MINING THE BASIC REPRODUCTION NUMBER (R0) FORECAST FOR THE COVID OUTBREAK

By:
ILLAYAKANTTHAN A/L RAJOGOVAL
(Matric no: 143484)

Supervisor:
ASSOCIATE PROFESSOR DR. LOH WEI PING

JULY 2022

This dissertation is submitted to

Universiti Sains Malaysia

As partial fulfilment of the requirement to graduate with an honours degree in

BACHELOR OF ENGINEERING (MECHANICAL ENGINEERING)

School of Mechanical Engineering

Engineering Campus

Universiti Sains Malaysia

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ....

(ILLAYAKANTTHAN A/L RAJOGOVAL)

Date…………………………………………………………….…...(25/7/2022)

Statement 1:

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by giving explicit references. Bibliography/references are appended.

Signed ....

(ILLAYAKANTTHAN A/L RAJOGOVAL)

Date…………………………………………………………………...(25/7/2022)

Statement 2:

I hereby give consent for my thesis, if accepted, to be available for photocopying and for interlibrary loan, and for the title and summary to be made available outside organizations.

Signed ....

(ILLAYAKANTTHAN A/L RAJOGOVAL)

Date…………………………………………………………………...(25/7/2022)

ACKNOWLEDGEMENT

First and foremost, I wish to express my gratitude and praise to God, the Almighty, for his blessings, which have given me hope to finish my research successfully despite the challenges of the epidemic era the world is currently experiencing. Additionally, I would like to express my profound thanks to my supervisor, Associate Professor Dr Loh Wei Ping, for her unwavering support and guidance throughout this research, which encouraged me to conduct the study as effectively as possible. She never gives up on me and constantly supports me whenever I have questions or issues, despite the difficult times the world is experiencing because of the COVID-19 pandemic. I also like to thank the technical staff and the management at the School of Mechanical Engineering of Universiti Sains Malaysia (USM) for assisting me with all my tasks during these four years of study at the university. I will never forget to thank my family members, who have supported me unconditionally since I enrolled at this university. I owe them a huge debt of gratitude because, without their support, I would not be the person I am today. Finally, I want to express my sincere appreciation to my friends for continuing to inspire me throughout this research and my years at university. They have been a great support system for me during my university studies, and I am very thankful for that. Lastly, I convey my utmost appreciation to everyone for their ongoing concern and support for me.

TABLE OF CONTENT

ABSTRAK

Kaedah popular yang digunakan untuk menganalisis penyebaran COVID-19 adalah dengan mengira "Basic Reproduction Number, R_o". Angka ini ialah metrik yang digunakan secara global untuk menggambarkan penularan wabak COVID-19. secara global. Kajian sedia ada menggunakan beberapa model matematik termasuk SIR, SIRD, SEIR, SPIR, MCMC, Statistical exponential growth, statistical likelihood estimation and dynamic transmission models to menilai COVID-19 R_o. Walau bagaimanapun, tiada model yang tepat untuk menilai COVID-19 R_o. Selain itu, atribut penting untuk mendapatkan model ramalan terbaik untuk COVID-19 R_o masih tidak jelas. Oleh itu, tujuan kajian ini adalah untuk mengenal pasti dan menilai atribut dan parameter yang berkaitan dengan penilaian COVID-19 R_o, mengklasifikasikan data yang digunakan dalam model COVID-19 R_o sedia ada, membangunkan model pengelasan ramalan untuk COVID-19 R_o dan untuk menilai dan meningkatkan ketepatan COVID-19 R_o. Dua kajian kes berkaitan wabak COVID-19 di Malaysia dan Malta telah digunakan untuk kajian ini. Atribut kajian terutamanya mengenai kes harian, kematian, kemasukan ke hospital, kemasukan ke ICU dan isolasi. Waikato Environment for Knowledge Analysis (WEKA) versi 3.8 telah diterima pakai untuk analisis perlombongan data pada 2 peringkat klasifikasi. Prosedur pra-pemprosesan data memastikan nilai outlier dan ekstrem serta atribut yang tidak berkaitan telah dibuang. Untuk analisis klasifikasi, J48, Naïve Bayes, Random Forest dan SMO telah digunakan pada tiga pilihan ujian yang berbeza: set latihan penuh, pengesahan silang 10 kali ganda dan 66% split untuk menghasilkan peratusan ketepatan pengelasan. Ketepatan yang diperoleh untuk klasifikasi tahap pertama adalah antara 70.6349% hingga 100% untuk dataset Malaysia dan 89.6266% hingga 100% untuk dataset Malta. Manakala bagi klasifikasi tahap kedua, ketepatan adalah antara 67.8751% hingga 100% untuk dataset

Malaysia manakala 58.9212% hingga 100% untuk dataset Malta. Semua ketepatan klasifikasi yang diperoleh adalah melebihi ketepatan garis dasar. COVID-19 R_o model ramalan dibangunkan menggunakan algoritma klasifikasi regresi linear untuk meramalkan COVID-19 R_o berdasarkan COVID-19 R_o sebenar. Kajian ini mengenal pasti kes baharu, kematian, kemasukan ke hospital dan kemasukan ke ICU sebagai atribut penting untuk mendapatkan COVID-19 R_o yang tepat.

ABSTRACT

The popular method used to measure the spread of COVID-19 is by calculating the Basic Reproduction Number, R_o. This number is a globally used metric to describe the COVID-19 outbreak globally. Existing studies applied different models including SIR, SIRD, SEIR, SPIR, MCMC, Statistical exponential growth, statistical likelihood estimation and dynamic transmission models to evaluate COVID-19 R_o. However, there is no exact best model to ensure accurate Basic Reproduction Number, R_o. Besides, the essential attributes to return the best prediction model for COVID-19 R_o remains unclear. Therefore, this study aims to identify and evaluate the attributes and parameters associated with the development of the Basic Reproduction Number, R_o models, classify the data used in existing Basic Reproduction Number R_o models, develop a predictive classification model for the Basic Reproduction Number, R_o and to assess and enhance the accuracy of the Basic Reproduction Number, R_o prediction. Two case studies related to the COVID-19 outbreak in Malaysia and Malta were used. The study attributes are mainly about daily cases, deaths, hospitalization, ICU admission and Isolation. Waikato Environment for Knowledge Analysis (WEKA) version 3.8 was adopted for data mining analysis at two levels of classification stages. Data pre-processing procedures ensure outlier and extreme values and irrelevant attributes are discarded. For classification analysis, J48, Naïve Bayes, Random Forest and SMO were used on three different testing options: full training set, 10-fold cross validation and 66% split to yield the percentage of classification accuracy. Accuracies obtained for first-level classification ranged from 70.6349% to 100% for the Malaysia dataset and 89.6266% to 100% for the Malta dataset.

Meanwhile, for second-level classification, the accuracies ranged from 67.8751% to 100% for the Malaysia dataset while 58.9212% to 100% for the Malta dataset. All the classification accuracies obtained were above the baseline accuracy. The COVID-19 Basic Reproduction Number, R_o a predictive model is developed using a linear regression classification algorithm to predict the COVID-19 Basic Reproduction Number, R_o based on the actual COVID-19 Basic Reproduction Number, R_o. The study identified new cases, deaths, hospitalization, and ICU admission as important attributes to derive accurate COVID-19 Basic Reproduction Number, R_o

CHAPTER 1

INTRODUCTION

1.1 Overview

This chapter presents an introduction to a research study related to mining the Basic Reproduction Number, R_o forecast for the COVID-19 outbreak. This study is mainly aimed to identify the suitable attributes associated with the development of the COVID-19 Basic Reproduction Number, R_o or known as COVID-19 R_o mathematical model and to develop a suitable mathematical model to assess the accuracy of the Basic Reproduction Number, R_o. obtained for the forecast of the COVID-19 outbreak. The study background is presented in section 1.5.

1.2 Problem Statement

The COVID-19 Basic Reproduction Number, R_o or known as COVID-19 R_o It is a measure that was popularly used to reflect the COVID-19 outbreak. Nevertheless, the primary concern is the choice of different attributes and parameters considered in the equation and the appropriate mathematical model to determine the COVID-19 R_o. Choosing different attributes and parameters that are relevant could affect the accuracy of the COVID-19 R_o obtained. Therefore, it is important to assess the accuracy of the attributes and parameters used. Past studies have considered several attributes, parameters, and different models. But it is unclear which attributes could return more accurate results. Therefore, it is important to analyse the common attributes and parameters used to assess the COVID-19 outbreak and classify them by levels of significance to train the mathematical model for higher accuracy of COVID-19 R_o Besides past studies have evident different types of mathematical used to determine the COVID-19 R_o but there was no conclusion as to which model is the better version.

1.3   Project Objectives

This project is aimed to

   I.   Identify and evaluate the attributes and parameters associated with the development of Basic Reproduction Number, R_o models.

  II.   Classify the data used in the existing Basic Reproduction Number R_o models.

 III.   Develop a predictive classification model for the Basic Reproduction Number, R_o .

 IV.   Assess and enhance the accuracy of the Basic Reproduction Number, R_o prediction

1.4   Scope of Project

The study of this project involves the analysis of suitable parameters needed to assess the accuracy of the COVID-19 Basic Reproduction Number, R_O for the Covid-19 outbreak. The data analysis requires applying data mining techniques aided by the Waikato Environment for Knowledge Analysis (WEKA) tool. Two cases study data related to the COVID-19 outbreak will be retrieved from the publicly available domain databases. The data mining analysis will be structured into five main stages: Data Collection, Data Preprocessing, Data Classification, Verification and Interpretation and Evaluation. The data is initially pre-processed to remove any outliers, extreme values, and irrelevant data attributes. The data classification approach involves the segregation of data based on the level of accuracy of the Basic Reproduction Number, R_O from high to low using classification algorithms embedded in the WEKA tool. The classification algorithms used are J48, Sequential Minimal Optimization (SMO), Naïve Bayes, and Random Forest benchmarked on the ZeroR baseline algorithm. The classification mode is trained using a full training set, 10-fold cross validation and 66% split. Predicted COVID-19 Basic Reproduction Number, R_O values are verified based on the actual COVID-19 R_O values.

1.5   Research background

In December 2019, a series of pneumonia cases with clinical symptoms like viral pneumonia emerged in Wuhan, Hubei Province, China, with no identified cause (Huang et al., 2020; National Health Commission of the People's Republic of China, 2020; World Health Organization, 2020a). Most cases involved patients who worked or lived near the local Huanan Seafood Wholesale Market, selling live animals (Chan et al., 2020). This new virus was initially named 2019 novel coronavirus, which was subsequently referred to as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The disease is called coronavirus disease 2019 (COVID-19) by the World Health Organization (WHO). In January 2020, WHO confirmed COVID-19 as a sustained human to human disease (Li et al., 2020b). Due to contagiousness and morbidity, WHO declared this disease a global pandemic (Zhu et al., 2020). More than 2000 confirmed cases of COVID-19 had been found in China by January 26, 2020, primarily in Wuhan (C. Wang et al., 2020), indicating the virus's high transmissibility. By February 15, 2020, the number of confirmed cases in China had risen to 66 580, including 1524 deaths (Zhong et al., 2003). Moreover, COVID-19 had swiftly spread over the world by March 2020, with many new cases being reported daily.

In Malaysia, the first Covid-19 cases were reported on the 25[th of] January 2020, which traced back to three Chinese nationals who previously had close contact with an infected person in Singapore (Elengoe, 2020). Meanwhile, on the 4[th of] February 2020, the first Malaysian diagnosed with Covid-19 was reported. When the number of positive cases increased beyond 553 on the 16[th of] March 2020, the Prime Minister of Malaysia announced a Movement Control Order (MCO) to curb the spread of this contagious virus.

The COVID Basic Reproduction Number, R_o or known as COVID-19 R_O is the most important term in disease dynamics, and it is one of the methods used to measure the spread of this virus. This method is used by the World Health Organization (WHO) and most countries to measure the Covid-19 outbreak in a country and globally. It is used to quantify the transmission dynamics of an infectious disease. It is defined as the number of secondarily infected people produced by one infected individual introduced into a completely susceptible population. Suppose that infectious individuals make an average of β infection-producing contacts per unit time, with a mean contagious period of τ. Then the basic reproduction number is: $(R_o = \beta\tau)$. The computation of the Basic Reproduction Number, R_o is significant for implementing prevention measures due to the high infectiousness of SARS-CoV-2 among the vulnerable population. The average number of new cases created by an infected person is represented by this index (Liu et al., 2020; Delamater, 2020.). When R_O value is below one, and each infected individual infects less than one individual on average; therefore, the disease will cease to persist in the community. However, if R_O is greater than one, the disease persists and can spread. In other words, a high R_O implies a highly contagious infectious agent.

The essential factor in determining the COVID-19 R_O are the attributes and parameters used to calculate it. The data collected related to the Covid-19 outbreak will be classified into relevant attributes and parameters, which will be used to calculate the Basic Reproduction Number. Therefore, it is essential to organise and determine the suitable attributes and parameters that should be used to build the mathematical model to assess the accuracy of the Basic Reproduction Number, R_o for the forecast of the COVID-19 outbreak.

Besides, generating a suitable mathematical model also plays an important role in determining the accuracy of the Basic Reproduction Number, R_o. Thus, it is important to determine a suitable mathematical model so that an accurate COVID-19 R_O can be obtained.

1.6   Thesis Outline

This thesis is structured into five chapters.

Chapter 1 presents the introduction to the research study concerning the Coronavirus disease (COVID-19) outbreak and Basic Reproduction Number, R_o. This includes the discussion of the project background, objectives, problem statement scope of work and research background. The background of the COVID-19 disease and the Basic Reproduction Number, R_o is the focus of this chapter.

Chapter 2 discusses the state-of-the-art review of the COVID-19, attributes and parameters used in the evaluation of Basic Reproduction Number, R_o and the mathematical model developed to assess the COVID-19 R_O. The significant findings of the previous studies are emphasized.

Chapter 3 delivers the entire methodology and data mining processes involved. The breakdowns include data collection, data transformation, data pre-processing, data classification, and classification error performance analysis.

The findings of this study are described in Chapter 4. The attributes utilised to create the classification model in WEKA and the prediction accuracy attained by the predictive models are presented in tables and graphs. The categorization results are reviewed, analysed, and compared. The main attributes used to evaluate the COVID-19

R_O are retrieved, and the predictive classification model of COVID-19 R_O was developed.

Lastly, Chapter 5 concludes the study's achievements and novel findings. Study contributions, how this study's objectives are met, and the future directions of the study are discussed.

CHAPTER 2

LITERATURE REVIEW

2.1   Overview

COVID-19, a disease caused by the SARS-CoV-2 infection, has affected many people's lives. Computing the COVID-19 Basic Reproduction Number, R_o have been one of the methods used by the World Health Organization (WHO) and most countries to measure the COVID-19 outbreak. Many research studies were carried out to study the accuracy of the COVID-19 R_o in assessing the COVID-19 outbreak. Based on the past research studies, different attributes were considered in the equation and the mathematical model used for developing the COVID-19 R_o. There were altogether 53 related works retrieved from Science Direct and Google Scholar databases from 2019 to 2022.

2.2   Search Strategy

International databases such as Science Direct and Google Scholar were accessed to search for eligible articles related to the COVID-19 Basic Reproduction Number, R_0. The articles were searched using the following keywords: "COVID-19" and "COVID-19 Basic Reproduction Number". The article publication period was taken from January 2020 to April 2022. This is because COVID-19 was only discovered in December 2019. Thus, the article related to the COVID-19 outbreak only began as early as January 2020. We identified 47,513 articles from Science Direct and 674,000 from Google Scholar databases (Figure 2.1). From these numbers, 721,417 have been identified as irrelevant papers for two reasons. Some research papers were duplicates (appear in both the databases). Several papers retrieved from the database do not meet the present study's scope. Thus, these irrelevant papers were removed and left behind

7

96 research articles for further eligibility assessments. From the 96 accesses, approximately 31 articles were found suitable and selected for review. Aspects that contribute to good estimates of COVID-19 R_o were reviewed. This includes the relevant attributes that were considered in the equation and the type of mathematical model used to assess the accuracy of COVID-19 R_o
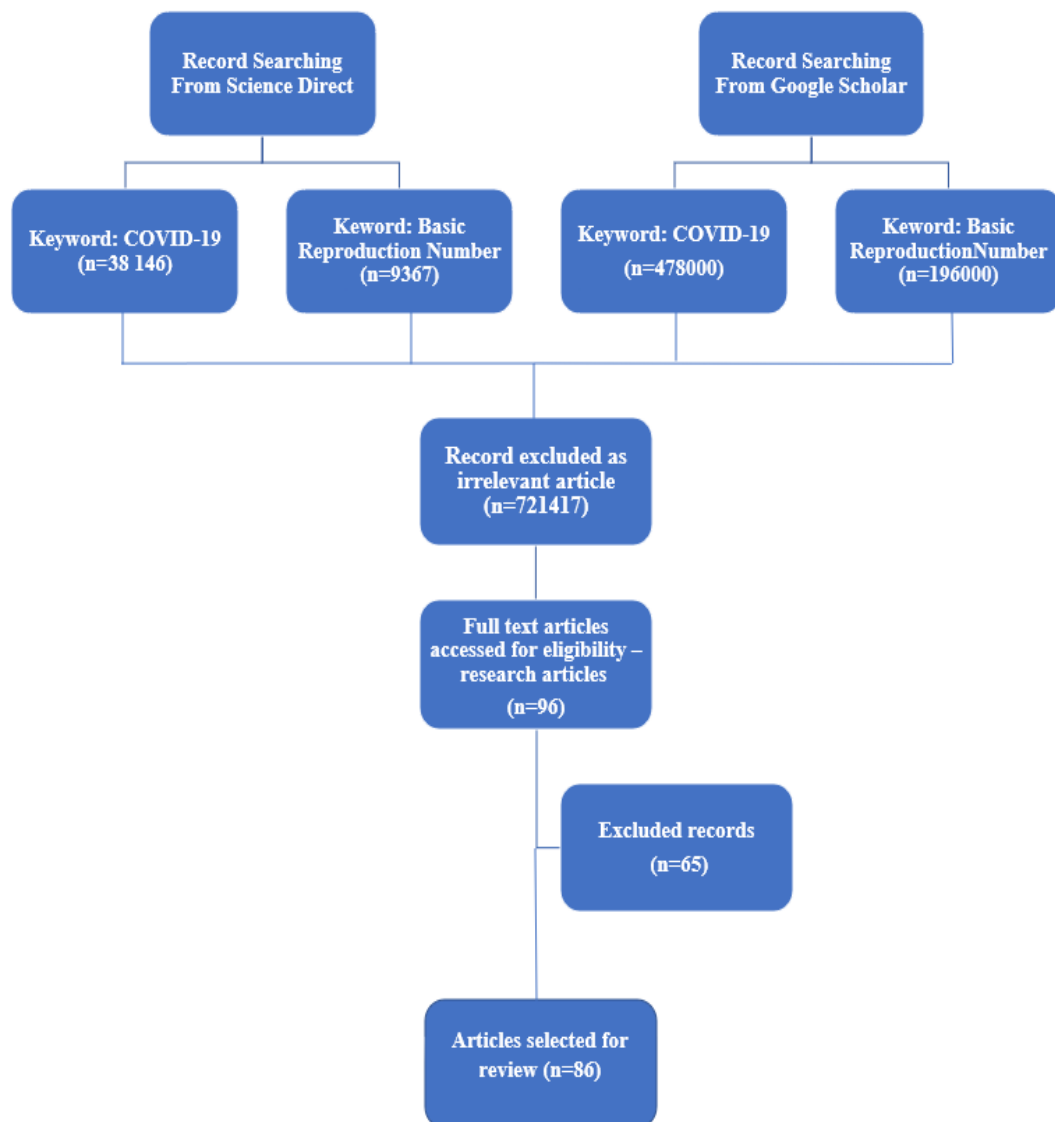

Figure 2.1: Schematic diagram of article search from Science Direct and Google Scholar engines

2.3 COVID-19

COVID-19 is a zoonotic disease in which an animal virus mutates to allow it to enter and replicate inside the human body, quickly spreading throughout the population. Asymptomatic and symptomatic infected people spread the virus through oral fluid droplets, transmitted through the air via coughing or sneezing (Adhikari et al., 2020; Singhal, 2019; Yang & Wang, 2020). Fever, cough, and exhaustion were reported as the symptoms of those infected with this virus, as well as sore throat, diarrhoea, headache, nausea, myalgia, chest pain, and loss of smell and taste (Lu, 2020). The genetic sequence of the 2019 novel coronavirus (2019-nCoV) allowed for the rapid development of 2019-nCoV-specific point-of-care real-time RT-PCR diagnostic tests (C. Wang et al., 2020). These RT-PCR diagnostic tests have been used to rectify whether people who are suspected of the COVID-19 disease.

2.4 Importance of R_o for COVID-19 outbreak

The COVID-19 R_o is the number of people directly infected by an infected person in a community. In other words, susceptibility is likely the most often used indicator of how severe an epidemic outbreak can be (Holme & Masuda, 2015). COVID-19 R_o have been used as an indicator globally to study the COVID-19 transmissibility rate. COVID-19 R_o evaluates whether the disease may persist in each population(Knobel et al., 2020). It represents the level of mitigation required to bring an epidemic under control. To avoid subsequent transmissions, mitigation strategies include rapid case identification, quarantine, and physical distance(Chaudhry, 2022). The higher the COVID-19 R_o , the more likely the disease will be rapidly spread. According to (Healthline, 2020.), three different possibilities are conveyed by the COVID-19 R_o:

1) The disease will not spread and will eventually die out if $R_o < 1$.

2) The disease will be stable in the community but will not spread to other people if

$R_o = 1$.

3) The disease will spread and may result in an epidemic if $R_o > 1$, .

According to (Sanche et al., 2020), the COVID-19 R_o was previously expected to be between 2.2 and 2.7. However, based on data from case reports across China, COVID-19 R_o is substantially higher. Early in the pandemic in Wuhan, the doubling time was 2.3 to 3.3 days, according to the studies. Researchers obtained a median COVID-19 R_o a score of 5.7 using this data. This indicates that each individual infected with the virus can transmit 5 to 6 people rather than the previously anticipated 2 to 3 people.

2.5   Attributes related to the COVID-19 outbreak

There are many attributes involved in the outbreak of the COVID-19 outbreak. The type of attributes used to evaluate the COVID-19 R_o plays an essential role in determining the accuracy of the R_o value obtained. Based on the past research studies, different attributes have been considered in the equation and the mathematical model to evaluate the COVID-19 R_o (Alimohamadi et al., 2020b) . For instance, (Movahedi et al. 2021) have conducted a study on the estimation of the COVID-19 R_o for the COVID-19 pandemic in Minnesota. They have focused on the COVID-19 R_o for the COVID-19 pandemic in Minnesota. The attributes used in (Movahedi et al. 2021) are the number of infections, hospitalizations, ICU admissions, number of recovery and deaths. The data is from a 6-month data set containing Minnesota data on the COVID-19 outbreak and is retrieved from the Minnesota

Department of Health (Minnesota Department of Health, 2020. Situation Update for COVID-19., 2020.).

Besides, (Dharmaratne et al., 2020) have studied the estimation of the COVID-19 R_o for the novel coronavirus disease in Sri Lanka. Based on their study, daily situation reports from the Epidemiology Unit, Ministry of Health, Sri Lanka, have been used to extract the total number of confirmed COVID-19 patients who tested positive with the reverse transcriptase polymerase chain reaction. Furthermore,(D'Arienzo & Coniglio, 2020) on the assessment of the SARS-CoV-2 COVID-19 R_o based on the early phase of the COVID-19 outbreak in Italy, have focused on attributes such as susceptible but not yet infected with the disease, number of infectious individuals, and individuals who have recovered from the disease and now have immunity to it.

Studies were also carried out to evaluate the effect of different rates of cases, clinical studies, and global health concerns on COVID-19 R_o. For example, Al-Raeei (2020) have carried out a study on the COVID-19 Basic Reproduction Number of the new coronavirus pandemic with mortality for India, the Syrian Arab Republic, the United States, Yemen, China, France, Nigeria, and Russia with different rate of cases, Clinical Epidemiology and Global Health. The attributes that were considered in their study are susceptible cases, infected cases, recovered cases and death cases.

The research was also carried out to study the spread of COVID-19 transmission in a specific community. For instance, Puglisi et al. (2021) have conducted research on the estimation of the COVID-19 Basic Reproduction ratio in a large urban jail in the United States. The attributes considered in this study are susceptible individuals due to contact with infected symptomatic or asymptomatic individuals and infected individuals. Moreover, Zhang et al. (2020) have conducted a study on estimating the

reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship. They studied the transmissibility of COVID-19 on the ship based on the "earlyR" package to estimate the R0 in the early outbreak stage (Polonsky et al., 2019). COVID-19 R_o estimate requires serial interval distribution, but there was insufficient information about cluster cases to perform serial interval estimation (Musa et al., 2021).

Studies were also carried out on COVID-19 R_o of the COVID-19outbreak when criteria like safety measures were taken to curb the spread of COVID-19. (Ahmad Alajlan et al., 2021) have carried out a study on the impact of lockdown strategies on the basic reproductive number of coronavirus (COVID-19) cases in Saudi Arabia. The COVID-19 R_o was estimated daily using data from the Saudi Ministry of Health's (MOH) daily reporting on the COVID-19 epidemic in Saudi Arabia (*Ministry of Health 2020*, n.d.). The attributes focused on in their paper are susceptible cases and confirmed cases, including patients having at least one of the following symptoms: headache, sore throat, diarrhoea, nausea, or diarrhoea, as well as laboratory confirmation of COVID-19 infection. Apart from that, studies on the influence of mask usage on the spread of COVID-19 during the pandemic in New York City were carried out by Ma et al. (2022). In their study, the focused attributes are the coverage of masks used by the public, which can affect the transmission rate of the COVID-19 disease.

Environmental conditions also play a role in affecting the COVID-19 R_o of the COVID-19 outbreak. For example, Milicevic et al. (2021)studied PM2.5 as a major predictor of COVID-19 Basic Reproduction Number in the USA, Environmental Research. The authors presented that pollution (mainly PM2.5 and PM10 and NO2)

significantly impacts COVID-19 transmissibility. It was believed that the droplets containing virus particles bind to Particulate Matter (PM), enhancing virus droplet diffusion in the air (Chen et al., 2010; Comunian et al., 2020; Contini & Costabile, 2020). The attributes that were focused on in their study are susceptible cases, exposed cases, infected cases, recovered cases and death cases. Following the same idea, de Angelis et al. (2021) studied the effects of pollution on COVID-19 incidence and mortality in Lombardy, Italy. In their study, an ecological study assesses the association between long-term exposure to particulate matter (PM) and nitrogen dioxide (NO2) on COVID-19 incidence and all-cause mortality after accounting for demographic, socioeconomic and meteorological variables.

2.6 Parameters used to model the COVID-19 R_o

The parameters considered to model Basic Reproduction Number, R_o of the COVID-19 outbreak were studied. Different studies have used other parameters in their mathematical models to evaluate the COVID-19 Basic Reproduction Number, R_o as summarized in Table 2.1.

Most studies have focused on the transmission or infection rate, which is the probability that an infection may spread among susceptible people within a group (Y. Liu, Eggo, et al., 2020). In the context of COVID-19, the transmission is mainly due to droplet and airborne transmission (World Health Organization, 2020b). Infected individuals can exhale particles and droplets of respiratory fluids containing the SARS CoV-2 virus into the air. A person can become exposed to COVID-19 while near an infected person coughing or speaking, as COVID-19 is spread by contact with respiratory fluids containing the contagious SARS-CoV-2 virus. Aerosol particles spreading away from the diseased person can also expose

them (Centers for Disease Control and Prevention, 2021b). Additionally, exposure can occur when respiratory fluids are sprayed or splashed directly onto someone's mucous membranes (United States Environmental Protection Agency (EPA), 2021).

Besides, asymptomatic and symptomatic rates, known as the probability of exposed individuals infected but not yet infectious and transitioning to either symptomatic or asymptomatic infection, respectively, are considered (Puglisi et al., 2021). People may exhibit a wide range of symptoms in COVID-19 symptomatic cases. Symptoms usually develop two to 14 days after contact, according to the Centers for Disease Control (CDC). Common symptoms include fever, coughing, and trouble breathing (Centers for Disease Control and Prevention, 2022a). In asymptomatic, they don't exhibit any disease symptoms in their body. According to the CDC, asymptomatic COVID-19 infections make up 35% of all cases (Centers for Disease Control and Prevention, 2021). People are asymptomatic when they are infected and are unaware of it.

The positive and negative testing rate is the ratio of either positive or negative individuals to the number of individuals tested (Hasan & Nasution, 2022b). These aspects are essential for individuals to isolate themselves, lowering the risk of spreading the disease to others and enabling them to seek treatment earlier, likely reducing the severity of their illness and the risk of long-term disability or death (National Institute on Aging, 2020).

The social distancing rate refers to the probability of individuals that adhere to social distancing (X. Ma et al., 2022). This parameter reflects the probability of individuals adhering to a community's social distancing measure.

Social distancing is one of the most effective measures to reduce the spread of the virus, which is transmitted by air droplets (Qian & Jiang, 2022).

The hospitalization rate is the number of people hospitalised in a particular location due to a positive COVID-19 laboratory test over the total population of that defined area (Centers for Disease Control and Prevention, 2020). The COVID-19 pandemic disrupted hospital care as hospitals had to deal with a highly infectious virus while also fulfilling the ongoing needs of their communities for health services. In Malaysia, the peak rate is 14753 individuals, recorded on 16 August 2021 (Our World in Data, 2021).

The recovery rate parameter gives the number of infected individuals recovered from COVID-19 to the number of recovery days (Sigfrid et al., 2021). The recovery rate plays a crucial role in assessing how well a country is doing at controlling the coronavirus. It depends on various variables, including the death rate, hospitalisation rates for cases, the quality of care, and discharge criteria (A. Singh & Chattopadhyay, 2021).

The quarantine rate parameter is the number of individuals quarantined due to close contact, return from travelling and confirmed cases during the quarantine period (Cromer et al., 2021). This parameter is important because quarantine reduces the ongoing transmission of the disease in two ways: preventing transmission before symptom onset and reducing the overall transmission from persistently asymptomatic individuals (Ashcroft et al., 2021).

ICU rate is the time lag between infection and ICU admission as well as the length of stay in ICU (Ritter et al., 2021.). This parameter is important in studying

the seriousness of the disease and how it affects other attributes, such as changing in-hospital mortality rates over time after accounting for individual patient characteristics and the risk of disease transmission

Other studies also considered the death rate parameter. This parameter describes the number of deaths due to COVID-19 during a specific period in the total population (Jacob & Ganguli, 2016). This parameter is important to study how contagious the spread of the disease is in a selected community and to analyse the impact of increasing death rates on the health system.

The population size parameter is the number of people in a population divided into distinct size classes (Benke & Huryn, 2017). The population size is essential in many research studies as it is a factor in the spread of COVID-19. It is an important indicator to draft standard operating procedures and execute nontherapeutic interventions to control the epidemic. It is also important to evaluate the risk of transmission of the disease and the seriousness of the condition in a community (Ganasegeran et al., 2021).

The rate of transmission from susceptible to infectious individuals parameter is the probability of a susceptible individual being contacted with the disease (Cooper et al., 2020). This parameter is essential so that measures can be taken to treat the susceptible individual in the early stages before the disease spreads and becomes severe.

Next, the recovery rate of asymptomatic infected individuals parameter is the number of asymptomatic infected individuals recovered from COVID-19 to the number of recovery days (Subramanian et al., 2021). Meanwhile, the recovery rate

of symptomatic infected individuals is the number of symptomatic infected individuals recovered from COVID-19 to the number of recovery days (Otunuga, 2021). Both asymptomatic and symptomatic recovery complement each other for an individual's recovery from COVID-19.

The rate of asymptomatic infected people becoming infected symptomatic is the total number of asymptomatic infected people who become infected symptomatic to the number of asymptomatic infected people (Q. Ma et al., 2021).

Besides, the recovery rate of quarantined infected individuals parameter is the number of quarantined infected individuals recovered from COVID-19 to the number of recovery days (Tang, Xia, et al., 2020). Similarly, the recovery rate of hospitalized individuals is the number of hospitalized infected individuals recovered from COVID-19 to the number of recovery days (Medical Association, 2020). Past studies have shown that the recovery rate could take 5 to 14 days to recover from the disease, depending on the severity and stage of infection (Centers for Disease Control and Prevention, 2022b).

Finally, the rate of recovered people becoming susceptible is the probability of recovered people becoming susceptible again to disease (Gou & Jin, 2017). This parameter is essential to study the rate and cause of individuals who recovered from the disease to be infected.

Table 2.1: Parameters used to evaluate the COVID-19 R_o

Parameters	Description	Article References
Transmission/Infection rate	the probability that an infection may spread among	(Ahmad Alajlan et al., 2021; Al-Raeei, 2021; Anggriani et al., 2022;

	susceptible people within a group	D'Arienzo & Coniglio, 2020; Dharmaratne et al., 2020; Hasan & Nasution, 2022b; X. Ma et al., 2022; Milicevic et al., 2021; Movahedi et al., 2021; Puglisi et al., 2021; Tang, Bragazzi, et al., 2020)
Asymptomatic rate	probability of exposed individuals infected but not yet infectious and transition to asymptomatic infection, respectively, were considered	(Puglisi et al., 2021)
Negative testing rate	the ratio of negative individuals to the number of individuals that were tested	(Hasan & Nasution, 2022b)
Positive testing rate	the ratio of positive individuals to the number of individuals that were tested	(Hasan & Nasution, 2022b)
Social distancing rate	probability of individuals that adhere to social distancing	(X. Ma et al., 2022)
Symptomatic rate	probability of exposed individuals infected but not yet infectious and transition to symptomatic infection respectively were considered	(Puglisi et al., 2021)
Hospitalization rate	number of people hospitalised in a particular location due to a positive COVID-19 laboratory test over the total population of that defined area	(Movahedi et al., 2021)

Recovery Rate	number of infected individuals recovered from COVID-19 to the number of recovery days	(Al-Raeei, 2021; D'Arienzo & Coniglio, 2020; Dharmaratne et al., 2020; Hasan & Nasution, 2022b; Movahedi et al., 2021; Puglisi et al., 2021; Shringi et al., 2021)
Quarantine rate	the number of individuals that were quarantined due to close contact return from travelling and confirmed cases during the quarantine period	(Ahmad Alajlan et al., 2021)
ICU Rate	the time lag between infection and ICU admission, as well as the length of stay in ICU	(Movahedi et al., 2021)
Mortality/Death Rate	number of deaths due to COVID-19 during a specific period to the total population during the period	(Al-Raeei, 2021; Hasan & Nasution, 2022b; Movahedi et al., 2021; Shringi et al., 2021)
Population Size	number of people in a population divided into distinct size classes	(Al-Raeei, 2021; D'Arienzo & Coniglio, 2020; Dharmaratne et al., 2020; Hasan & Nasution, 2022b; Movahedi et al., 2021; Puglisi et al., 2021; Shringi et al., 2021)
Rate of transmission from susceptible to infectious individuals	probability of a susceptible individual to be contacted with the disease	(Gebremeskel et al., 2021)
The recovery rate of asymptomatic infected individuals	number of asymptomatic infected individuals	(Anggriani et al., 2022; Gebremeskel et al., 2021;

	recovered from COVID-19 to the number of recovery days	X. Ma et al., 2022; Tang, Wang, et al., 2020)
Recovery rate of symptomatic infected individuals	number of symptomatic infected individuals recovered from COVID-19 to the number of recovery days	(Anggriani et al., 2022; Gebremeskel et al., 2021; X. Ma et al., 2022; Tang, Wang, et al., 2020)
Recovery rate of quarantined infected individuals	number of quarantined infected individuals recovered from COVID-19 to the number of recovery days	(Anggriani et al., 2022; Tang, Wang, et al., 2020)
Recovery rate of hospitalized individuals	number of hospitalized individuals recovered from COVID-19 to the number of recovery days	(X. Ma et al., 2022)
Rate of asymptomatic infected people become infected symptomatic	total number of asymptomatic infected people become infected symptomatic to the number of total numbers of asymptomatic infected people	(Anggriani et al., 2022; Inayaturohmat et al., 2022)
Rate of recovered people become susceptible	probability of recovered people becoming susceptible again to being infected by the disease	(Anggriani et al., 2022)

2.7 Mathematical Model to evaluate the COVID-19 R_o

Many mathematical models have been developed to evaluate the COVID-19 R_o. The models derived focused on the important aspects of a disease, identifying threshold values for disease survival, and assessing the influence of certain control strategies. Accordingly, mathematical modelling can significantly quantify possible disease control strategies (Heffernan et al., 2005; van den Driessche, 2017).

To evaluate the COVID-19 R_o Mathematical models used are the SIR model, also known as a compartmental model, MCMC model, statistical exponential growth model, statistical maximum likelihood estimation model and dynamic transmission model. The majority of studies reported the use of compartmental models for evaluating the COVID-19 R_o while the minority applied the MCMC model (Table 2.2), figure 2.2 shows the percentage of studies that used the specific models.

The SIR Model represents the population as a single system with three functions: Susceptible, Infected, and Removed (Lord, 2021). The basic SIR model consists of variations: SIRD, SEIR and SPIR.

SIRD model: This model considers the number of persons who are Susceptible, Infected, Recovered, and Deceased. According to this model, a susceptible individual who contacts an infected person likely to get infected (Anastassopoulou et al., 2019; Wang & Jia, 2019; Yuan et al., 2020). An infected person might either recover or die because of the disease. As a result, the total of S(t), I(t), R(t), and D(t) are assumed to be constant in this model. Furthermore, it is considered that everyone who is exposed to the virus becomes infected immediately, with no latent period between exposure and infection. The disadvantage of this model is that there is no consideration of the effects of confinement or quarantine (Sen & Sen, 2021).

SEIR model: The model divides the total human population size at time t (N(t)) into Susceptible S(t), Exposed E(t), Asymptomatic Infectious IA(t), Symptomatic Infectious IS(t), and recovered R(t) (Mwalili et al., n.d.). The total number of people is given by N(t)=S(t)+E(t)+I+R(t). The SEIR model's basic hypothesis is that, over time, all the individuals in the model will be classified based on these four parameters (Yuexi & Sun, 2021.). The advantage of the SEIR model is that it is one of the most widely

used mathematical models used to describe the dynamics of an epidemic and forecast potential transmission scenarios. Since the infectious disease epidemic, the SEIR model can be used to evaluate the efficacy of various measures, such as lock-down (Godio et al., 2020). However, the disadvantage of the SEIR model is that it is complex and requires assumptions to be made before modelling (Harjule et al., 2021).

SPIR model: This model differs from the rest by denoting the Susceptible case, Probable case, Infectious case, and Recovered case, respectively. The system satisfies $S(t) + P(t) + I(t) + R(t) = N(t)$, which expresses the population constancy mathematically (Hasan & Nasution, 2022a). The advantage of the SPIR model is that it is easy to use and can forecast potential transmission scenarios. The difference between the SPIR model and the SEIR model is that in the SEIR model, the exposed case parameter is used, while in the SPIR model, the probable case parameter is used.

Apart from SIR (compartment model), the statistical exponential growth model is another standard model used in which the population constantly grows over time. Depending on whether reproduction is assumed to be continuous or periodic, the relationship can be represented in either way. Statistical exponential growth produces a constant curve of rising or decrease, the slope of which varies by direct proportion to population size (B. P. Singh, 2021). The advantage of using a statistical exponential growth model in evaluating the COVID-19 R_o naturally exhibit exponential behaviour in the early stages of an outbreak, when the number of infections is already substantial, but recoveries and deaths are still negligible (Bertozzi et al., 2020).

On the other hand, the Bayesian Markov chain Monte Carlo (MCMC) model includes Poisson autoregressive as a function of both a short-term and a long-term dependency (Thong et al., 2019). methods comprise a class of algorithms for sampling

from a probability distribution. The advantage of MCMC is that only the likelihood function with the unobserved variables must be considered. This suggests that Bayesian parameter estimate is typically faster than maximum likelihood estimation (Paap, 2002). However, the disadvantage of this model is that evaluating the accuracy and convergence can be highly difficult.

The statistical maximum likelihood estimation model is a statistical method to estimate parameters by maximizing their posterior probability densities based on the deep-rooted Bayesian formalism (Gao et al., 2017). It simplifies the problem of parameter estimation as a problem of maximizing a log-likelihood function (Gao et al., 2018). The advantage of the statistical maximum likelihood estimation model is a consistent approach to parameter estimation problems is provided by maximum likelihood. As a result, estimates using the maximum likelihood method can be generated for a wide range of estimation situations. However, the disadvantage is that it can be heavily biased for small samples. The optimality properties may not apply to small samples.

On the other hand, the dynamic transmission model reproduces the direct and indirect impacts that a communicable disease control strategy may have (Pitman et al., 2012). This model describes how the likelihood that vulnerable hosts will get infected depends on the existing (and typically changing over time) prevalence of infectious people (or vectors or fomites) and how this has an impact on the spread of infection and the (cost-) efficacy of interventions (Cohen & White, 2016). The advantage of the dynamic transmission model is that it serves as a simplified representation of more complex systems compared to purely statistical models.

Table 2.2 shows the type of mathematical models that were used in each of the articles obtained to evaluate the COVID-19 R_0. The table consists of 2 columns: the mathematical model column, which contains the list of mathematical models, and the paper column, which contains the paper that used the respective mathematical model.

Table 2.2: Type of mathematical model used to evaluate the COVID-19 R_o

Mathematical Model	Paper
SIR Model (Compartmental Model)	(Ahmad Alajlan et al., 2021; D'Arienzo & Coniglio, 2020; Dharmaratne et al., 2020)
SIRD Model (Compartmental Model)	(Al-Raeei, 2020; Shringi et al., 2021)
SEIR Model (Compartmental Model)	(X. Ma et al., 2022; Milicevic et al., 2021; Movahedi et al., 2021; Puglisi et al., 2021; Tang, Wang, et al., 2020; Zhou et al., 2020)
SPIR Model (Compartmental Model)	(Hasan & Nasution, 2022b)
Markov chain Monte Carlo Model (MCMC)	(Dushoff et al., 2020; Jung et al., n.d.; Wu et al., 2020)
Statistical exponential growth model	(Du et al., 2020; Jung et al., n.d.; Lai et al., 2020; Li et al., 2020a; T. Liu et al., n.d.; Muniz-Rodriguez et al., 2020; Sanche et al., 2020; Z. Wang et al., 2021; Zhao et al., 2020a)
Statistical maximum likelihood estimation model	(Alimohamadi et al., 2020a; Li et al., 2020c; Tang, Bragazzi, et al., 2020; S. Zhang et al., 2020)
Dynamic transmission model	(Giordano et al., 2020; Puglisi et al., 2021)