Using the SIR Model to Analyze the Epidemiological Situation in Kazakhstan and Neighboring Countries

Saya Sapakova¹, Askar Sapakov², and Madina Ipalakova¹

¹ International Information Technology University, Manas St. 34/1, Almaty, 050000, Kazakhstan
² Kazakh National Agrarian Research University, Valikhanov St. 137, Almaty, 050000, Kazakhstan

Abstract
In this paper, we study the epidemic situation in Kazakhstan and neighboring countries, taking into account territorial features in emergency situations. As you know, the excessive concentration of the population in large cities and the transition to a world without borders created ideal conditions for a global pandemic. The article also provides the results of a detailed analysis of the solution approaches to modeling the development of epidemics by types of models (basic SIR model, modified SEIR models) and the practical application of the SIR model using an example (Kazakhstan, Russia, Kyrgyzstan, Uzbekistan and other neighboring countries). The obtained processing results are based on statistical data from open sources on the development of the COVID-19 epidemic. The result obtained is a general solution of the SIR-model of the spread of the epidemic according to the fourth-order Runge-Kutta method. The parameters β, γ, which are indicators of infection, recovery, respectively, were calculated using data at the initial phase of the Covid 2019 epidemic. An analysis of anti-epidemic measures in neighboring countries is given.

Keywords
COVID-19, epidemic prediction SIR model; SEIR model; prediction and analysis; forecasting

1. Introduction
In today’s world, the globalization of travel and trade, free urbanization and environmental issues such as climate change have a significant impact on the spread of disease. Issues of urbanization are becoming increasingly important in connection with the task of entering Kazakhstan into the 30 highly developed countries of the world [1]. Today, high rates of urbanization are observed in Astana, Almaty, and the Almaty region. It is expected that over the next 20 years urbanization will increase by another 10%, which, naturally, will increase the burden on the social sector of the country [2, 3].

At this time, the relevance of the work lies in the fact that obtaining accurate forecasts about the course of the epidemic is of paramount importance for taking the right action, and the society needs adequate data on the current epidemiological situation. A subjective reason for the increased attention to the epidemic consists in the enormous informational pressure on people’s consciousness. For example, there are many sources on the Internet where users can obtain daily statistics for the entire world and for individual regions. This data is necessary and needed, but it is important to interpret it correctly. Thus, the ratio of fatalities to infections varies greatly among countries with roughly the same level of health care. The reasons for this variation are related not only to differences in the real severity of the epidemic situation, but also to the peculiarities of case detection and the practice of registering the virus as lethal. The fatality rate is closely related to the main urbanization factors, and these have worsened due to worldwide quarantine. In the current epidemiological situation, the extent of this deterioration and its duration cannot be assessed. The shutdown of the businesses, the closure of
enterprises increases the number of unemployed and socially vulnerable people. As we know, in countries with good social protection, the number of fatalities is lower. Besides, the number of people on planned treatment has sharply decreased, while it is impossible to evaluate the consequences of this fact. Unlike "regular" seasonal flu epidemics, this virus dramatically increases the burden on the health care system, so realistic predictive estimates are fundamental to understanding the damage and are used as a benchmark for developing new policies to control viral infection of any type and to assess quarantine conditions. It helps in measuring and predicting the health needs of the population and in determining how to allocate and manage health care resources.

In order to predict disease dynamics, assess threats, and select measures to control disease incidence, there is a need for mathematical modeling of the processes that occur during epidemics. Methods of modeling infectious diseases have been actively developed since the beginning of the 20th century. In recent years, the number of works on this topic has grown rapidly due to the deployment of information surveillance systems and the emergence of large volumes of statistics available for analysis. Epidemiological forecasts are performed for different time frames and serve different purposes depending on them.

The object of the study is a modified model of SIR epidemic. The aim of this work is to analyze a modified model of SIR epidemic, which considers the division of infected into two subgroups - quarantine and non-quarantine, the introduction of restrictive measures and the testing factor. In the course of the work, existing epidemic models, and their application to simulate the spread of the current COVID-19 virus were studied. A program was developed in the Python programming language to estimate model parameters and build trajectories. The results of the program are demonstrated using statistical data for one of the regions. It was found that the modified epidemic model is well suited for predicting the spread of the disease due to its simple modification and the possibility to consider different scenarios. The results of the work can be useful in predicting a new wave of disease or used in modeling the spread of a new virus strain. For the epidemiological situation of the country, in addition to internal emigration, immigration between neighboring countries also plays an important role. As we know, Kazakhstan borders Russia to the east, north and northwest, Uzbekistan, Kyrgyzstan and Turkmenistan to the south, and China to the southeast. It also shares the Caspian Sea coastline with Iran and Azerbaijan. Therefore, in this paper, we investigated the course of Covid situation in Kazakhstan and the neighboring countries such as: China, Kyrgyzstan, Uzbekistan, Iran, Azerbaijan, also with Turkey. For Kazakhstan, Turkey is one of the most important and reliable partners in the Eurasian continent, as well as a popular destination for Kazakhstan's tourists due to the intersection of common history and spiritual heritage. Therefore, this paper is devoted to the study of the epidemiological situation in these countries, considering the geographical location and the flow of movement.

Statistical and machine learning methods were applied using epidemiological data, air passenger traffic volumes, vector habitat suitability data, socioeconomic and demographic data for all affected regions of the country. Model performance will be quantified based on model prediction accuracy. The application of ANN in epidemic prediction requires analytical solutions that improve prediction performance. These decisions require the selection of ANN algorithm approaches or techniques. They include decisions about data preprocessing, network architecture or structure, number of input, hidden layers or output nodes, training algorithms, training parameter specification, number of epoch runs, and means of measuring accuracy.

Epidemic forecasting uses past incidence data to predict future epidemic size, peak periods, and duration. This process does not need to understand the details of disease dynamics; it simply aims to accurately predict future epidemics using appropriate methods. Predicting epidemics is critical to implementing effective infectious disease prevention and control measures. Forecasting problems arise in all areas of life. Predicting future events based on past history is performed using various methods in order to plan and evaluate disease control. Predictions allow users to gain insight before making decisions and taking actions that will affect the future course of an epidemic.

As shown by a preliminary study, there is a lack of Kazakhstani model structures and algorithms, brought to computer implementation and tested on the data of a particular city or the country as a whole. In this regard, the practice of creating neural network models to analyze and forecast the development of Kazakh cities should be widely spread, also to use them as a decision support system for administrative authorities in complex epidemiological situations [2, 3].
In this article we study the process of urbanization and its impact on the spread of infectious diseases in general settings. Various types of linear and nonlinear models are used to predict epidemics; these models use epidemiological time series data to predict short- and long-term outcomes from viral infections [4, 5, 6]. Each prediction method aims to achieve high accuracy in predicting future epidemiological conditions. Although the performance measurement tools for prediction models may differ, preventive medical and control decisions to contain an outbreak must be provided with highly accurate predictions. Achieving high-precision predictions depends on using the right tools and analysis methods. Making the right choice among the various methods for analyzing epidemic growth is complicated. Methodologies that must fully account for nonlinearity to represent the reality of disease dynamics with a more accurate prediction and that address underlying urbanization factors are to be preferred often.

2. Methods and models

In the works of the foreign scientists there are several epidemiological models that are used to predict the number of infected people and the rate of lethal outcomes from an outbreak of viral infections [6, 7]. Artificial neural network (ANN) is a convenient apparatus for predicting epidemics, considering factors of different nature. It will help to make analytical important decisions when forecasting epidemics with their help. Application of neural networks for epidemic risk prediction has been previously used for dengue prediction and risk classification [6,7, 8, 9, 10], simulation of epidemiological time series by fusion of neural networks, fuzzy systems, and genetic algorithms [11].

Most studies have considered the multilayer direct coupled perceptron neural network (MLPFFNN), which does not imply that epidemic prediction cannot be modeled according to other network architectures. To improve the accuracy of the ANN, hybrid ANNs are usually used [12].

The recurrent neural network is the most frequently used model for predicting epidemics compared to the non-recurrent architecture [13].

2.1. SIR-models

Mathematical modeling in epidemiology began with the work of D. Bernoulli in 1760, which demonstrated the effectiveness of vaccinating the population against chickenpox [D. Bernoulli, 1760]. Subsequently, a series of mathematical models based on the law of mass balance appeared (see review articles [N. Bacaer, 2011, F. Brauer, 2017] and the literature cited therein). The works of R. Ross in 1911 [R. Ross, 1911], A.J. Lotka in 1920 [A.J. Lotka, 1920] and V. Volterra in 1926 [V. Volterra, 1926] (the "predator-prey" model), led to the creation of the chamber SIR model by W.O. Kermack and A.G. McKendrick [A.G. McKendrick, 1926, W.O. Kermack, A.G. McKendrick, 1927]. In the works of E.N. Pelinovsky and colleagues [E. Pelinovsky et al., 2020, P. Wang et al., 2020, E.M. Koltsova, 2020] apply a generalized logistic equation to model the spread of COVID-19, describing the increase in the population of the disease. The assumption of a singular outbreak peak \( I(t) \) epidemic limits the application of the logistic model to describe long pandemic periods and consider restrictive measures. To account for the incubation period of the COVID-19 course, a modification of the Kermack-McKendrick SEIR-type model (see Figure 3.3b for a schematic of the SEIR model), of which more than 100 models have been developed to date (see, for example, [Y. Chen et al., 2020, M.V. Tamm, 2020, E. Unlu et al., 2020, O.I. Krivorotko et al, 2020, A.I. In these models, the population is divided into groups (besides \( S, I, R; E - asymptomatic carriers, H - hospitalized, C - critical cases requiring ventilator connection, D - dead due to COVID-19, Q - placed under quarantine and others). This allows us to clarify the epidemiological picture in the region by varying a more detailed set of coefficients in the equations. The disadvantage of SIR models is their lack of flexibility - the inability to consider changes in parameters (new virus and strain mutations, restrictive measures, vaccination). At an attempt to introduce these changes into SIR models (e.g., make the transmission rate \( \alpha = \alpha(t) \) variable) [S. Margenov et al., 2021], we are confronted with the nonuniformity and instability of the solution to the inverse problem of identifying this parameter \( \alpha(t) \). Note that SIR models are also used to predict pandemic development control outcomes [C.J. Silva et al., 2021], i.e., a piecewise constant control
function (restrictive measures: mask-wearing, social distance, quarantine) is added to the right-hand side of the equations. However, even in these cases, the problem of refining the coefficients of SIR models remains open and requires the application of inverse problem theory methods. Also, in SIR models it is possible to consider the influence of super predators on the spread of COVID-19 (infected individuals with an increased viral concentration) [F. Ndairou et al., 2020]. However, theoretical determination of this phenomenon requires modeling at the scale of individuals (ASM, see Section 4.3) [14-22]. Further development of mathematical models can be divided into two components: introduction of spatial coordinate into logistic equations and consideration of discrete spatial heterogeneity. In the first case we get a new class of "reaction-diffusion" mathematical models (see Section 3.3), and in the second one - a new approach, in which systems of differential equations are connected in space by a graph structure. All reasonable models of coronavirus epidemic spread are in one way, or another derived from the SIR model. Each person can be in one of the states (vulnerable, infected, sick, cured/immune, dead, etc.), and transition rules between them are defined. A distinction is made between mean-field models and agent-based models.

\[
\begin{align*}
\frac{dS}{dt} &= -\frac{\beta SI}{N}, \\
\frac{dI}{dt} &= \frac{\beta SI}{N} - \gamma I, \\
\frac{dR}{dt} &= \gamma I.
\end{align*}
\]

In which the population of \(N\) individuals is divided into three groups (chambers): \(S\) - susceptible, \(I\) - infected (\(I \ll N\)) and \(R\) - cured and dead, interconnected by probabilistic transitions \(\gamma, \beta\).

The initial conditions for system (1) have the form:

\[S(0) = S_0, \quad I(0) = I_0, \quad R(0) = N - S_0 - I_0.\]

One of the important results of [W.O. Kermack, A.G. McKendrick, 1927] was the introduction of the reproduction index (infectivity)

\[U_0 = \frac{\gamma}{\beta}\]

here, \(\beta\) represents the infection parameter that causes the susceptible population to decrease, when the population of the infected group is considered. If the population of the susceptible group is considered, \(\beta\) then causes the infected population to increase. When the removed population is considered, the removal parameter, \(\gamma\), is introduced, causing a portion of the infected population to move to the removed population. If an individual is infectious for an average time period \(D\), then \(\gamma = 1/D\). \(U_0\) is the most important characteristic of the disease and the parameter of the spread of the epidemic. The virus reproduction index \(U_0\) is defined as the average number of people who are infected by an active infected person in a completely non-immunized environment in the absence of special epidemiological measures aimed at preventing the spread of the disease. Also, \(U_0\) is the stability boundary of the equilibrium state of the SIR system in the absence of infected people. If \(U_0 > 1\), then at the initial stage the number of cases will grow exponentially. If \(U_0 \in (0, 1)\), then a small number of infected people who fall into a completely susceptible population will, on average, not be able to maintain their group, and there will be no epidemic.

Group immunity occurs when the fraction immune is equal to \((1 - 1/U_0)\). If the epidemic develops freely, then the total proportion of those who become sick is equal to \((1 - \exp(-U_0))\).

### 2.2. The model SEIR (Susceptible-Exposed-Infected-Recovered)

In SEIR model, individuals are classified into four infection stages, namely susceptible, exposed, infectious, and removed. All individuals in the population are assumed to be susceptible to the virus before the pandemic begins.
\[
\begin{align*}
\frac{dS}{dt} &= \mu N - \mu S - \frac{\beta}{N} SI; \\
\frac{dE}{dt} &= \frac{\beta}{N} SI - (\mu + \alpha) E; \\
\frac{dI}{dt} &= \alpha E - (\gamma + \mu) I; \\
\frac{dR}{dt} &= \gamma I - \mu R,
\end{align*}
\]

where \( S \) – susceptible individuals, \( E \) – infected individuals without symptoms, \( I \) – infected individuals with symptoms, \( R \) – recovered individuals, \( N \) – whole population.

We have \( S + E + I + R = N \), but this is only constant because of the simplifying assumption that birth and death rates are equal; in general \( N \) is a variable. \( S(0) = S_0, E(0) = E_0, I(0) = I_0, R(0) = R_0 \).

\[
U_0 = \left( \frac{\alpha}{\mu + \alpha} \right) \left( \frac{\beta}{\mu + \gamma} \right),
\]

where \( \beta, \gamma, \mu \) are the rates of infection, recovery, and mortality, respectively. The latency period is a random variable with exponential distribution with parameter \( \alpha \) (i.e. the average latency period is \( \alpha^{-1} \)).

The classical SIR is written as a system of differential equations. This implicitly assumes that the lifetime in each state is exponentially distributed, and its variance is equal to the mean. This does not agree well with experimental data, e.g., on the distribution of the incubation period, the time between successive infections (serial interval), the duration of the disease, etc. Due to this fact, many models (the Neyer model, perhaps, remains a rare exception) use integral operators, most often with gamma-distributed kernels, rather than local in time differential equations, to describe transitions. This slows down the counting, but not significantly.

3. Implementing the SIR model

In this study, we used the Runge-Kutta numerical method to solve the system of differential equations. As can be seen Figure 1, the number of those infected increases over a period of time and then eventually decreases as people recover/die from the disease. The susceptible portion of the population decreases as the virus is transmitted and eventually goes to an absorptive state of 0. The opposite happens for the cured/dead case. Note that other initial conditions and parameter values will lead to different scenarios, feel free to play with these numbers to explore the system.

![Figure 1: Fraction of population over the world](image)

The SIR model used is strictly theoretical; we need a real approximation of COVID-19 propagation to extract the information and understand the virus transmission. Consequently, we need to extract the parameters \( \beta \) and \( \gamma \) for each case to predict the system values.
In most of the studied works on the topic, the value of the parameter $I_0$ (the number of infected at the time of the beginning of the epidemic) is taken as approximately known, and the value of the parameter $S_0$ (the initial number of susceptible) is often uncertain.

**Figure 2:** Fit of SIR model for Kazakhstan infected cases

**Optimal parameters:** beta = 1.0 and gamma = 1.0

**Figure 3:** Fit of SIR model for Turkey infected cases

**Optimal parameters:** beta = 1.0 and gamma = 1.0
Figure 4: Fit of SIR models for Russia and Uzbekistan

Figure 5: Fit of SIR model for China infected cases

Figure 6: Global confirmed cases excluding China

In the following chart, we examine the analysis for 9 countries that are geographically located in the border areas with Kazakhstan.
As you know, the countries of the region (Kazakhstan, Kyrgyzstan, Tajikistan) directly border China, where the virus first appeared, as well as (Turkmenistan) Iran, where the epidemic has become extremely widespread. If we consider the actions of the leadership in individual countries at the initial stage of the epidemic, then, for example, in Turkmenistan, mass quarantine was not announced in the country, therefore, in March-April 2020, Turkmenistan continued to hold mass events and celebrate holidays with a large crowd of people (Novruz celebration). And, as of April 28, 695 infected people were registered in neighboring Kyrgyzstan, therefore, starting from March 25, the President of Kyrgyzstan declared a state of emergency in Bishkek, Osh, Jalal-Abad, Nookat, Kara-Suu and Suzak regions.

In Uzbekistan, the first case of coronavirus infection was recorded on March 15. As of April 28, 1904 cases of infection have been registered. Since March 16, Uzbekistan has closed communication with other countries, mass events and meetings in state bodies have been canceled, educational institutions have been transferred to remote work, a regime of self-isolation and quarantine has been established, and residents of the republic are prohibited from using personal vehicles.

In our country, the first case of infection with the coronavirus COVID-19 was detected on March 13. Since March 16, a state of emergency has been in effect in our country. As of April 28, according to official data, 2835 infected people were registered in Kazakhstan. The epicenters of the spread of coronavirus are Nur-Sultan, Almaty, Shymkent. In this regard, all foreign flights were canceled and arriving passengers were sent to quarantine. As we can see, during that period, similar measures were taken in neighboring countries to reduce the degree of infection with this virus. This can be seen from the results of this study, for these countries we received similar graphs, unlike Russia, China and Turkey where the population there is very different from the above countries.

4. Conclusion

Especially the Kazakh government has taken many emergency actions in order to help reduce the contagion probability and cure the infectious individuals. On the other hand, the imported disease cases unavoidably affects the exploration of COVID-19 in Kazakhstan. All lead to it that the system for this epidemic transmission should be considered within the on-extensive application rather than the classical ideal one. We have used the SIR model to investigate the time evolution of infectious individuals in Kazakhstan and 8 neighboring countries with it. Using the SIR model, the graphs in Figures 2-5 are obtained and the optimal values of the beta and gamma parameters for these countries are calculated.

We can also notice that according to the graphs of neighboring countries with Kazakhstan, the epidemiological situation is similar, except for China and Turkey, which differ due to the population. The results of this work can be used to obtain assessments in the country and abroad, to apply anti-epidemic measures against COVID-19, which include a set of measures aimed at preventing the importation and
spread of infection, and are organized by authorized territorial bodies, taking into account the factors of urbanization, external and internal migration of Kazakhstan.

5. References

[1] Concept for Kazakhstan’s entry into the 30 most developed countries in the world. Approved by the Decree of the President of the Republic of Kazakhstan dated January 17, 2014 No. 732.