ng, O. T., Wong, M. S. Y., & Marimuthu, K. (2020). Air, surface environmental, and personal protective equipment contamination by severe acute respiratory syndrome coronavirus 2 (sars-cov-2) from a symptomatic patient. *Jama*, *323*(16), 1610–1612.

Phillips, N. (2021). *The coronavirus is here to stay — here's what that means.* Retrieved from `https://www.nature.com/articles/d41586-021-00396-2` (Last accessed 1 January 2022)

Presidency, T. (2020). *Speeches.* Retrieved from `https://presidency.gov.gh/index.php/briefing-room/speeches.org`

R Core Team. (2022). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Ren-LL, W. Y., Wu, Z., et al. (2020). Identification of a novel coronavirus causing severe pneumonia in human. *Chin Med J*, *133*(9), 1015–1024.

Rooki, R. (2016). Application of general regression neural network (grnn) for indirect measuring pressure loss of herschel–bulkley drilling fluids in oil drilling. *Measurement*, *85*, 184–191.

Rothan, H. A., & Byrareddy, S. N. (2020). The epidemiology and pathogenesis of coronavirus disease (covid-19) outbreak. *Journal of autoimmunity*, *109*, 102433.

Serfozo, R. (2009). *Basics of applied stochastic processes.* Springer Science & Business Media.

Shi, Z., & Fang, Y. (2020). Temporal relationship between outbound traffic from wuhan and the 2019 coronavirus disease (covid-19) incidence in china. *MedRxiv*.

Shumway, R. H., Stoffer, D. S., & Stoffer, D. S. (2000). *Time series analysis and its applications* (Vol. 3). Springer.

Singh, R., & Balasundaram, S. (2007). Application of extreme learning machine method for time series analysis. *International Journal of Intelligent Technology*, *2*(4), 256–262.

Sookaromdee, P., Wiwanitkit, V., et al. (2020). Imported cases of 2019-novel coronavirus (2019-ncov) infections in thailand: Mathematical modelling of the outbreak. *Asian Pac J Trop Med*, *13*(3), 139–140.

Specht, D. F., et al. (1991). A general regression neural network. *IEEE transactions on neural networks*, *2*(6), 568–576.

Suryaningrat, W., Munandar, D., Maryati, A., Abdullah, A., & Ruchjana, B. (2021). Posted prediction in social media base on markov chain model: twitter dataset with covid-19 trends. In *Journal of physics: Conference series* (Vol. 1722, p. 012001).

Takele, R. (2020). Stochastic modelling for predicting covid-19 prevalence in east africa countries. *Infectious Disease Modelling*, *5*, 598–607.

Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International journal of forecasting*, *16*(4), 437–450.

Team, I. (2020). First case of 2019 novel coronavirus in the united states. *N Engl J Med*.

Twumasi, C. (2018). *Statistical modeling of hiv, tuberculosis, and hepatitis b transmission in ghana* [Master's Thesis]. Retrieved from `http://ugspace.ug.edu.gh/handle/123456789/28099`

Twumasi, C., Asiedu, L., & Nortey, E. N. (2019). Markov chain modeling of hiv, tuberculosis, and hepatitis b transmission in ghana. *Interdisciplinary Perspectives on Infectious Diseases, 2019*.

Twumasi, C., & Twumasi, J. (2021). Machine learning algorithms for forecasting and backcasting blood demand data with missing values and outliers: A study of tema general hospital of ghana. *International Journal of Forecasting*.

Wan, Y., Shang, J., Graham, R., Baric, R. S., & Li, F. (2020). Receptor recognition by the novel coronavirus from wuhan: an analysis based on decade-long structural studies of sars coronavirus. *Journal of virology*, *94*(7), e00127–20.

Wang, W., Tang, J., & Wei, F. (2020). Updated understanding of the outbreak of 2019 novel coronavirus (2019-ncov) in wuhan, china. *Journal of medical virology*, *92*(4), 441–447.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 359–372.

WHO. (2014). *Report of the who-china joint mission on coronavirus disease 2019 (covid-19)*. Retrieved from `https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf`

WHO. (2020a). *Archived: Who timeline - covid-19*. Retrieved from `www.who.int`

WHO. (2020b). *Coronavirus disease 2019 (covid-19) situation report*. Retrieved from `https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports`

Worldometer. (2020). *Current number of corona virus cases in ghana*. Retrieved from `https://www.worldometers.info/coronavirus/country/ghana/`

Wu, K., Darcet, D., Wang, Q., & Sornette, D. (2020). Generalized logistic growth modeling of the covid-19 outbreak: comparing the dynamics in the 29 provinces in china and in the rest of the world. *Nonlinear dynamics*, *101*(3), 1561–1581.

Wu, K., Zheng, J., & Chen, J. (2020). Utilize state transition matrix model to predict the novel corona virus infection peak and patient distribution. *Available at SSRN 3539658*.

Yonar, H., Yonar, A., Tekindal, M. A., & Tekindal, M. (2020). Modeling and forecasting for the number of cases of the covid-19 pandemic with the curve estimation models, the box-jenkins and exponential smoothing methods. *EJMO*, *4*(2), 160–165.

Zaki, A. M., Van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D., & Fouchier, R. A. (2012). Isolation of a novel coronavirus from a man with pneumonia in saudi arabia. *New England Journal of Medicine*, *367*(19), 1814–1820.

Zhang, W., Zhang, C., Bi, Y., Yuan, L., Jiang, Y., Hasi, C., ... Kong, X. (2021). Analysis of covid-19 epidemic and clinical risk factors of patients under epidemiological markov model. *Results in Physics*, *22*, 103881.

Zipkin, E. F., Jennelle, C. S., & Cooch, E. G. (2010). A primer on the application of markov chains to the study of wildlife disease dynamics. *Methods in Ecology and Evolution*, *1*(2), 192–198.

Appendix A

5.4 R Codes for Markov Chain Modelling

Transition counts in 2021

```
Total_population<-NULL; Total_susceptible<- NULL;
Transition_counts<- NULL;Total_death<- NULL
Total_infected<- NULL; Total_recovered<- NULL;Total_induced_death<-
NULL
Total_population[[2021]]<- 31732129
Total_infected[[2021]]<- 129440
Total_recovered[[2021]]<- 125839
Total_induced_death[[2021]]<- 1167
Total_death[[2021]]<- 0.0063*Total_population[[2021]]

Total_susceptible[[2021]]<- Total_population[[2021]]-Total_infected[[2021]]-
Total_death[[2021]]

Transition_counts[[2021]]<- matrix(NA,nrow=2, ncol=3)
rownames(Transition_counts[[2021]])<- c("S","I")
colnames(Transition_counts[[2021]])<- c("S","I","D")

Transition_counts[[2021]][1, 1]<-Total_susceptible[[2021]]
Transition_counts[[2021]][1, 2]<- Total_infected[[2021]]
Transition_counts[[2021]][1, 3]<- Total_death[[2021]]
```

Transition counts in 2021

```
Transition_counts[[2021]][2, 1]<- Total_recovered[[2021]]
Transition_counts[[2021]][2, 2]<- Total_infected[[2021]]-Total_recovered[[2021]]
Transition_counts[[2021]][2, 3]<- Total_induced_death[[2021]]


Transition_counts[[2021]]
```

Transition counts in 2020

```
Total_population[[2020]]<- 31072940

Total_infected[[2020]]<- 54711

Total_recovered[[2020]]<- 53594

Total_induced_death[[2020]]<- 335

Total_death[[2020]]<- 0.00658*Total_population[[2020]]


Total_susceptible[[2020]]<-

Total_population[[2020]]-Total_infected[[2020]]-Total_death[[2020]]


Transition_counts[[2020]]<- matrix(NA,nrow=2, ncol=3)

rownames(Transition_counts[[2020]])<- c("S","I")

colnames(Transition_counts[[2020]])<- c("S","I","D")


Transition_counts[[2020]][1, 1]<-Total_susceptible[[2020]]

Transition_counts[[2020]][1, 2]<- Total_infected[[2020]]

Transition_counts[[2020]][1, 3]<- Total_death[[2020]]


Transition_counts[[2020]][2, 1]<- Total_recovered[[2020]]

Transition_counts[[2020]][2, 2]<- Total_infected[[2020]]-Total_recovered[[2020]]

Transition_counts[[2020]][2, 3]<- Total_induced_death[[2020]]


Transition_counts[[2020]]
```

Transition probability based on Maximum likelihood Estimation

```
Row1_prop<-NULL; Row2_prop<- NULL


#Year 2020 and 2021
for(year in c(2020,2021)) Row1_prop[[year]]<- Transition_counts[[year]][1,
]/sum(Transition_counts[[year]][1, ])
Row2_prop[[year]]<-Transition_counts[[year]][2,
]/sum(Transition_counts[[year]][2, ])



# Define the transition probabilities
p00 <- NULL; p01<- NULL; p02<- NULL; p10<- NULL;
p11<- NULL; p12<-NULL; p20<- NULL; p21<- NULL; p22<- NULL

for(year in c(2020,2021)){ p00[[year]]<- Row1_prop[[year]][1]
p01[[year]]<- Row1_prop[[year]][2]
p02[[year]]<- Row1_prop[[year]][3]

p10[[year]]<- Row2_prop[[year]][1]
p11[[year]]<- Row2_prop[[year]][2]
p12[[year]]<- Row2_prop[[year]][3]
```

Transition probability based on Maximum likelihood Estimation

```
p20[[year]]<- 0
p21[[year]]<- 0
p22[[year]]<- 1
# Enter the transition probabilities into a matrix P
P<- NULL

for(year in c(2020,2021)){
P[[year]]<- matrix(c(p00[[year]],p01[[year]],p02[[year]],
p10[[year]],p11[[year]],p12[[year]],
p20[[year]],p21[[year]],p22[[year]]),3,3,byrow=TRUE)
rownames(P[[year]])<- c("S","I","D")
colnames(P[[year]])<- rownames(P[[year]])
}
#Estimated transition probability matrix in year 2020
print(paste("Year 2020")) P[[2020]]
#Estimated transition probability matrix in year 2021
print(paste("Year 2021")) P[[2021]]
```

Maximum likelihood estimates with Standard error and 99% Confidence intervals for year 2020 & 2021

```
parameters<- c("P00","P01","P02","P10","P11","P12")
Estimates<- NULL;std_err<- NULL
for(year in c(2020,2021)) Estimates[[year]]<- numeric(length(parameters))
for(year in c(2020,2021)) std_err[[year]]<- numeric(length(parameters))

for(year in c(2020,2021)){ Estimates[[year]][1]<- P[[year]][1,1]
std_err[[year]][1]<-
sqrt((Estimates[[year]][1]*(1-Estimates[[year]][1]))/sum(Transition_counts[[year]][1,]))
Estimates[[year]][2]<- P[[year]][1,2]
std_err[[year]][2]<-
sqrt((Estimates[[year]][2]*(1-Estimates[[year]][2]))/sum(Transition_counts[[year]][1,]))

Estimates[[year]][3]<- P[[year]][1,3]
std_err[[year]][3]<-
sqrt((Estimates[[year]][3]*(1-Estimates[[year]][3]))/sum(Transition_counts[[year]][1,]))

Estimates[[year]][4]<- P[[year]][2,1]
std_err[[year]][4]<-
sqrt((Estimates[[year]][4]*(1-Estimates[[year]][4]))/sum(Transition_counts[[year]][2,]))
Estimates[[year]][5]<- P[[year]][2,2]
std_err[[year]][5]<-
sqrt((Estimates[[year]][5]*(1-Estimates[[year]][5]))/sum(Transition_counts[[year]][2,]))
Estimates[[year]][6]<- P[[year]][2,3]
std_err[[year]][6]<-
sqrt((Estimates[[year]][6]*(1-Estimates[[year]][6]))/sum(Transition_counts[[year]][2,]))
}
```

Maximum likelihood estimates with Standard error and 99% Confidence intervals for year 2020 & 2021

```
Confi_int99<- function(Est,SE)
lower<- numeric(length(Est));upper<- numeric(length(Est))
for(i in seq_along(Est))
lower[i]<- Est[i]- (2.575*SE[i])
upper[i]<- Est[i]+ (2.575*SE[i])

return(list(lower=lower,upper=upper))

C.I<- NULL
for(year in c(2020,2021)) C.I[[year]]<- Confi_int99(Est= Estimates[[year]],SE=
std_err[[year]])

MLE_transProb<- NULL
for(year in c(2020,2021)) MLE_transProb[[year]]<- data.frame(parameters= param-
eters,
MLE_estimates=Estimates[[year]],std_err=std_err[[year]],
C.I_lower99=C.I[[year]]lower,C.I_upper99=C.I[[year]]upper)
```

Maximum likelihood estimates with Standard error and 99% Confidence intervals for year 2020 & 2021

```
#Saving MLE estimates with their 99% confidence intervals for year 2020 and 2021
for (year in c(2020,2021))
print(paste("Estimates for year","",year))
print(MLE_transProb[[year]])
#saving results as a csv file
write.csv(MLE_transProb[[year]],paste0("MLE_transProb_",year,".csv") )
```

Using likelihood ratio test (LRT) or AIC to compare between the two fitted Markov Models

```
#Function for calculating the log-likelihood
log_likehood<- function(transition_count,transition_prob)
loglik<- sum(transition_count *log(transition_prob[1:2, ]))
return(loglik)

AIC_stats<- function(parameters,loglik) (-2*loglik)+ (2*parameters)
LogLik<- NULL; AIC_values<- NULL
for (year in c(2020,2021))
LogLik[[year]]<- log_likehood(transition_count=
Transition_counts[[year]],transition_prob=P[[year]])
print(paste("Loglik at year:",year))
print(LogLik[[year]])

AIC_values[[year]]<-AIC_stats(parameters=6,loglik=LogLik[[year]])
print(paste("AIC at year:",year))
print(AIC_values[[year]])

#Test statistics of LRT
lRT_teststat <- -2 * (as.numeric(LogLik[[2020]])-as.numeric(LogLik[[2021]]))
lRT_teststat

#p-value
(p.val <- pchisq(lRT_teststat, df = 1, lower.tail = FALSE))
```

Determining recurrent and transient states (& absorbing states)

```
#Install packages
library(markovchain)
library(igraph)
library(diagram)
library(expm)
#library(markovchain)
disease_states<- c("S","I","D")
year<- 2020
tmat_2020 <- matrix(c(p00[[year]],p01[[year]],p02[[year]], p10[[year]],
p11[[year]],p12[[year]], p20[[year]],p21[[year]],p22[[year]]),
nrow = 3, byrow = TRUE,dimnames = list(disease_states,disease_states))

dtmcA_2020<-new("markovchain",transitionMatrix=tmat_2020,
states=disease_states,
name="MarkovChain P")
print(summary(dtmcA_2020))

year<- 2021
tmat_2021 <- matrix(c(p00[[year]],p01[[year]],p02[[year]], p10[[year]],
p11[[year]],p12[[year]], p20[[year]],p21[[year]],p22[[year]]),
nrow = 3, byrow = TRUE,dimnames = list(disease_states,disease_states))

dtmcA_2021<-new("markovchain",transitionMatrix=tmat_2021,
states=disease_states,
name="MarkovChain P")
#print(summary(dtmcA_2021))
```

Transition diagram

```
options(repr.plot.width=8, repr.plot.height=8,repr.plot.res = 300) #Setting plot size
#For year 2020
par(mar=c(0,0,0,0))
plotmat(t(tmat_2020),pos = c(2,1),
lwd = 1, box.lwd = 2,
cex.txt = 0.8,
box.size = 0.1,
box.type = "circle",
box.prop = 0.5,
box.col = "light yellow",
#arr.length=.1,
#arr.width=.1,
self.cex = .4,
#self.shifty = -.01,
#self.shiftx = .13,
shadow.size=0,self.arrpos=0.5,
endhead = F,
name=disease_states,
main ="")
```

Transition diagram

```
mtext("Year 2020",side=3,col="black",line=-3,font=2)
#For year 2021
par(mar=c(0,0,0,0))
plotmat(t(tmat_2021),pos = c(2,1),
lwd = 1, box.lwd = 2,
cex.txt = 0.8,
box.size = 0.1,
box.type = "circle",
box.prop = 0.5,
box.col = "light yellow",
#arr.length=.1,
#arr.width=.1,
self.cex = .4,
#self.shifty = -.01,
#self.shiftx = .13,
shadow.size=0,self.arrpos=0.5,
endhead = F,
name=disease_states,
main ="")

mtext("Year 2021",side=3,col="black",line=-3,font=2)
```

Calculate the probability of first infection for susceptible individuals & first recovery for infected individuals in year 2020 and 2021

```
### Calculate the probability of first infection for susceptible individuals
#and first recovery for infected individuals in year 2020 and 2021


# Create the number of time steps, n
n=c(1:50)


#Probability of first transition (Probability of first infection)
f01n<- NULL; f10n<- NULL


for (year in c(2020,2021))
f01n[[year]]<- rep(NA,length(n))
f10n[[year]]<- rep(NA,length(n))


#Probability of first transition (Probability of first recovery)
#for a infected individuals
for (year in c(2020,2021)){
for (m in 1:length(n)) {
f01n[[year]][m] <- p01[[year]] * p00[[year]]^(m-1)
f10n[[year]][m] <- p10[[year]] * p11[[year]]^(m-1)
} }
```

Calculate the probability of first infection for susceptible individuals & first recovery for infected individuals in year 2020 and 2021

```
# print the values
#print(f01n)
#print(f10n)


o<-par(mar=c(0,4,2,2))#Run this before the code below
nf<-layout(matrix(1:2, nrow=2,ncol=1))
par(o)# o<-par(mar=c(0,4,2,2))
plot(n,f01n[[2020]],xlab="",ylab="Probability of first infection",type="l",main=""
,lwd=3,col="blue",xaxt = "n",ylim=c(0.0005,.005))
lines(n,f01n[[2021]],xlab="",lwd=3,col="red",xaxt = "n")
text(25,.0045,"Among susceptible individuals",col="black",font=2)
par(mar=c(4,4,0,2))
plot(n,f10n[[2020]],col="blue",xlab="Time steps",ylab="Probability of first
recovery",type="l",lwd=3,
ylim = c(0, 1))
lines(n,f10n[[2021]],col="red",xlab="Time steps",type="l",lwd=3,
ylim = c(0, 1))
text(25,.95,"Among COVID-19 infected individuals",col="black",font=2)


text <- c("Year 2020","Year 2021")
legend(x=25,y=.75,legend = text,col=c("blue","red"),cex=1, horiz = FALSE,pt.cex
= 1,
box.lwd = 2,fill=c("blue","red"),ncol=1,title="Infection period")
```

Cummulative probability of infection

```
#Alternatively fptMc <- new("markovchain", transitionMatrix=P[[2020]],
name="FistPassageTime",
states=c("S" ,"I","D"))
#fptMc
FirstP<-firstPassage(fptMc,state = "S",50)
print(unlist(FirstP[,2])) #First passage or transition from state S to
pinfected<- NULL; precovery<- NULL

for (year in c(2020,2021))
pinfected[[year]]<- rep(0,length(n))
precovery[[year]]<- rep(0,length(n))

# Write a loop that sums the probability of first infection at each
# time step
for (year in c(2020,2021))
for (j in 1:length(n))
pinfected[[year]][j] = sum(f01n[[year]][1:j])
precovery[[year]][j] = sum(f10n[[year]][1:j])

plot(n,pinfected[[2020]],xlab="Time steps",ylab="Cummulative probability of
infection",type="l",ylim = c(0,.2),
main="" ,lwd=3,las=1,col="blue")
```

Cummulative probability of infection

```
lines(n,pinfected[[2021]],xlab="Time steps",ylab="Cummulative probability of
infection",type="l",col="red",lwd=3)


text <- c("Year 2020","Year 2021")
legend(x=10,y=.2,legend = text,col=c("blue","red"),cex=1, horiz = FALSE,pt.cex
= 1,
box.lwd = 2,fill=c("blue","red"),ncol=1,title="Infection period")


print(f01n[[2020]])
```

Estimating overall probability of infection Pr(i→ j)

```
Overall_prob_infection<- NULL;Overall_prob_recovery<-
NULL;std_err_infection<-NULL;
std_err_recovery<-NULL; C.Int_infection<- NULL;C.Int_recovery<- NULL
for(year in c(2020,2021)){
#from state 0 (S) to 1 (I) Overall_prob_infection[[year]]<- p01[[year]]/(1-
p00[[year]]) std_err_infection[[year]]<-sqrt((Overall_prob_infection[[year]]*(1-
Overall_prob_infection[[year]]))
/sum(Transition_counts[[year]][1,]))
C.Int_infection[[year]]<- Confi_int99(Est= Overall_prob_infection[[year]],SE=
std_err_infection[[year]])

#from state 1 (I) to 0 (S)
Overall_prob_recovery[[year]]<- p10[[year]]/(1-p11[[year]])
std_err_recovery[[year]]<-sqrt((Overall_prob_recovery[[year]]*(1-
Overall_prob_recovery[[year]]))
/sum(Transition_counts[[year]][2,]))

C.Int_recovery[[year]]<- Confi_int99(Est= Overall_prob_recovery[[year]],SE=
std_err_recovery[[year]])
}
```

Estimating overall probability of infection Pr(i→ j)

```
Results_overall_estimates<- NULL
metric<-c("Overall infection prob","Overall recovery prob")
for(year in c(2020,2021))
Results_overall_estimates[[year]]<- data.frame(metric= metric,
Estimates=c(Overall_prob_infection[[year]],Overall_prob_recovery[[year]]),
std_err=c(std_err_infection[[year]],std_err_recovery[[year]]),
C.I_lower99= c(C.Int_infection[[year]]lower,C.Int_recovery[[year]]lower),
C.I_upper99= c(C.Int_infection[[year]]upper,C.Int_recovery[[year]]upper))
print(paste("Year:",year))
print(Results_overall_estimates[[year]])
#saving results as a csv file
write.csv(Results_overall_estimates[[year]],
paste0("Results_overall_estimates_",year,".csv") )
```

Estimating the expected time to infection and recovery

```
Expected_time_to_infection<- NULL;Expected_time_to_recovery<- NULL
for(year in c(2020,2021)){
#from state 0 (S) to 1 (I)
Expected_time_to_infection[[year]]<- 1/(1-p00[[year]])
print(paste(" Expected time to infection for Year:",year))
print(Expected_time_to_infection[[year]])
#from state 1 (I) to 0 (S)
Expected_time_to_recovery[[year]]<- 1/(1-p11[[year]])
print(paste(" Expected time to recovery for Year:",year))
print(Expected_time_to_recovery[[year]]) }
```

Estimating the life expectancy for infected and susceptible individuals, W

```
for(year in c(2020,2021)){
Q <- P[[year]][1:2,1:2]
# Define an identity matrix called Identity
Id <- diag(1,2,2)
# Find the inverse of the Identity matrix minus Q using
# the command "solve" and call it InvsIminusQ
IminusQ <- Id - Q
InvsIminusQ <- solve(IminusQ)
# Create a 2 by 1 vector with the value 1 in each element
one <- matrix(c(1,1),2,1,byrow=T)
# find life expectancies
W <- (InvsIminusQ)%*%one

# Maximum number of years susceptible and infected individuals can live (during
COVID-19 era)
print(paste("Estimated life expectancy during the infection Year:",year))
print(W) }
```

P^n transition probability matrix

```
P_nth_components<- function(transition_matrix,n,printeigen=TRUE,d.p=3){
#Finding the eigenvalues
if(printeigen==TRUE){
print(eigen(transition_matrix))
}
eigenvalue<- eigen(transition_matrix)$values
Q_eigenvectors<- eigen(transition_matrix)$vectors

x<-NULL; v<- NULL
#Eigen values of transition matrix P
lambda_0<- eigenvalue[1]; lambda_1<- eigenvalue[2]; lambda_2<- eigenvalue[3]
Q1<- Q_eigenvectors[1, ]; Q2<- Q_eigenvectors[2, ]; Q3<- Q_eigenvectors[3, ]
#Entries of the Q matrix of eigenvectors
x[[11]]<-Q1[1];x[[12]]<-Q1[2];x[[13]]<-Q1[3]
x[[21]]<-Q2[1];x[[22]]<-Q2[2];x[[23]]<-Q2[3]
x[[31]]<-Q3[1];x[[32]]<-Q3[2];x[[33]]<-Q3[3]

#Entries of the inverse of Q matrix (denoted by matrix V)
Q_inverse<- solve(Q_eigenvectors)
V<- Q_inverse
V1<-V[1, ]; V2<-V[2, ]; V3<-V[3, ]
v[[11]]<-V1[1];v[[12]]<-V1[2];v[[13]]<-V1[3]
v[[21]]<-V2[1];v[[22]]<-V2[2];v[[23]]<-V2[3]
v[[31]]<-V3[1];v[[32]]<-V3[2];v[[33]]<-V3[3]
P_n<- matrix(rep(NA,9), nrow = 3, dimnames = list(c("S","I","D"),
c("S","I","D")),byrow=T)
```

P^n transition probability matrix

```
P_n[1,1]<- paste(round(x[[11]]*v[[11]],3), "(",round(lambda_0,3),")",paste("n̂+"),
round(x[[12]]*v[[21]],3),"(",round(lambda_1,3),")",paste("n̂+"),
round(x[[13]]*v[[31]],3),"(",round(lambda_2,3),")",paste("n̂"))
P_n[1,2]<- paste(round(x[[11]]*v[[12]],3), "(",round(lambda_0,3),")",paste("n̂+"),
round(x[[12]]*v[[22]],3),"(",round(lambda_1,3),")",paste("n̂+"),
round(x[[13]]*v[[32]],3),"(",round(lambda_2,3),")",paste("n̂"))
P_n[1,3]<- paste(round(x[[11]]*v[[13]],3), "(",round(lambda_0,3),")",paste("n̂+"),
round(x[[12]]*v[[23]],3),"(",round(lambda_1,3),")",paste("n̂+"),
round(x[[13]]*v[[33]],3),"(",round(lambda_2,3),")",paste("n̂"))
P_n[2,1]<- paste(round(x[[21]]*v[[11]],3), "(",round(lambda_0,3),")",paste("n̂+"),
round(x[[22]]*v[[21]],3),"(",round(lambda_1,3),")",paste("n̂+"),
round(x[[23]]*v[[31]],3),"(",round(lambda_2,3),")",paste("n̂"))
P_n[2,2]<- paste(round(x[[21]]*v[[12]],3), "(",round(lambda_0,3),")",paste("n̂+"),
round(x[[22]]*v[[22]],3),"(",round(lambda_1,3),")",paste("n̂+"),
round(x[[23]]*v[[32]],3),"(",round(lambda_2,3),")",paste("n̂"))
P_n[2,3]<- paste(round(x[[21]]*v[[13]],3), "(",round(lambda_0,3),")",paste("n̂+"),
round(x[[22]]*v[[23]],3),"(",round(lambda_1,3),")",paste("n̂+"),
round(x[[23]]*v[[33]],3),"(",round(lambda_2,3),")",paste("n̂"))
```

P^n transition probability matrix

```
P_n[3,1]<- paste(round(x[[31]]*v[[11]],3), "(",round(lambda_0,3),")",paste("n̂+"),
round(x[[32]]*v[[21]],3),"(",round(lambda_1,3),")",paste("n̂+"),
round(x[[33]]*v[[31]],3),"(",round(lambda_2,3),")",paste("n̂"))
P_n[3,2]<- paste(round(x[[31]]*v[[12]],3), "(",round(lambda_0,3),")",paste("n̂+"),
round(x[[32]]*v[[22]],3),"(",round(lambda_1,3),")",paste("n̂+"),
round(x[[33]]*v[[32]],3),"(",round(lambda_2,3),")",paste("n̂"))
P_n[3,3]<- paste(round(x[[31]]*v[[13]],3), "(",round(lambda_0,3),")",paste("n̂+"),
round(x[[32]]*v[[23]],3),"(",round(lambda_1,3),")",paste("n̂+"),
round(x[[33]]*v[[33]],3),"(",round(lambda_2,3),")",paste("n̂"))
P_n_matrix<- matrix(rep(NA,9), nrow = 3, dimnames = list(c("S","I","D"),
c("S","I","D")),byrow=T)
P_n_matrix[1,1]<- (x[[11]]*v[[11]]*lambda_0n̂)+(x[[12]]*v[[21]]*lambda_1n̂)
+(x[[13]]*v[[31]]*lambda_2n̂)
P_n_matrix[1,2]<- (x[[11]]*v[[12]]*lambda_0n̂)+(x[[12]]*v[[22]]*lambda_1n̂)+
(x[[13]]*v[[32]]*lambda_2n̂)
P_n_matrix[1,3]<- (x[[11]]*v[[13]]*lambda_0n̂)+(x[[12]]*v[[23]]*lambda_1n̂)+
(x[[13]]*v[[33]]*lambda_2n̂)
P_n_matrix[2,1]<-
(x[[21]]*v[[11]]*lambda_0n̂)+(x[[22]]*v[[21]]*lambda_1n̂)+
(x[[23]]*v[[31]]*lambda_2n̂)
```

P^n transition probability matrix

```
P_n_matrix[2,2]<-
(x[[21]]*v[[12]]*lambda_0n̂)+(x[[22]]*v[[22]]*lambda_1n̂)+
(x[[23]]*v[[32]]*lambda_2n̂) P_n_matrix[2,3]<-
(x[[21]]*v[[13]]*lambda_0n̂)+(x[[22]]*v[[23]]*lambda_1n̂)+
(x[[23]]*v[[33]]*lambda_2n̂)
P_n_matrix[3,1]<-
(x[[31]]*v[[11]]*lambda_0n̂)+(x[[32]]*v[[21]]*lambda_1n̂)+
(x[[33]]*v[[31]]*lambda_2n̂)
P_n_matrix[3,2]<-
(x[[31]]*v[[12]]*lambda_0n̂)+(x[[32]]*v[[22]]*lambda_1n̂)+
(x[[33]]*v[[32]]*lambda_2n̂)
P_n_matrix[3,3]<- (x[[31]]*v[[13]]*lambda_0n̂)+(x[[32]]*v[[23]]*lambda_1n̂)+
(x[[33]]*v[[33]]*lambda_2n̂)
return(list(P_n_exact_form=P_n,P_n_matrix=round(P_n_matrix,d.p)))        }
#**Testing at n=6 for example**
Year<- 2020
P_nth_components(transition_matrix=P[[Year]],n=6,printeigen=TRUE,d.p=3)  #
Testing at
n=6 for example
#**Testing at n=6 for example**
Year<- 2021
P_nth_components(transition_matrix=P[[Year]],n=6,printeigen=TRUE,d.p=3)  #
Testing at
n=6 for example
```

Predicting the proportion of infected individuals who become susceptible, remain infected or die from COVID-19 over time from the estimated P^n transition matrix

```
Prop_I_S<-NULL;Prop_I_I<- NULL; Prop_I_D<- NULL
total_years<-200
for(year in c(2020,2021)) Prop_I_S[[year]]<- rep(NA,length=total_years)
for(year in c(2020,2021)) Prop_I_I[[year]]<- rep(NA,length=total_years)
for(year in c(2020,2021)) Prop_I_D[[year]]<- rep(NA,length=total_years)
for(year in c(2020,2021)){
for(t in 1:total_years){
Prop_I_S[[year]][t]<- P_nth_components(transition_matrix=P[[year]],
n=t,printeigen=FALSE,d.p=6)$P_n_matrix[2,1]
Prop_I_I[[year]][t]<- P_nth_components(transition_matrix=P[[year]],
n=t,printeigen=FALSE,d.p=6)$P_n_matrix[2,2]

Prop_I_D[[year]][t]<- P_nth_components(transition_matrix=P[[year]],
n=t,printeigen=FALSE,d.p=6)$P_n_matrix[2,3] }
}
o<-par(mar=c(0,4,2,2))#Run this before the code below
nf<-layout(matrix(1:2, nrow=2,ncol=1))
Year=2020
plot(Prop_I_S[[Year]],type="l",col="blue",lwd=3,ylim=c(0,1),xaxt = "n",
ylab="Transition probability from infected state",las=1)
lines(Prop_I_I[[Year]],type="l",col="red",lwd=3)
lines(Prop_I_D[[Year]],type="l",col="green",lwd=3)
abline(v=105, col="black",lty=3,lwd=0.5)
text(100,.99,"Predictions from the nth-step transition matrix starting from year
2020",col="black",font=2)
```

Predicting the proportion of infected individuals who become susceptible, remain infected or die from COVID-19 over time from the estimated P^n transition matrix

```
text <- c("S","I","D")
legend(x=1,y=.7,legend = text,col=c("blue","red","green"),cex=1, horiz = FALSE,pt.cex = 1,
box.lwd = 2,fill=c("blue","red","green"),ncol=1,title="From state I to")
par(mar=c(4,4,0,2))
Year=2021
plot(Prop_I_S[[Year]],type="l",col="blue",lwd=3,ylim=c(0,.98),xlab="Time steps (in years)",
ylab="Transition probability from infected state",las=1)
lines(Prop_I_I[[Year]],type="l",col="red",lwd=3)
lines(Prop_I_D[[Year]],type="l",col="green",lwd=3)
abline(v=109, col="black",lty=3,lwd=.5)
text(110,.9,"Predictions from the nth-step transition matrix starting from year 2021",col="black",font=2)
```

Predicting the proportion of infected individuals who become susceptible, remain infected or die from COVID-19 over time from the estimated P^n transition matrix

```
Year=2020
almost_eql<-
bazar::almost.equal(round(Prop_I_S[[Year]],2),round(Prop_I_D[[Year]],2))
which(almost_eql==TRUE)
datafr=data.frame(Year=1:200,Prop_I_S=Prop_I_S[[Year]],
Prop_I_D=Prop_I_D[[Year]])
datafr[which(almost_eql==TRUE), ]
Year=2021
almost_eql<-
bazar::almost.equal(round(Prop_I_S[[Year]],2),round(Prop_I_D[[Year]],2))
which(almost_eql==TRUE)
datafr=data.frame(Year=1:200,Prop_I_S=Prop_I_S[[Year]],
Prop_I_D=Prop_I_D[[Year]])
datafr[which(almost_eql==TRUE), ]
```

Appendix B

5.5 R Codes for Time series Forecasting using ARIMA & Machine learning algorithms

Calling Covid-19 Data

```
daily_data=read.csv("Daily_Covid19Casestill19thOct.csv")
head(daily_data)
tail(daily_data)
#Loading packages library(xts) #library(zoo) library(forecast)
library("tsfknn")#knn
library("imputeTS") #time series imputation
library("tsfgrnn")# grnn
library("tsoutliers") #detecting and correcting outliers
library("seastests") #seasonaity test
library(FSA)
library(nortsTest)
library("car")
library(grid) # visualizations
library(ggplot2)
library(gridExtra)
library("nnfor")# time series with neural networks
library("astsa")
```

Calling Covid-19 Data

```
#declaring daily data
time_index<- seq(from = as.Date("2020-03-14"), to = as.Date("2021-10-18"), by =
"days")
#Cases <- zoo(daily_data$Cases,time_index)
Cases <- xts(daily_data$Cases, time_index)
print(paste("Data length=",length(time(Cases))))
periodicity(to.period(Cases,'days'))
tsplot(1:length(time_index),Cases, ylab="Daily COVID-19 Infection
Cases",col="black",xaxt="n",nxm=0,xlab="Date",lwd=1.7)
axis(side=1,at=seq(1,length(time_index),by=100),
labels=time_index[seq(1,length(time_index),by=100)])
```

Data Preprocessing (Outlier detection and data cleaning)

```
#No missing value
statsNA(Cases)
Cases_covid<- ts(daily_data$Cases,start = c(2020,
as.numeric(format(time_index[1],
"%j"))),
frequency = 365)


#print(Cases_covid)
#Detected outliers tsoutliers(Cases_covid)
#Outlier correction
Cases_covid_cleaned<- tsclean(Cases_covid)
time_index<- seq(from = as.Date("2020-03-14"), to = as.Date("2021-10-18"), by =
"days")
tsplot(1:length(time_index),Cases_covid, ylab="Daily COVID-19 Infection
Cases",col="black",xaxt="n",nxm=0,xlab="Date",lwd=.7)
#tsplot(Cases_covid, ylab="COVID-19 Cases",col="blue",xm=1)
axis(side=1,at=seq(1,length(time_index),by=100),
labels=time_index[seq(1,length(time_index),by=100)])
#axis(side, at=, labels=, pos=, lty=, col=, las=, tck=, ...)
lines(1:length(time_index),Cases_covid_cleaned,col="red",lwd=1.7)
legend('topleft', col=c("black","red"), lwd=3, legend=c("Original series", "Adjusted
series"), bg='white')
# ACF and PACF for determining seasonality
Corrected_data<- Cases_covid_cleaned
acf_plot<- ggAcf(Corrected_data,main="")+ theme_bw()+
theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
panel.background = element_blank())
```

Data Preprocessing (Outlier detection and data cleaning)

```
pacf_plot<- ggPacf(Corrected_data,main="")+ theme_bw()+
theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
panel.background = element_blank())
grid.arrange(acf_plot,pacf_plot,nrow=2)
#Testing for weekly seasonality
summary(wo(Corrected_data,freq=7)
#Testing for monthly seasonality
summary(wo(Corrected_data,freq=12))
#Testing for daily seasonality
summary(wo(Corrected_data,freq=365))
```

Aggregating data into weekly, monthly and quartely for seasonality plots

```
Data<-NULL
Week <- as.Date(cut(time_index, "week"))
Data[["weekly"]]<- aggregate(Corrected_data  Week, FUN=sum)
Month<- as.Date(cut(time_index, "month"))
Data[["monthly"]]<- aggregate(Corrected_data  Month, FUN=sum)
p1<- ggseasonplot(ts(Data[["weekly"]]$Corrected_data,frequency=52, start=c(2020,
as.numeric(format(Week[1], "%j")))),
year.labels=TRUE, year.labels.left=TRUE) +
ylab("Cases") +
ggtitle("Weekly seasonal plot")+ theme_bw()+
theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
panel.background = element_blank())+ theme(axis.text.x = element_text(angle =
300))

p2<-              ggseasonplot(ts(Data[["monthly"]]$Corrected_data,frequency=12,
start=c(2020,3,1)),
year.labels=TRUE, year.labels.left=TRUE) +
ylab("COVID-19 Infection Cases") +
ggtitle("Monthly seasonal plot")+ theme_bw()+
theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
panel.background = element_blank())
```

Aggregating data into weekly, monthly and quartely for seasonality plots

```
p2<- ggseasonplot(ts(Data[["monthly"]]$Corrected_data,frequency=12,
start=c(2020,3,1)),
year.labels=TRUE, year.labels.left=TRUE) +
ylab("COVID-19 Infection Cases") +
ggtitle("Monthly seasonal plot")+ theme_bw()+
theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
panel.background = element_blank())
p3<- ggseasonplot(Corrected_data,
year.labels=TRUE, year.labels.left=TRUE) +
ylab("COVID-19 Infection Cases") +
ggtitle("Daily seasonal plot")+ theme_bw()+
theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
panel.background = element_blank()) + theme(axis.ticks.x = element_blank(),
axis.text.x = element_blank())
grid.arrange(p3,p1,p2,nrow=3)
```

Fitting time ARIMA Model

```
horiz<-120
Observed_times<- time_index[1:(length(time_index)-horiz)]
#Computational time
time<- NULL

#Script of functions for rolling-origin evaluation for ARIMA model (forecasts)
setwd("C:/Users/user/Desktop/Pearl Covid19 Modelling")
source("ro-eval-ARIMA-script.r")
# 120-day forecast
options(warn=-1)
time0<- proc.time()
ARIMA_results_forecasts<- ro_eval_ARIMA_forecast(series=Corrected_data,
forecast_length=120,time_index=time_index,min_origin=90)
time[["ARIMA"]]<-proc.time()-time0
#Computational time
time[["ARIMA"]]
# 120-day rolling-origin evaluation with update and recalibration
#view first model
ARIMA_results_forecasts$ARIMA_model[[1]]
#Estimated Error at each forecast origin
dim(ARIMA_results_forecasts$ARIMA_errors)
head(ARIMA_results_forecasts$ARIMA_errors,n=15)
write.csv(ARIMA_results_forecasts$ARIMA_errors,
"ARIMA_results_forecasts_ARIMA_errors.csv")
```

Fitting time ARIMA Model

```
autoplot(ARIMA_results_forecasts$ARIMA_forecast[[1]])+
ggtitle("ARIMA model 120-day forecasts")+labs(y="Daily COVID-19 Infection Cases")+
theme_bw()+
theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
panel.background = element_blank())+
theme(plot.title = element_text(hjust=0), axis.text.x = element_text(size = 12)
,axis.text.y = element_text(size = 12), axis.title.x = element_text(size = 13),
axis.title.y = element_text(size = 13))+
theme(plot.title = element_text(face = "bold"))
```

Rolling-origin forecast evaluation for KNN models

```
#Script of the function to choose k and the lag value which results in minimum RMSE
#or better forecasts
source("KNN-k-lag-best-script.r")
#Script of the function for rolling-origin forecast evaluations for the KNN model
source("ro-eval-KNN-script.r")


# 120-day forecast
time0<- proc.time()
KNN_results_forecasts<-
ro_eval_KNN_forecast(series=Corrected_data,forecast_length=120,
time_index,min_origin=90)
time[["KNN"]]<-proc.time()-time0


#Computational time
time[["KNN"]]
forecast_origins<-rownames(KNN_results_forecasts$KNN_errors)
par(mfrow=c(3,3),mar=c(4,4,1,1))
k_values<- 52
lead_times<- 1:9
lwd_width<-3
for(i in lead_times){
plot(1:k_values,KNN_results_forecasts$k_best[[i]][1,],
type="l",xlab="K",xlim=c(1,k_values),
ylab="RMSE",col="blue",ylim=c(0,max(KNN_results_forecasts
$k_best[[i]])+3),lwd=lwd_width)
legend("bottom",c("lag 1","lag 2" ,"lag 3","lag 4","lag 5","lag 6"),
col=c("blue","red","green","yellow","black","brown"),bty="n",cex=1,box.lwd = 2,
fill=c("blue","red","green","yellow","black","brown"),horiz=F,ncol=2)
```

Rolling-origin forecast evaluation for KNN models

```
lines(1:k_values,KNN_results_forecasts$k_best[[i]][2,],
col="red",type="l",lwd=lwd_width)
lines(1:k_values,KNN_results_forecasts$k_best[[i]][3,],
col="green",type="l",lwd=lwd_width)
lines(1:k_values,KNN_results_forecasts$k_best[[i]][4,],
col="yellow",type="l",lwd=lwd_width)
lines(1:k_values,KNN_results_forecasts$k_best[[i]][5,],
col="black",type="l",lwd=lwd_width)
lines(1:k_values,KNN_results_forecasts$k_best[[i]][6,],
col="brown",type="l",lwd=lwd_width)

text(30,max(KNN_results_forecasts$k_best[[i]])+3,paste("Forecast
origin",":",forecast_origins[i]),col="black", font=2) }
# 120-day rolling-origin evaluation with update and recalibration
#view first model
#Estimated Error at each forecast origin
dim(KNN_results_forecasts$KNN_errors)
head(KNN_results_forecasts$KNN_errors,n=15)
write.csv(KNN_results_forecasts$KNN_errors,
"KNN_results_forecasts_KNN_errors.csv")
```

Rolling-origin forecast evaluation for KNN models

```
KNN1<-autoplot(KNN_results_forecasts$KNN_Model[[1]]
, highlight = "neighbors", faceting =FALSE)+
ggtitle("KNN model 120-day forecasts")+ theme_bw()+
theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
panel.background = element_blank(),legend.position=c(0.8, .8))+labs(y="Daily
COVID-19
Infection Cases")+
theme(plot.title = element_text(hjust=0), axis.text.x = element_text(size = 12)
,axis.text.y = element_text(size = 12), axis.title.x = element_text(size = 13),
axis.title.y = element_text(size = 13))+
theme(plot.title = element_text(size = 12, face = "bold"),
legend.title=element_text(size=12),
legend.text=element_text(size=12))
KNN1+ theme(plot.title = element_text(size=15))
```

Neural Network Time Series Forecasts

```
#script of functions for choosing the non-seasonal lag (p) and the average number of networks
#for the NNAR model forecasts based on mimimum RMSE
#source("NNAR-tuning-script.r")
#script of functions for rolling-origin forecast evaluations for the NNAR model
source("ro-eval-NNAR-script.r")
# 120-day forecast
time0<- proc.time()
NNAR_results_forecasts<- ro_eval_NNAR_forecast(series=Corrected_data, forecast_length=120,seed.number=1,time_index=time_index,min_origin=90)
time[["NNAR"]]<-proc.time()-time0
autoplot(NNAR_results_forecasts$NNAR_forecast[[1]])+
ggtitle("NNAR model 120-day forecasts")+labs(y="Daily COVID-19 Infection Cases")+
theme_bw()+
theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
panel.background = element_blank())+
theme(plot.title = element_text(hjust=0), axis.text.x = element_text(size = 12)
,axis.text.y = element_text(size = 12), axis.title.x = element_text(size = 13),
axis.title.y = element_text(size = 13))+
theme(plot.title = element_text(face = "bold"))
```

Generalized Regression Neural Networks

```
#script of functions for choosing the lags used as autoregressive variables
#for the GRNN model forecasts based on mimimum RMSE
#The smooting parameter is automatically chosen via optimisation
source("GRNN-lag-best-script.r")
#Script of the function for rolling-origin forecast evaluations for the GRNN model
source("ro-eval-GRNN-script.r")

#Script of function for GRNN autoplots (forecast)
source("autoplot-GRNN-script.r")
# 120 forecast
time0<- proc.time()
GRNN_results_forecasts<- ro_eval_GRNN_forecast(series=Corrected_data,
forecast_length=120,time_index=time_index,min_origin=90)
time[["GRNN"]]<-proc.time()-time0
# 120-day rolling-origin evaluation with update and recalibration
#view first model
#Estimated Error at each forecast origin
dim(GRNN_results_forecasts$GRNN_errors)
head(GRNN_results_forecasts$GRNN_errors,n=15)
write.csv(GRNN_results_forecasts$GRNN_errors,
"GRNN_results_forecasts_GRNN_errors.csv")
```

Generalized Regression Neural Networks

```
G1<-autoplot.grnnForecast(GRNN_results_forecasts$GRNN_Model[[1]],highlight
=
c("points"))+
ggtitle("GRNN model 120-day forecasts")+ theme_bw()+
theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
panel.background = element_blank())+labs(y="Daily COVID-19 Infection Cases")+
theme(plot.title = element_text(hjust=0), axis.text.x = element_text(size = 12)
,axis.text.y = element_text(size = 12), axis.title.x = element_text(size = 13),
axis.title.y = element_text(size = 13))

G1+theme(legend.position = c(0.8,
0.9),legend.title=element_text(size=13),legend.text=element_text(size=12))+
theme(plot.title = element_text(size=15))+theme(plot.title = element_text(face =
"bold"))
```

Multilayer Perceptron for time series forecasting

```
#Script of function for MLP plots (forecast)
source("plot-MLP-script.r")
##script of functions for rolling-origin forecast evaluations for the MLP model
# sel.lag=TRUE automatically select lags
#difforder=NULL automatically determines the differencing order
#hd.auto.type="cv" automatically selects the number of hidden nodes using 5-fold
cv (hd)
if hd=NULL
source("ro-eval-MLP-script.r")

# 120-day forecast
time0<- proc.time()
MLP_results_forecasts<-
ro_eval_MLP_forecast(series=Corrected_data,
forecast_length=120,seed.number=NULL,
time_index=time_index,min_origin=90)
time["MLP"]]<-proc.time()-time0
# 120-day rolling-origin evaluation with update and recalibration
#view first model
#Estimated Error at each forecast origin
dim(MLP_results_forecasts$MLP_errors)
head(MLP_results_forecasts$MLP_errors,n=15)
write.csv(MLP_results_forecasts$MLP_errors,
"MLP_results_forecasts_MLP_errors.csv")
```

Multilayer Perceptron for time series forecasting

```
autoplot(MLP_results_forecasts$MLP_forecasts[[1]])+
ggtitle("MLP model 120-day forecasts")+labs(y="Daily COVID-19 Infection Cases")+
theme_bw()+
theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
panel.background = element_blank())+
theme(plot.title = element_text(hjust=0), axis.text.x = element_text(size = 12)
,axis.text.y = element_text(size = 12), axis.title.x = element_text(size = 13),
axis.title.y = element_text(size = 13))+
theme(plot.title = element_text(face = "bold"))
```

Extreme learning machines (ELM) neural networks for time series forecasting

```
#Script of function for ELM plots (forecast)
source("plot-ELM-script.r")


##script of functions for rolling-origin forecast evaluations for the ELM model
source("ro-eval-ELM-script.r")
# 120-day forecast
time0<- proc.time()
ELM_results_forecasts<-
ro_eval_ELM_forecast(series=Corrected_data,
forecast_length=120,seed.number=NULL,
time_index=time_index,min_origin=90)
time[["ELM"]]<-proc.time()-time0


# 120-day rolling-origin evaluation with update and recalibration
#view first model
#Estimated Error at each forecast origin
dim(ELM_results_forecasts$ELM_errors)
head(ELM_results_forecasts$ELM_errors,n=15)
write.csv(ELM_results_forecasts$ELM_errors,
"ELM_results_forecasts_ELM_errors.csv")
```

Comparing computational times of the fitted time-series models

```
computational_time<- numeric(6)
models<- c("ARIMA","KNN","NNAR","GRNN","MLP","ELM")
for(k in seq_along(models)) computational_time[k]<- sum(time[[models[k]]][1:2])
plot(computational_time,type="b",lwd=3,ylab="Computational time (in secs)",
xlab="Fitted time-series models",
col="blue",xaxt="n",ylim=c(0,20000), pch=7)
axis(1, at=1:6, labels=models)

text(1,                    computational_time[1],round(computational_time[1],2),
pos=1,col="red",font=2)
text(2,                    computational_time[2],round(computational_time[2],2),
pos=1,col="red",font=2)
text(3,                    computational_time[3],round(computational_time[3],2),
pos=2,col="red",font=2)
text(4,                    computational_time[4],round(computational_time[4],2),
pos=2,col="red",font=2)
text(5,                    computational_time[5],round(computational_time[5],2),
pos=2,col="red",font=2)
text(6,                    computational_time[6],round(computational_time[6],2),
pos=1,col="red",font=2)
```

Finding the average RMSE for comparison between the time series models

```
rownames(ARIMA_results_forecasts$ARIMA_errors)
#Forecast errors
ARIMA_results_forecasts_ARIMA_errors
<-read.csv("ARIMA_results_forecasts_ARIMA_errors.csv")
KNN_results_forecasts_KNN_errors
<-read.csv("KNN_results_forecasts_KNN_errors.csv")
NNAR_results_forecasts_NNAR_errors
<-read.csv("NNAR_results_forecasts_NNAR_errors.csv")
GRNN_results_forecasts_GRNN_errors
<-read.csv("GRNN_results_forecasts_GRNN_errors.csv")
MLP_results_forecasts_MLP_errors
<-read.csv("MLP_results_forecasts_MLP_errors.csv")
ELM_results_forecasts_ELM_errors
<-read.csv("ELM_results_forecasts_ELM_errors.csv")
error_length<- length(ARIMA_results_forecasts_ARIMA_errors[,2])
error_length
max(MLP_results_forecasts_MLP_errors[,2])
# Comparing the forecast errors
par(mar=c(4,4,1,1))
plot(ARIMA_results_forecasts_ARIMA_errors[,2],type="b",col="red",lwd=3,
ylim=c(0,2000),xaxt="n",xlab="Forecast origins",ylab="RMSE",main="")
lines(KNN_results_forecasts_KNN_errors[,2],type="b",col="blue",lwd=3)
lines(NNAR_results_forecasts_NNAR_errors[,2],type="b",col="green",lwd=3)
lines(GRNN_results_forecasts_GRNN_errors[,2],type="b",col="brown",lwd=3)
lines(MLP_results_forecasts_MLP_errors[,2],type="b",col="magenta",lwd=3)
lines(ELM_results_forecasts_ELM_errors[,2],type="b",col="black",lwd=3)
```

Finding the average RMSE for comparison between the time series models

```
axis(1, at=1:31, labels=rownames(ARIMA_results_forecasts$ARIMA_errors))
mtext(text = "Rolling-origin forecast evaluation",
side = 3, #side 2 = left
line = -1,cex=1,font=2,adj =0.01)
text <- c("ARIMA","KNN","NNAR","GRNN","MLP","ELM")
colour<- c("red","blue","green","brown","magenta","black")

legend(x=2,y=2100,legend = text,col=colour,box.lwd = 2,fill=colour,bty="n",
,text.width = strwidth(text)[1]*0.7, cex=1,title="",horiz = T)
```

Finding the average RMSE for comparison between the time series models

```
#Function to compute the median MAPE across the time series models
Median_RMSE<-function(RMSE_lead_times,Methods){
median_RMSE<- apply(RMSE_lead_times,2,median)# median RMSE
return(data.frame(median_RMSE= median_RMSE,Models=Methods))
}
RMSEs_forecasts<-data.frame(ARIMA=ARIMA
_results_forecasts_ARIMA_errors[,2],
KNN=KNN_results_forecasts_KNN_errors[,2]
,NNAR=NNAR_results_forecasts_NNAR_errors[,2],
GRNN=GRNN_results_forecasts_GRNN_errors[,2],
MLP=MLP_results_forecasts_MLP_errors[,2],
ELM=ELM_results_forecasts_ELM_errors[,2])

RMSEs_forecasts$Lead_times<-ARIMA_results_forecasts_ARIMA_errors[,1]
RMSEs_forecasts
```