Predicting the Stagnation Time of Covid-19 Pandemic Using Bass Diffusion and Mini-Batch Gradient Descent Models

Mohamed Atef Mosa,
Department of Information Technology, Institute of Public Administration, Riyadh, KSA
Egyptian Space Agency, Assembly, Integration & Testing Center, Cairo, Egypt.
mosamo@ipa.edu.sa, mohammedatefmosa@gmail.com

Abstract: The coronavirus (SARS-CoV-2), which first appeared in Wuhan, China, in December of 2019, spread quickly around the world, eventually categorizing it as a global "Epidemic". In early 2020, the SARS-CoV-2 emerging virus had devastating effects on all aspects of daily life, public health, and even the global economy. During that epidemic, much effort had been made to predict the number of confirmed and deaths, and when the epidemic would subside. However, the prediction of epidemic indications (COVID-19) was highly uncertain and different from what happened next. Multiple and rapid virus mutations, and late detection of infection in many cases of people, have made the prediction process complicated and difficult, with some of the proposed models appearing to be largely misleading. In this research paper, we reviewed the analytical and statistical methods to extrapolate the most important data and indicators about the infection (COVID-19) and the rate of confirmed, recovery, and deaths during the past few months in some countries of the world, especially in the Kingdom of Saudi Arabia. On the other hand, we proposed the time for the infection to subside in the Kingdom of Saudi Arabia and some other countries. In the proposed prediction model, the Bass diffusion model was adopted by combining with the mini-batch Gradient descent algorithm to obtain the optimum values for the Bass algorithm parameters. The model was trained on about 85% of the available historical data and tested on the rest of the data. The proposed model indicated that the Kingdom of Saudi Arabia will face an increase in the coming days in terms of the high number of confirmed cases. Moreover, the rate of increase in injuries will decrease over time until it reaches its lowest levels in January of the next year. The model also showed that the curved flattening point for confirmed figures will be at the mentioned month, which is the expected date for the epidemic to recede in Saudi Arabia in the absence of other aftershocks.

Keywords: Data analysis; prediction; COVID-19; Bass diffusion model; mini-batch gradient descent.

توقع زمن الركود لجائحة كورونا (كوفيد -١٩) باستخدام نماذج انتشار باس ونماذج نزول التدرج ذات الدفعة المصغرة

الملخص: انتشر الفيروس التاجي، فيروس كرونا المستجد (SARS-CoV-2)، والذي ظهر لأول مرة في مدينة ووهان الصينية في ديسمبر من العام ٢٠١٩، لكن سرعان ما توغل الي جميع أنحاء العالم ليصنف في نهاية الأمر على أنه "جائحة" (Epidemic) عالمية. في مطلع العام ٢٠٢٠، كان لفيروس كرونا المستجد SARS-CoV-2 تأثيرات مدمرة على كل جوانب الحياة اليومية والصحية، بل وحتى الاقتصاد العالمي. خلال تلك الجائحة بُذلت جهود كثيرة للتنبؤ بأعداد الإصابات والوفيات وموعد انحسار الوباء. ومع ذلك، فإن التنبؤ بمؤشرات عدوي (COVID-19) كانت غير مؤكدة بدرجة كبيرة ومغايرة لما حدث بعد ذلك. طفرات الفيروس المتعددة والسريعة، والاكتشافات المتأخرة للعدوي في حالات كثيرة، جعلت من عملية التنبؤ أمرا معقدا وعسيرا، بحيث ظهرت بعض النماذج المقترحة على أنها مضللة. في هذه الورقة البحثية استعرضنا بعد الطرق التحليلية والإحصائية لاستقراء أهم البيانات والمؤشرات حول عدوي (COVID-19) ومعدل الإصابات والتعافي والوفيات خلال الأشهر القليلة الماضية في بعض دول العالم وبالأخص في المملكة العربية السعودية. من ناحية أخري قمنا بتسليط الضوء على إمكانية التنبؤ بمعدلات انتشار عدوي (COVID-19) في المملكة العربية السعودية وبعض الدول الأخرى في الأيام القادمة والتنبؤ بموعد انحسار هذا الوباء. تم الاعتماد في نموذج التنبؤ المقترح على خوارزم باس لقياس الانتشار (Bass diffusion model) وذلك بالاندماج مع خوارزم الانحدار التدريجي (mini-batch gradient descent) للحصول على القيم المثلي لمعاملات خوارزم باس. تم تدريب النموذج على حوالي ٨٥٪ من البيانات التاريخية المتاحة حول عدوى (COVID-19)، واختباره على بقية البيانات. أوضح النموذج المقترح أن المملكة العربية السعودية سوف تواجه تزايدا في الأيام القادمة من حيث ارتفاع حصيلة أعداد الإصابات. وأن معدل الزيادة في الإصابات سينخفض مع الوقت حتى يصل الي معدلاته الدنيا في شهر يناير على أقل تقدير من العام القادم وذلك في حالة عدم ارتداد الموجه. كما أظهر النموذج أن نقطة تسطيح المنحني لأعداد الإصابات ستكون خلال الشهر ذاته، وهو الموعد المتوقع لانحسار الوباء بالمملكة العربية السعودية وذلك في حالة عدم وجود موجات ارتدادية أخرى.

1. Introduction:

Coronaviruses are a wide group of viruses that can cause many human losses, ranging from a common cold to severe acute respiratory syndrome. Also, viruses from this group cause many different animal diseases. Since many early cases were associated with a large market for marine and animal food in Wuhan, China, the virus is believed to have an animal origin, but so far it has not been confirmed [1].

Since the emergence of the emerging SARON virus (SARS-CoV-2) in December last year 2019, researchers around the world have developed various models that rely on the data monitored about (COVID-19) infection to predict the damage and injuries that will be left shortly around the countries of the world. Some models have tried hard to find a date for the outbreak of the virus globally or even in certain regions. However, this type of prediction, which is "time-related prediction", may develop some emerging factors that may make it a very accurate prediction. This is because the values of the observed readings change periodically and around the clock. Among the efforts made was a model developed by the Institute for Health Metrics and Evaluation (IHME) at the University of Washington [2], which is updated around the clock. Another model was developed by the MRC Center for Analysis of Global Infectious Diseases in London [3]. Also, some models focused on predicting future mortality numbers and numbers that will need intensive care [4-6]. While many other studies focused on answering two main questions: 1) How many expected cases of infection will be 2) and what is the maximum number of injuries that can be reached [7-9]. On the other hand, the extent of the impact of social separation, travel restrictions, mitigation strategies, and embargoes imposed by governments on societies has been highlighted, to study the impact of these measures on the expected numbers and statistics [8-11]. Some studies published [4-7] also attempted to verify the accuracy of some of the proposed prediction models. However, it was found that the cited models that were published by the IHME had errors and significant deviations from the real numbers [6, 12]. Where it was found that the rate of diffraction in the numbers of real deaths is very large and completely different from the expected numbers from the proposed model [13]. The IHME team later revised the model [5], but prediction errors are still high and out of real numbers. Over time, researchers learn and develop their prediction methods and algorithms to reach near-accurate or largely unrealistic numbers. Despite the intrinsic complexity and uncertainty of predictions of a COVID-19 infection, some of the work presented significantly affected some of the precautionary policies and procedures in some respects [14, 15].

Uncertainty and ambiguity about the COVID-19 infection increase the desire of countries and governments to know the future expectations and effects that the infection may have, especially those whose economies have been greatly worsened by the pandemic. At the same time, it becomes clear that it is also difficult to complete the prediction accurately due to the mystery surrounding the pandemic. The primary challenge rooted in the disease (COVID-19) as a "wicked problem", which was formulated by researchers Rittel and Webber [16]. Where troublesome problems are described as those new, unique, complex, and evolving problems with incomplete, contradictory, and changing requirements, often difficult to identify. Here are some characteristics of troublesome problems: 1) They are often closely related to ethical, political, economic, or professional issues. 2) It cannot be solved by traditional analytical method. 3) No solution can be objectively tested or evaluated and then confirmed as right or wrong. 4) It makes no sense to talk about optimal solutions to these kinds of problems. 5) The proposal designed to solve the problem may lead to the emergence of other hidden problems.

COVID-19 infection is one of these bothersome problems, which are highly mysterious and not naturally predictable in the general sense. However, this does not mean that objective science-based analyzes and forecasts are completely futile. This only means that the stereotypes of traditional solutions to optimization, accuracy in modeling, and predictions about this pandemic should be avoided. For example, in some of the proposed prediction models for COVID-19 infection [4, 5, 12, 13], there are clear intentions or goals to improve prediction models based on the variance between real and predicted values, but the answer will still be inaccurate and not final to some extent. So talking about the timing of recession may be meaningless and may create a false sense of certainty that is not present. Moreover, when real-world scenarios change in terms of government interventions and human behaviors, it will be naive to assess the accuracy of a model that has been trained using data collected in conditions different from the one it is being tested on and that has been created under completely different conditions. Therefore, to ensure prediction more closely to realism in light of these procedures and changing decisions periodically, different prediction models must be fed with recent data on infection to reduce the error gap that may be present.

Consequently, my infection (COVID-19) requires an innovative strategy to elicit insights based on periodically updated data. This is called predictive monitoring: it is the monitoring of variable forecasts that are constantly updated with the latest data, along with real data monitoring. Dealing with time-bound predictive monitoring (COVID-19) is not clear as mentioned above, due to several other additional factors: 1) Historical data on the infection (COVID-19) that have been collected, may not be sufficient to allow for a prediction long term. 2) Results may differ according to the precautionary measures emerging by the governments, which may change periodically and around the clock. Add to this the general awareness of the people and their commitment to these measures. For example, Countries that have taken harsh and strict measures against this pandemic, as China has done, have had different results than other countries that initially tolerated precautionary measures, or who recently discovered it. 3) Rapid mutations created by the virus to disguise and adapt to different environments by copying new copies that are difficult to identify. Analysis of more than 5,300 genomes of the Coronavirus in 62 countries showed that although the virus was somewhat stable, some "genomes" were gaining mutations, including a mutation in "Spike protein," the protein the virus uses to infect human cells. Scientists say the genetic change in "Spike protein" is a sign of the virus adapting to its human host. According to researchers at the London School of Hygiene and Tropical Medicine, it is unclear how mutations affect the virus, but these mutations that have arisen independently in different countries may help the virus to spread more easily. 4) Possible bouncing wave scenarios that might strike again and forcefully. As part of the analysts' believes that China did not completely control the virus. As Dr. Li Languan, who led a medical team to combat the virus in Wuhan, said that there are still many patients in critical condition, she also indicated that it could lead to the opening of the Chinese borders with other countries to a second wave of the spread of the virus in China. According to a study conducted by a hospital in Wuhan, it was revealed that 3-14% of those recovering from the virus contract it again without showing any symptoms, and they are called "Silent Carrier". In the same context, "Wang Wei" - director of the "Tongji" Hospital in Wuhan - told CCTV channel that 5 out of 147 recoverees are infected with the virus again. This scenario also reinforces some historical facts, such as the influenza pandemic that spread in 1889 and 1918 and spread in three waves, and each wave was stronger than the one that preceded it as a result of the mutating of the virus. Therefore, if these factors were previously investigated, it would affect the accuracy of the prediction models.

This research paper aims to analyze the data of the COVID-19 pandemic worldwide, especially in the Kingdom of Saudi Arabia, and compare it with some countries to extract some important information from it. It is known that the Kingdom of Saudi Arabia is special, therefore, because it incubates dozens of different nationalities from all over the world. So, its steps should be more careful than other Middle Eastern countries due to their different nature. Besides, this paper proposes a model for predicting the numbers of casualties expected in the coming days and when the epidemic will subside. The Bass diffusion model was used in combination with the mini-batch Gradient descent model to obtain optimal values for the Bass algorithm coefficients. Predictive monitoring can guide decision-makers in their vision, plans, and future actions to shape a safer and more interactive future for a pandemic.

2. Methodology of work

The work methodology of this paper is to summarize infection data analysis (COVID-19) using the most famous Python language analysis and display libraries (NumPy, Pandas, and Matplotlib). The data was described at the beginning, then the pandemic data was reviewed and analyzed at the level of the countries of the world combined. After that, the analysis was carried out at the level of the Kingdom of Saudi Arabia and compared to some other countries and some neighboring countries to extract some important information and indications from it. In the end, a model based on time-related prediction was built on the data on (COVID-19) infection, which included determining the number of cases of confirmed, death and recovery expected to be reached and determining when the epidemic will recede.

2.1. Data description

Johns Hopkins University has prepared a smart, dynamic (Dashboard) report that is updated around the clock using affected case data. Johns Hopkins University relied on data collected from the World Health Organization (WHO) website and the CDC website for infection (COVID-19). This dataset contains information on the number of cases and deaths and the number of people recovering from (new COVID-19) disease for the year 2019. The data have the following component:

- History of notes (data)
- State or state
- Country / Territory - Observation Country
- Latest update - time UTC
- The daily and cumulative number of confirmed cases as of that date
- The daily and cumulative number of deaths as of that date
- Daily and cumulative number of cases recovered to that date:

2.2. Covid-19 data analysis in world

In this paragraph, some basic information about the infection that was monitored until 22/9/2020 is shown. Table 1 shows the number of countries that have been affected by the epidemic so far, and which exceeded two hundred countries. The table also shows the total number of confirmed cases, which have approached thirteen million, cases of disease recovery, deaths, and active cases, so far, cumulatively. From Table 1, it is possible to deduce the number of active cases, where the active cases are interpreted as the number of confirmed cases - the number of cases recovered - the number of deaths.

Table 1 also shows the number of active cases, which are cases whose condition has not changed from either injury to recovery or death until now. It may be possible that the increase in the number of active cases is a dependable indicator of the numbers of recovering cases or numbers of deaths compared to the number of confirmed cases, as the proportion between them is inverse. In Figure 1, the number of injuries and deaths, and the numbers of recovering cases are shown daily. On the other hand, Figure (2) shows the number of closed cases. Where closed cases mean, they are cases that have been decided and changed from being active cases to cases that have been recovered or cases of death. The number of closed cases is calculated as = the number of recovered cases + the number of deaths. The report notes that closed cases are increasing markedly with higher recovery rates and lower death rates.

Table 1: Basic information about a COVID-19 infection

Parameter	Value
The total number of countries where the disease is spread:	223
The total number of confirmed cases worldwide:	31779835
The total number of cases that have recovered around the world	21890442
Total number of deaths worldwide:	975104
The total number of active cases worldwide:	8914289
The total number of closed cases worldwide:	22865546
The approximate number of cases confirmed daily around the world:	129186
The approximate number of cases recovering per day around the world:	88986
The approximate number of deaths per day worldwide:	3964
The approximate number of confirmed cases per hour worldwide:	5383
The approximate number of cases recovering per hour around the world:	3708
The approximate number of deaths per hour worldwide:	165

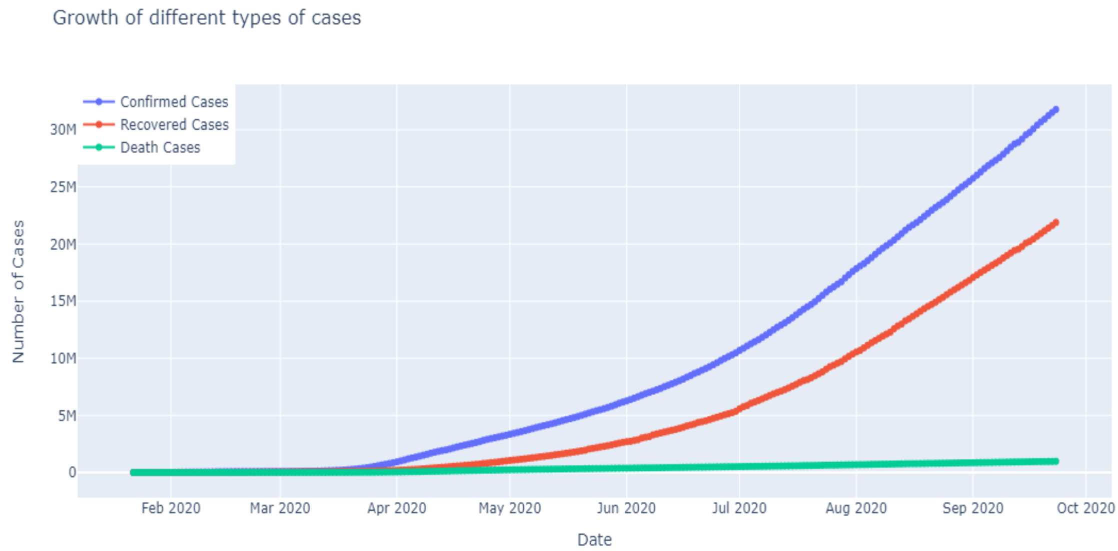

Figure 1: Numbers of infected, recovered, and deceased cases about COVID-19 infection

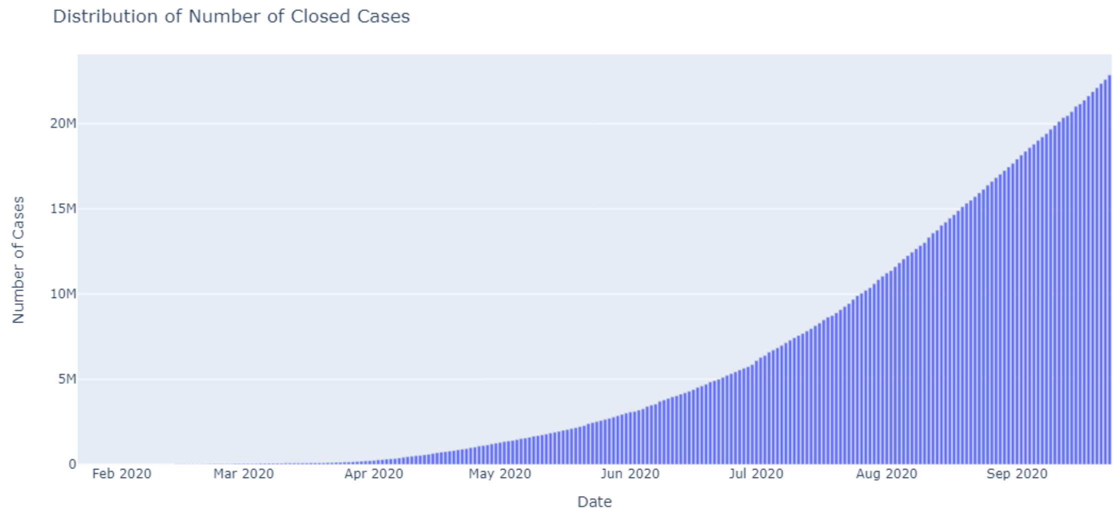

Figure 2: Number of closed cases around COVID-19 pandemic

2.3. Analysis of the Kingdom of Saudi Arabia data

In this section, some basic information about the numbers of COVID-19 infections in Saudi Arabia is presented. Table 2 shows the cumulative number of confirmed and recovered cases, deaths, and active cases as of 22/9/ 2020. Also shown in the table are some statistics of the mean numbers of approximate daily numbers and hourly numbers.

Table 2: Basic information about the Kingdom of Saudi Arabia COVID-19 infection

Parameter	Value
The total number of confirmed cases:	331359
The total number of cases recovered	313786
The total number of deaths:	4569
Total number of active cases	13004
Total number of closed cases	318355
The approximate number of cases confirmed daily:	1609
The approximate number of cases recovering daily:	1523
The approximate number of deaths per day:	22

From Table (2), it turns out that the total number of cases in the Kingdom of Saudi Arabia reached 331359 as of 22/9/2020, while the numbers recovered exceeded 313 thousand, or 94.6%. Looking at the data, we notice that the recovery rate is significantly high and promising. While the data showed that the number of deaths due to the pandemic in the Kingdom reached 4569 cases or 1.3% of the total cases. It is somewhat too low for many countries. The table also shows the approximate average averages for each day and for each hour .

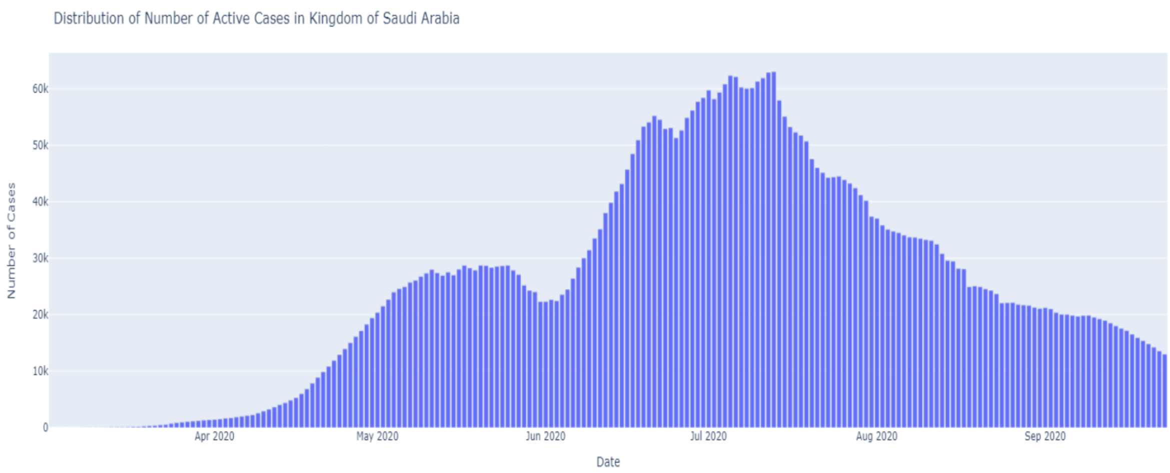

Figure 3: Distribution of numbers of active cases in the Kingdom of Saudi Arabia

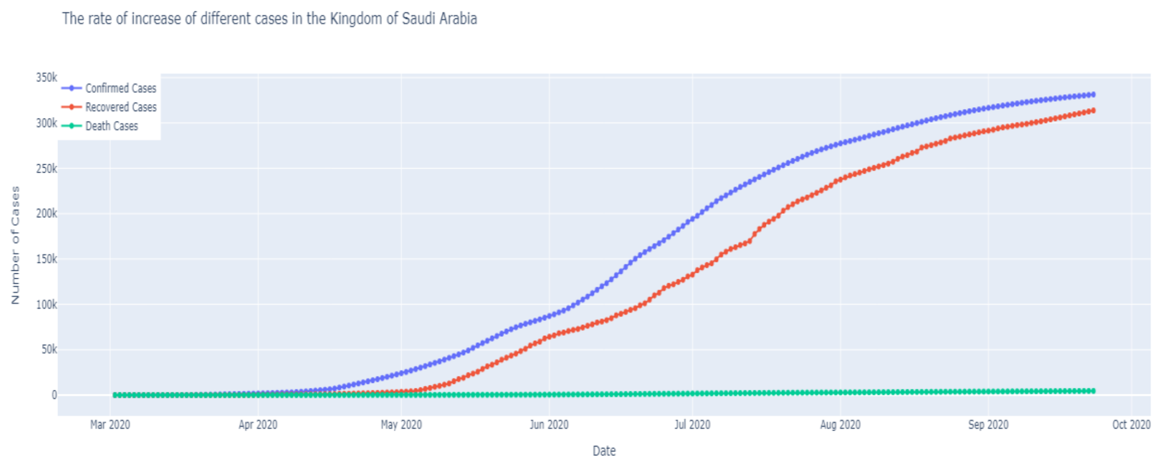

Figure 4: Daily Increase Numbers on the COVID-19 Pandemic Cases in the Kingdom of Saudi Arabia

In Figure (3), the data showed that the rate of active cases was increasing daily, until July 2020, and showed relatively stable stability, until it started to decline at the end of July till now. Figure (4) shows an increase in the number of cases of injury, recoveries, and deceased cases in the Kingdom of Saudi Arabia. The figure shows that at the end of May, the rate of casualty monitoring started to decline significantly and started in early June to the end of July.

3. Prediction

3.1. Bass diffusion model

Predicting the prevalence of COVID-19 is a very complex process. Many algorithms were used for prediction and relied on several techniques including time series [18], support vector machine [19], neural network [20], and other machine learning prediction models [21], but nothing is superior to the Bass diffusion model [22 - 24] is intended to predict prevalence. The Bass model is characterized by its simplicity and accuracy in predicting the average number of consumers for products related to time. There are thousands of academic articles on the Bass model and its applications. Frank M [22] papers one of the most important articles that have cited several citations to be the highest ever in predictive research in general. In this research paper, we seek to employ the Bass diffusion model in combination with mini-batch gradient descent to predict the prevalence of a COVID-19 epidemic in the Kingdom of Saudi Arabia and some other countries, by determining the number of infections that are likely to be reached during the coming months and then determining the date of the epidemic recession.

The nature of the Kingdom of Saudi Arabia differs from most of the countries of the world, due to the presence of the two holy mosques in it, and to attract a large number of expatriate workers of various specialties. In theory, each of these people is vulnerable to a COVID-19 infection. So, we'll assume that there is a fixed potential population for infection denoted by the symbol m. Where m represents the population of a country or city. When the epidemic first sweeps, everyone is at risk of infection. And with many infected people, the number of people exposed to the infection thus decreases. (Here we assume that no one can get COVID-19 infection twice and everyone can get infected at most once) to complete the prediction. The number of infections is denoted by the symbol S, and it changes as a function of time t. The number of cumulative injuries is also denoted by the symbol Y and it also changes as a function of time t.

$\boldsymbol{m}$	*The number of people nominated for infection*
$\boldsymbol{S}_{(\boldsymbol{t})}$	*The number of people candidate for infection per day*$\boldsymbol{t}^{\boldsymbol{th}}$
$\boldsymbol{Y}_{(\boldsymbol{t})}$	*The number of cumulative persons who are candidates for infection to day*$\boldsymbol{t}^{\boldsymbol{th}}$

After defining these variables, we will have the following simple definition. That is, the number of cumulative infections is equal to the sum of all infected in days t.

$$\boldsymbol{Y}_{(\boldsymbol{t})} = \sum_{\boldsymbol{\tau}=\boldsymbol{1}}^{\boldsymbol{t}} \boldsymbol{S}_{(\boldsymbol{t})} \tag{1}$$

We can calculate the number of people who are not yet infected, and who are denoted by the symbol P_((t)) by equation (2). These people are defined as the number of people at risk of infection m-people who were infected Y_((t))

$$\boldsymbol{P}_{(\boldsymbol{t})} = \boldsymbol{m} - \boldsymbol{Y}_{(\boldsymbol{t})} \tag{2}$$

The Bass diffusion model was used to predict the number of buyers in stores based on their historical data at a specific time. So, the buyers were categorized into two categories: Immediate Buyers and Counterfeit Buyers. As for the buyers 1) The initiators: they are the most receptive buyers of the new products and who have a desire to acquire the new products when they are issued. While buyers (2) imitators do not have the lead in purchasing new products, but they do not buy it. In the end, buyers buy the product, but some initially buy it, and some buy it after that. The mathematical model of the Bass diffusion model is known as Equation (3), where the model predicts the number of potential buyers of their types at time t.

$$\boldsymbol{S}_{(\boldsymbol{t})} = \boldsymbol{\rho}(\boldsymbol{0}) + \frac{\boldsymbol{q}}{\boldsymbol{m}} \boldsymbol{Y}_{(\boldsymbol{t})} \tag{3}$$

Where the initiators are marked with the symbol ρ (0), they were the ones who initially purchased at zero time. Q indicates the number of imitators who purchased at a later time. To adjust the Bass model and reformulate it to fit the problem of predicting the number of pandemic injuries: We assume that at the beginning, there were certainly many unknown injuries whose owners were injured without their knowledge of the pandemic. These correspond to the numbers of proactive buyers in the procurement prediction model. These individuals who have been exposed to COVID-19 infection because they were unaware of the pandemic, and who did not initially know it, will be symbolized ρ (0) and they are the people who were infected in time zero. Also, these people marked with the symbol ρ (0) are not the only ones at risk. Other uninfected people deal with these infected people and become infected as well. This second type of person, who was aware of the pandemic after announcing it, and despite these people being aware of the dangers of social participation with the injured, they chose to violate the instructions and not to take into account the precautionary social divergence procedures and allow themselves to be infected with this infection. This group of people is denoted by q. It should be noted that the individuals in the aforementioned categories have clear differences in the mechanism of infection, but all of them are ultimately infected.

The Bass diffusion model is shown in its traditional form, shown in equation (3). But the model in its image is not suitable for application directly in our problem. To be compatible with the problem of infection, the model must be reformulated in another appropriate forum. As for predicting the numbers of buyers, it is somewhat different from predicting the numbers of people injured in the event of a pandemic. In purchases, the number of buyers increases, while in our case the number of injured people increases and decreases at the same time. The truth is that the number of injured increases, but at the same time decreases daily. Once a person has been infected, it is classified as infected and counted as infected. However, the outcome of the injury does not keep the same condition as in the buyers' prediction model. The patient may change his condition to two other cases in the end, or to keep his condition infected as is. The patient may turn from injury to death or recovery. But surely the rate of injuries will be greatly affected by these results. Accordingly, the total number of recovery cases and deaths is referred to as closed cases. From the above, we can say mathematically that closed cases o are divided into two parts. Reciprocal cases of death and deaths.

o The number of closed cases - who switched from injury to another

$O_{(t)}$ *The number of closed cases - who have changed from infection to another until today*$\boldsymbol{t^{th}}$

Therefore, equation (3) for the prediction model will be modified to appear in its new form, which is compatible with the prediction of the pandemic numbers, as with equation (4): where the total number of injuries will be affected by cases of recovery and deaths alongside cases with new cases of infections.

$$\boldsymbol{S_{(t)} = \rho(0) + \frac{q}{m}[Y_{(t)} - O_{(t)}]} \tag{4}$$

3.2. gradient descent

In the fourth equation for the pandemic injury prediction model, the values of the coefficients ρ (0), q, remain unknown, which relates to the number of casualties for people who are ignorant of the world. M is defined as the number of people at risk in the country for which it is predicted. Y_((t)), O_((t)) are defined as the cumulative number of closed and active cases respectively. While the coefficient values ρ (0), q remains unknown. What is required is to determine the optimal values for them in each country to perform a correct and realistic forecasting process. There are many gradient descent algorithms, which differ in how they work to calculate the gradient of a specific goal function. A gradient slope algorithm is an optimization algorithm that is used to reach the minimum coefficients for a given target function. According to Ruder, S. [25], and after making several comparisons between linear regression algorithms, it was found that the most accurate results and the lower number of turns to reach the optimal coefficients are mini-batch gradient descent. The objective here of applying the linear regression algorithm is to determine the optimum values for the coefficients ρ (0), q so that the prediction values for the incidence cases in the previous period for which we have their data are identical and consistent with the real data in that period in the training phase.

The mechanism of work is divided into two parts: a section for training and a section for forecasting. In the training phase, the model is trained on historical data to arrive at the optimal values for the coefficients, which make the real and predictive readings of that period identical. As for the stage of forecasting, it is the stage in which the numbers of injuries and death are predicted in the coming future days, and the date of flattening of the curve is determined, which is the time of receding of the epidemic and the absence of any additional injuries.

4. Implementation and result analysis

In this section, the parameter setting for the Bass-MBGD model is discussed and computational results are presented.

4.1. Simulated instances

In the implementation phase, the work is divided into three phases: 1) the model training phase. 2) prototype testing stage. 3) Prediction stage. As for the first stage, which is the training phase, the aim was to obtain the optimum values for the coefficients $\rho(0)$ and q for each country which make the real readout values approximately equal to the values predicted by the model. The bass model was implemented to predict future readings related to time for each country as in equation (4). The inputs to the model are the population of each country, which is what we get from Table (3). As for the population of the Kingdom of Saudi Arabia, the numbers of expatriates were added to the number of citizens due to their large number. We obtain the numbers of confirmed, recovery, and mortality for each country from the data described in paragraph (1.1). Before making the prediction and diffusion model, the stepwise regression algorithm assumes values for the coefficients ρ (0), q, and is offset in equation (4), the actual values compared with the inferred values, and the error amount calculated. The process is done in many circuits until the lowest error value is reached. During the training period of the model on the data of each country, the optimum values for the coefficients ρ (0), q, some of which are mentioned in Table (3), were obtained .

Table 3: Some countries' population numbers and transaction values $\rho(0)$، q

Country	Number of populations	$\boldsymbol{\rho(0)}$	$\boldsymbol{q}$
Brazil	16436526	0.011	12.4579
India	2975443	0.013	13.124
South Africa	2045931	0.007	8.7911
US	1900541	0.046	52.8236
Kuwait	1899508	0.011	15.0557
Russia	734909	0.039	89.7219
Qatar	482567	0.016	15.2599
Egypt	396549	0.009	8.4178
Bahrain	357503	0.008	7.4706
Saudi Arabia	356251	0.022	23.1464
United Kingdom	314587	0.047	60.8451
Italy	236313	0.059	48.2137
Spain	231939	0.087	352.029
Bangladesh	194387	0.024	33.2524
France	186878	0.083	406.142
Turkey	160517	0.073	377.597
Iran	137544	0.041	18.8444
Canada	108062	0.044	49.8065
United Arab Emirates	65392	0.028	25.0833

We assumed that no one could be infected with COVID-19 twice and each person could be infected at most once. We know that this assumption is not realistic, but it is more realistic than to assume that everyone will be infected two or three times, or that 40%, for example, of those infected persons will be infected again. This is because the data for that are not available. It is possible to say that this hypothesis is an indicator of the expected minimum number of confirmed cases, not the maximum. As other factors affect the number of confirmed cases, including 1) If a person is infected more than once, the actual number of populations increases in each country. 2) If there will be a second and third backlash striking the country, the model will give different results certainly. Therefore, it was necessary to set some convenient assumptions on the basis of which we say that this is the minimum expectation.

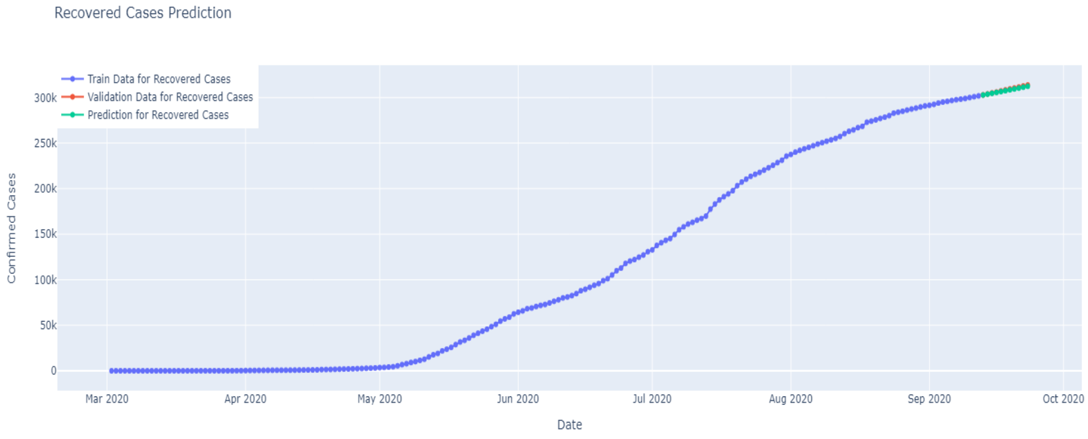

Figure 5: Training and testing period for the model on cases of recovered cases in the Kingdom of Saudi Arabia

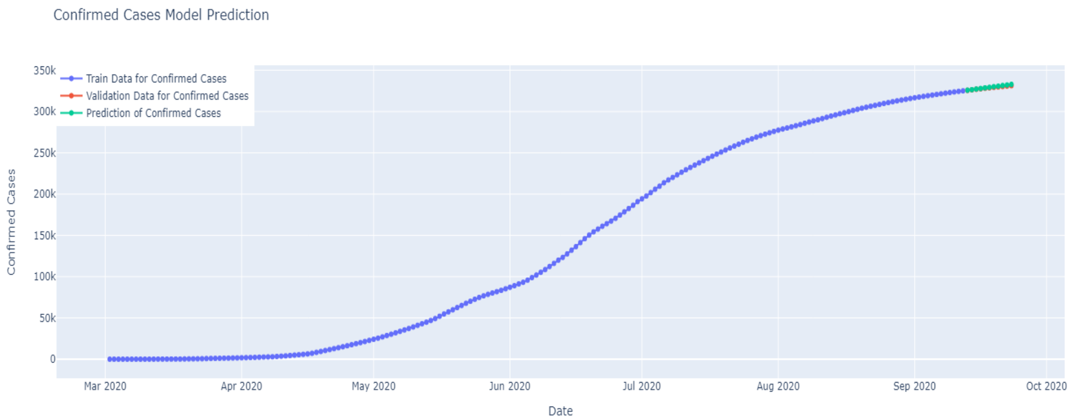

Figure 6: Training and testing period for the model on cases of confirmed cases in the Kingdom of Saudi Arabia

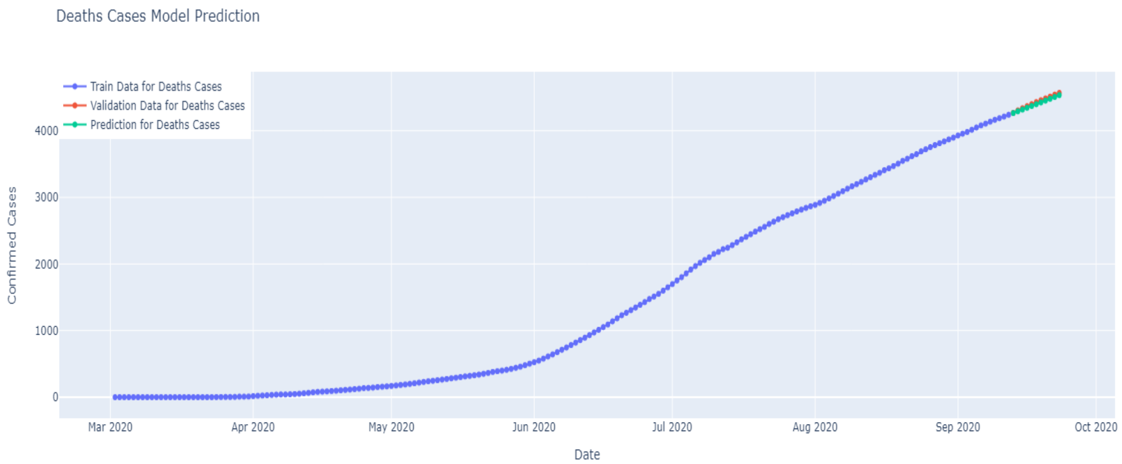

Figure 7: Training and testing period for the model on cases of mortality cases in the Kingdom of Saudi Arabia

4.2. Computational results

The model was trained and tested on numbers for each country separately. where confirmed, active, recovered, and mortality cases will be predicted for the next 500 days. In Figure (5, 6, and 7) the numbers of confirmed, recovered, and mortality cases in the Kingdom of Saudi Arabia are shown as an example. The range of the data in the Kingdom of Saudi Arabia is from the date of the start of monitoring until September 20, 2020. The data is divided into two parts, in a training section the model is represented in about 85% of the data and the rest of the data is used for testing. The blue curve represents the training period of the model in which the optimum values for the parameters of the Bass model for propagation are determined by the gradual regression algorithm. The test period appears later in both the green and red curves, in which the Bass algorithm is applied with the values of the coefficients inferred to predict recovered, confirmed, and death numbers represented by the green line. The readings obtained from the model in the green line appear in the testing phase to a large extent actual data represented in the red line. Note that the model was able to predict the data for this period properly, almost close to the actual readings.

To assess the effectiveness of the Bass diffusion model, we also compared the model in terms of root mean square error (RMSE) with other forecasting algorithms like ARIMA (2,1,1) [26] and the Bass model [27] [28]. In this work, [26] the authors employed the Autoregressive Integrated Moving Average (ARIMA) model to forecast the expected daily number of COVID-19 cases in Saudi Arabia for four weeks. Bass Model [27] is developed for deaths for the period, March 21 to April 30 for the USA as a whole and as the US States of New York, California, and West Virginia. And finally, in this study, [28] the authors adapted the bass diffusion model to determine the time when the COVID-19 curve flattens in the Philippines. Further, it also determined the possible incidence of the second wave of infection.

Table 4: ARIMA order selection based on AIC and BIC approaches in Saudi Arabia for confirmed cases

	AIC	BIC
ARIMA (2,1,1)	3133.5	3144.7
ARIMA (2,1,2)	3134.2	3141.5
ARIMA (1,2,1)	3110.8	3119.8
ARIMA (1,2,2)	3101.9	3111.9
ARIMA (2,2,2)	3138.2	3149.6

All of the comparison algorithms have been re-implemented and experimented on the same data. Moreover, Akaike information criterion (AIC), and Bayesian information criterion (BIC) criterion are employed for model selection among a finite set of models ARIMA in Kingdom of Saudi Arabia for confirmed cases. As we see in table 4. The best model of ARIMA is ARIMA (1, 2, 2). On the other hand, we compared all models in terms of Root Mean Square Error RMSE. As presented in table 5, based on our results, the prediction methods of our model performed better than other models. According to table 5. we can see that the Bass-GD model got the lowest RMSE in the most sample, and therefore it should be able to predict the number of confirmed, recovered, and deaths cases of COVID-19 in Saudi Arabia and other countries in the next coming months better than the other models.

Table 5: RMSE for different models in some countries for confirmed cases

		ARIMA (1, 2, 2) [26]	Bass model [27]	Bass model [28]	Bass-MBGD
Kingdom of Saudi Arabia	Confirmed	435.20	441.19	443.05	428.06
	Recovered	891.59	900.73	901.64	878.05
	Deaths	82.63	30.82	33.06	27.30
Russia	Confirmed	855.94	8731.46	8806.50	8560.80
	Recovered	20169.92	20308.41	20487.75	19926.80
	Deaths	767.30	748.55	776.31	750.30
UAE	Confirmed	149.13	146.14	149.44	144.40
	Recovered	300.11	302.16	304.79	292.60
	Deaths	13.30	16.36	13.43	8.20
Egypt	Confirmed	129.77	124.62	133.72	122.30
	Recovered	224.03	232.57	228.54	219.40
	Deaths	99.90	108.02	112.93	101.10
United Kingdom	Confirmed	1279.27	1322.29	1330.89	1288.80
	Recovered	2657.67	2623.20	2699.62	2622.20
	Deaths	116.33	109.10	121.09	110.01

After training the model and adjusting the values of the transactions, the data of five hundred days in the future was extrapolated through the last stage, the stage of predicting the numbers of confirmed, recovered, active, and mortality for several countries during the coming months, as shown in Figure (8-12). The dark lines represent the actual numbers of confirmed, deaths, recovered, and active respectively. The faint lines belong to the model. In Figure (8) for the Kingdom of Saudi Arabia, the model predicts that the number of confirmed numbers in the Kingdom will be at least more than 355 thousand cases, in case no second wave. But if the second wave comes, the figures will be increased greatly. and it indicated that the number of confirmed cases will decline significantly during December 2020 until reaching the curve flattening of injuries in January of the year 2021, which is the expected date for the recession of the pandemic in the Kingdom. The model also showed that the death rate will remain significantly low about the number of injuries in case no second wave. In figures (9-12), expected confirmed, recovered, deaths and active figures for some countries are plotted.

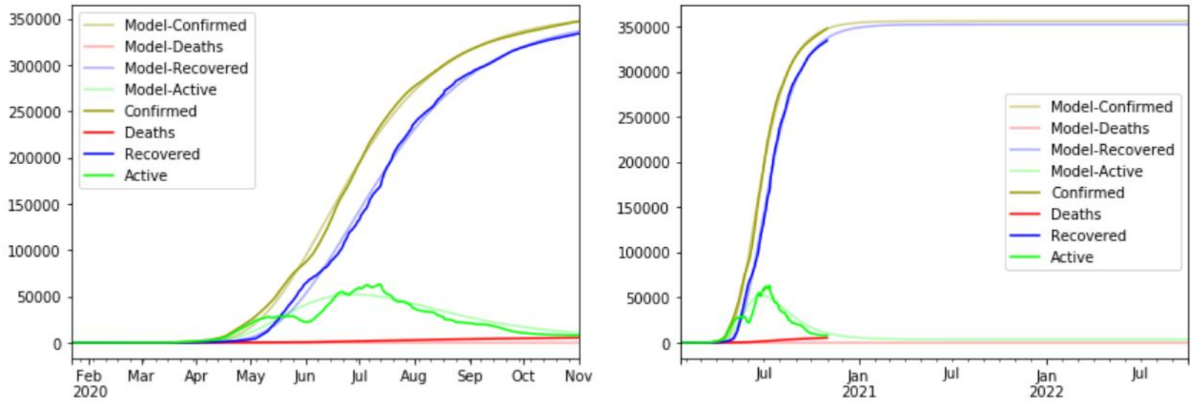

Figure 8: Predicting the number of future cases in the Kingdom of Saudi Arabia

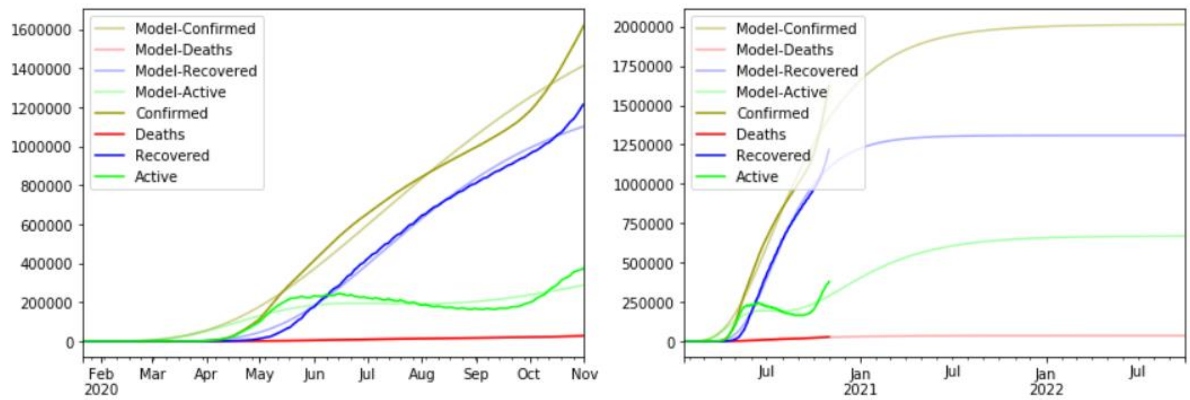

Figure 9: Predicting the number of future cases in the Russia

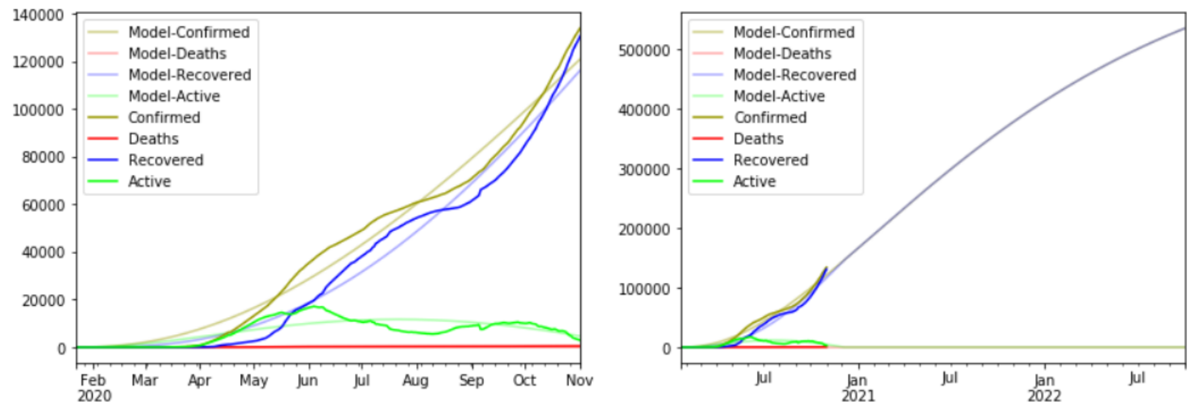

Figure 10: Predicting the number of future cases in the UAE

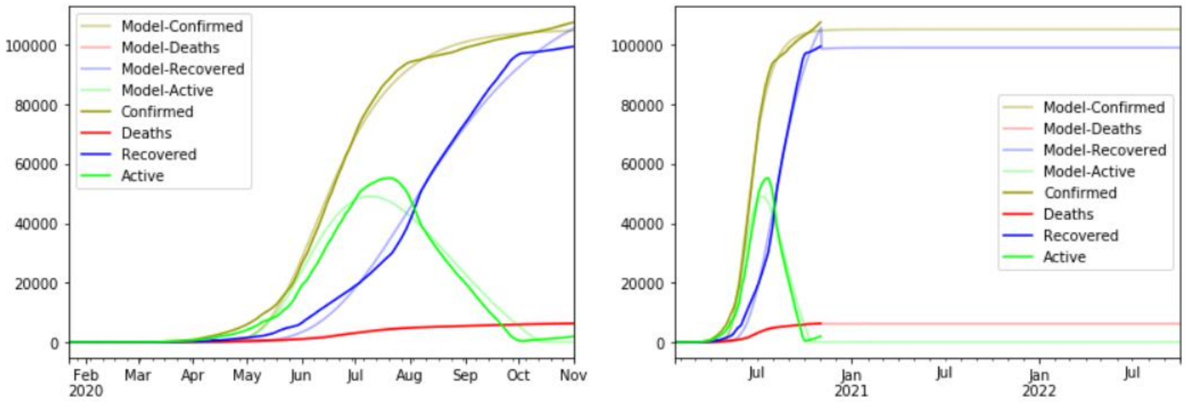

Figure 11: Predicting the number of future cases in Egypt

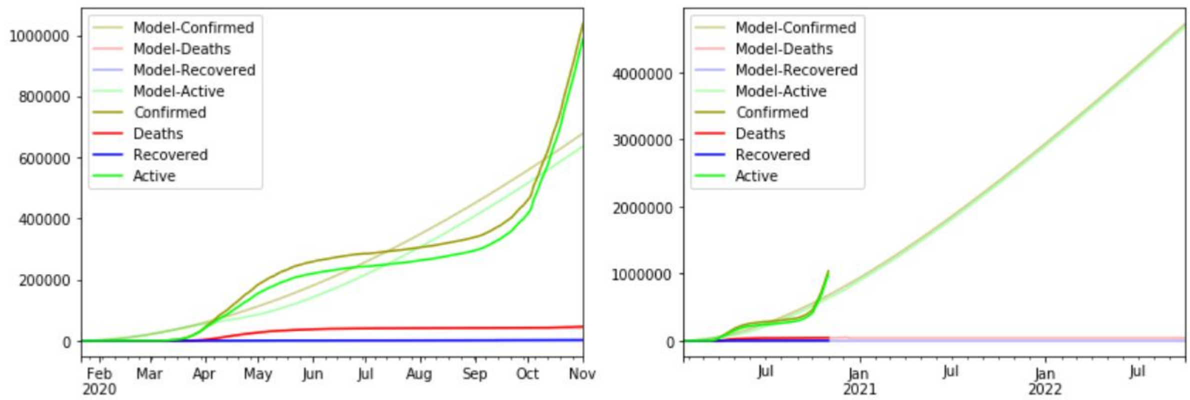

Figure 12: Predicting the number of future cases in the United Kingdom

5. Conclusion and future work:

In this research paper, we reviewed the analytical and statistical methods to extrapolate the most important data and indicators about the COVID-19 infection and the rates of confirmed, recovery, and mortality during the past few months, especially in the Kingdom of Saudi Arabia. On the other hand, the process of predicting the rates of infection prevalence (COVID-19) in the Kingdom of Saudi Arabia and some other countries during the coming days was done by relying on the Bass diffusion model in combination with the mini-batch Gradient descent model. The model was trained on 85% of historical data from the beginning of the actual and significant change around my infection (COVID-19) and tested on the rest of the data. The proposed model showed that the Kingdom of Saudi Arabia will face an increase in the coming days in terms of an increase in the number of confirmed, deaths, recovery, and active cases. And that the rate of increase in injuries will decline over time until it reaches its lowest level in January of next year at the very least, knowing that it is possible that this period will increase a little. The model also showed that the curved flattening point for the numbers of injuries is likely to be during the month of January 2021, which is the expected date for the epidemic to recede in the Kingdom of Saudi Arabia and to record a rate of zero injuries in the absence of other aftershocks.

In the next few days, the model will be developed to predict other aftershocks, and other models will be developed for the prediction.

6. References

[1] "Update and Interim Guidance on Outbreak of 2019 Novel Coronavirus (2019-nCoV) in Wuhan, China CDC Health Update". New Jersey Department of Health. 18 January 2020

[2] Institute of Health Metrics and Evaluation (IHME) at University of Washington https://covid19.healthdata.org/united-states-of-america

[3] MRC Centre for Global Infectious Disease Analysis at the Imperial College. https://mrc-ide.github.io/covid19-short-term-forecasts/index.html

[4] IHME COVID-19 Health Service Utilization Forecasting Team and Christopher JL Murray. “Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months,” 2020. MedRxiv https://doi.org/10.1101/2020.03.27.20043752

[5] IHME COVID-19 health service utilization forecasting team and Christopher JL Murray. “Forecasting the impact of the first wave of the COVID-19 pandemic on hospital demand and deaths for the USA and European Economic Area countries,” 2020. MedRxiv https://doi.org/10.1101/2020.04.21.20074732

[6] Woody S., et al. “Projections for first-wave COVID-19 deaths across the US using social-distancing measures derived from mobile phones,” 2020. MedRxiv https://doi.org/10.1101/2020.04.16.20068163

[7] Yang Z., et al. “Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions.” J Thorac Dis, 12(3): 165-174, 2020. DOI: 10.21037/jtd.2020.02.64

[8] Ferguson N.M., et al. “Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College London.” 16 March 2020. Accessed at www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-03-16-COVID19-Report-9.pdf.

[9] Petropoulos F., Makridakis S. “Forecasting the novel coronavirus COVID-19,” PLoS ONE, 15(3): e0231236, 2020. https://doi.org/10.1371/ journal.pone.0231236

[10] Kissler S,M. et al. “Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period.” Science, 14 April 2020. DOI: 10.1126/science.abb5793

[11] Chinazzi M., et al. “The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak.” Science, 24 April 2020, 395-400

[12] Jewell N.P., et al. “Caution Warranted: Using the Institute for Health Metrics and Evaluation Model for Predicting the Course of the COVID-19 Pandemic.” Ann Intern Med, 14 April 2020. DOI: https://doi.org/10.7326/M20-1565

[13] Marchant R., et al. “Learning as we go: An examination of the statistical accuracy of covid19 daily death count predictions,” 2020. DOI: https://doi.org/10.1101/2020.04.11.20062257

[14] Grey S. and MacAskill A. "Special Report: Johnson listened to his scientists about coronavirus - but they were slow to sound the alarm." Reuters, April 7, 2020. https://reut.rs/2Rk3sa7

[15] Resnick B. "The White House projects 100,000 to 200,000 Covid-19 deaths." Vox, March 31, 2020. https://www.vox.com/science-and-health/2020/3/31/21202188/us-deaths-coronavirus-trump-white-house-presser-modeling-100000

[16] Rittel H.W.J. and Webber M.M. "Dilemmas in a General Theory of Planning." Policy Sciences, 4 (2): 155–169, (1973). doi:10.1007/bf01405730

[17] Kermack W. O. and McKendrick A. G. (1927) "A Contribution to the Mathematical Theory of Epidemics," Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, Vol. 115, 772, pp. 700-721.

[18] Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159-175.

[19] Du, X. F., Leung, S. C., Zhang, J. L., & Lai, K. K. (2013). Demand forecasting of perishable farm products using support vector machine. International journal of systems Science, 44(3), 556-567.

[20] Kuo, R. J. (2001). A sales forecasting system based on fuzzy neural network with initial weights generated by genetic algorithm. European Journal of Operational Research, 129(3), 496-517.

[21] Mahajan, Vijay, Muller E, and Wind,Yoram. (2000) New-Product Diffusion Models, vol. 11, New York: Springer Science & Business Media.

[22] Bass, Frank M. (1969) "A New Product Growth for Model Consumer Durables" Management Science, Vol. 15, 5, pp. 215–227.

[23] Bass, Frank M., Krishnan, Trichy V. and Jain, Dipak C. (1994) "Why the Bass Model Fits without Decision Variables" Marketing Science Vol. 13, 3, pp. 203-223.

[24] Bass, Frank M. (2004) "Comments on 'A New Product Growth for Model Consumer Durables the Bass Model'" Management Science Vol. 50, 12, pp. 1833–1840.

[25] Ruder, S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.

[26] Alzahrani, S. I., Aljamaan, I. A., & Al-Fakih, E. A. (2020). Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. Journal of infection and public health, 13(7), 914-919.

[27] Gurumurthy, K., & Mukherjee, A. (2020). The Bass Model: a parsimonious and accurate approach to forecasting mortality caused by COVID-19. International Journal of Pharmaceutical and Healthcare Marketing.

[28] Zeny L. Maureal, Jovelin M. Lapate, Madelaine S. Dumandan, Vanda Kay B. Bicar, Derren N. Gaylo (2020). Adapted Bass Diffusion Model for the Spread of COVID-19 in the Philippines: Implications to Interventions and Flattening the Curve. International Journal of Innovation, Creativity and Change. www.ijicc.net Volume 14, Issue 3, 2020.