Study on the Risk of Transmission of COVID-19 Based on Population Migration

Xiaofeng Hu\(^1,2\)

\(^1\)School of Management/Institute of Finance Engineering/School of Emergency Management, Jinan University, Guangzhou, 510632 Guangdong, China
\(^2\)Guangdong Emergency Technology Research Center of Risk Evaluation and Pre-Warning on Public Network Security, Guangzhou, 510632 Guangdong, China

Correspondence should be addressed to Xiaofeng Hu; hxfsummer@stu2017.jnu.edu.cn

Received 19 April 2022; Revised 31 May 2022; Accepted 9 June 2022; Published 30 June 2022

Academic Editor: Kuruva Lakshman

Copyright © 2022 Xiaofeng Hu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Since the emergence of COVID-19, migration of people has transferred the virus to new locations, causing the epidemic to expand, and local governments have put in place control measures to prevent the virus from spreading further. As of January 24, 2020, we calculated the population immigration from Wuhan to the rest of mainland China using migration statistics from the Gaode Map. In addition, we utilized machine learning methods to simulate the curve of the COVID-19 epidemic in different regions and over different time periods. Furthermore, we used machine learning methods to simulate the COVID-19 epidemic curve in various regions and over various time periods. Based on the Wuhan exodus, we built a migration transmission risk model. From January 24 to February 19, 2020, we predicted the location, severity, and timing of epidemics in various parts of mainland China. We showed how discrepancies in model predictions might be utilized to measure transmission load in different parts of the country. Higher transmission risk indices suggest more community transmission in the region. According to the study, states with lower transmission risk indices but fewer cases than expected may have taken highly effective public health measures.

1. Introduction

In the past few months, COVID-19 death tolls have risen across 114 countries following the spread of SARS-CoV-2. In March 2020, the World Health Organization declared the outbreak to be a pandemic. SARS-CoV-2-caused COVID-19 has been treated by local and national government agencies in a significant way.

Some governments have implemented regulations akin to “epidemic lines” to prevent the transmission of viruses from previously affected areas to other nations and territories. In addition, epidemic lines are established to prevent the importation of infectious diseases from high density areas. Blockade restrictions are intended to keep sick visitors from moving to areas with high susceptibility populations, where new transmission chains could form. They are, however, almost always reactionary, and only after a number of communication channels have been created. Therefore, the effectiveness of blockade control in preventing or delaying outbreaks in other areas remains to be studied. Previous studies have shown that “urban blockades” consisting of various nonpharmaceutical interventions can slow down the growth of outbreaks for a certain period of time [1–4]. In Italy and Spain, the first lockdown of citizens and the recommendation to keep social distance in public places were introduced. Initial analyses in Italy and Spain showed that the restriction of social activities during the first urban quarantine led to a significant reduction in the number of daily confirmations, mortality, and admissions to intensive care units [5]. As of February 11, 2020, Kucharski et al. [6] estimated the prevalence of daily COVID-19 outbreaks in Wuhan. Although the incidence of infectious diseases has decreased, the growth of this epidemic continues unabated. As a result, a second blockade has further restricted the movement of domestic residents and shut down all unwanted commercial activities. Although the result of these
multiple policies was to slow the growth of the epidemic, the one confirmed case only reflected a diagnosis after a period of incubation, so its overall effect has not been proven. Other techniques have also been proposed to predict epidemic levels in advance in specific populations—whether using online search behavior [7–9]. Alternatively, network sensors could be used (i.e., monitoring those at higher risk of disease due to their network location) [10].

On 23 January 2020, China imposed a blockade on Wuhan as a result of the COVID-19 outbreak, preventing anyone from leaving the city unsupervised. The provinces of Hubei were also subject to similar controls by the end of January 2020 [11]. This regulation, which went into effect on the eve of and throughout the Chinese New Year, is also among the world’s major yearly population shifts [12]. Other public health measures, such as "physical distance," have been established in China at the same time [13]. Although early control measures in Wuhan may have had a stronger impact, it may be impossible to enforce such highly disruptive lockdown regulations in the early phases of an outbreak, before isolated spread to other cities, particularly locations with significant levels of interstate transportation. The data imply that endemic infectious diseases in some cities began before Wuhan’s blockade controls were implemented. The COVID-19 had not been confirmed to be transferred between humans until January 8, 2020 [14]. In the months following the outbreak, COVID-19 outbreaks in other Chinese cities continued to decline, in large part due to additional public health actions implemented to reduce the outbreak’s transmission, that is, to reduce the number of illnesses to one or less [6, 15, 16]. Because of the modest amount of population movements worldwide, blockade control may be more effective in delaying outbreaks [13, 17], and Tian et al. found that the same is true for low-traffic destinations within China from Wuhan, where outbreak management can better delay or stop outbreaks. Other recent COVID-19 investigations have estimated case output during the present outbreak using historical population migration data (e.g., spring migration data from past years) [17–21].

COVID-19 outbreak in Wuhan was reduced through nonpharmacological measures, for example, school closing hours and working distance between work places are significant factors to consider [22]. In response to other respiratory infectious disease epidemics, these measures drastically changed the previous pattern of mixing age-specific groups of the population [23, 24]. Although blockade management has likely contributed to lowering infection export outside of Wuhan, and outbreaks in other cities were postponed [25], alterations to the mixing pattern have had an impact on the outbreak’s course in Wuhan. This study will utilize Wuhan as an example to assess whether distance measures have an impact on the growth of the COVID-19 outbreak, in the hopes of providing some insight for other parts of the world. One of the studies was conducted on COVID-19 data to analyze the risk of disease transmission considering the viral load. The research emphasized on exploring the various factors responsible for the transmission of COVID-19. Regression analysis was conducted to assess risk transmission and understanding of incubation dynamics pertaining to the disease [26]. Similar studies were conducted in focused on identifying novel factors relevant to transmission of COVID-19 disease. The study used regression analysis model to predict COVID-19 deaths in various countries [27].

Early research into school closures, parental notification, and drug-screening campaigns during COVID-19 suggest that large-scale strategies like school closures, case isolation, household quarantines, internal lockdowns, and border controls can slow the spread of infectious diseases and/or shorten the duration of outbreaks. Given that a COVID-19 epidemic in humans would have a substantial influence on human life and society’s security, the goal of this thesis is to propose a set of effective control techniques to slow the virus’s spread.

Though COVID-19 remains an active pandemic, several cities are showing signs of recovery. The impact of urban closures on population mobility and the COVID-19 outbreak must be assessed. Since the pandemic in China has been contained since March, it is essential that the outbreak is investigated in terms of population movements and NCD efforts (e.g., embargo control). The goal of this study is to see how effective blockade restrictions in Wuhan (the COVID-19 pandemic’s epicenter) as a measure to reduce morbidity and delays outbreaks, and the actions taken aimed at reducing outbreaks of disease in other large, well-connected cities on mainland China. To conduct simulations, this paper uses publicly available location-based service- (LBS-) based mobility data supplied by Gaode Map Migration Data. This research forecasts outbreak cases in a variety of places around the country in order to estimate the likelihood of long-term local transmission.

2. Data Description

2.1. Population Migration Data. Cell phone use among Chinese 15–65 years old is close to 100%, according to prior statistics, and this figure covers the whole population in this age group. In light of the COVID-19 outbreak, tracking population mobility in China and elsewhere across the globe is especially important. The study of population migration can help researchers better understand the distribution of epidemics in space and related policies. Modeling the transmission of diseases, detecting patterns and hotspots, and predicting future outbreaks are possible through the analysis of population migration data [28]. A correlation exists between population mobility and COVID-19 transmission, which is caused by human movement. During the incubation period, infected individuals are likely to transmit to other closely related persons, with an average incubation period of 5.1 to 2 weeks, as observed in 99 percent of COVID-19 patients [29]. Early patients with COVID-19 may have no symptoms or very small signs, and statistically, 86% of patients go undetected [20]. Therefore, these characteristics of the COVID-19 outbreak are extremely well predicted and forecasted in both regional and international targets. COVID-19’s occurrence and spread across the globe can help us better understand the virus’s epidemiological characteristics and the nonpharmaceutical intervention
(NPI) measures now in use to control and contain the disease. The use of cordon-style measures alone in the vicinity of COVID-19 outbreak centers and other easily accessible regions with high population concentrations will have little effect on the epidemic’s epidemiological trend. In cities, strict nonpharmaceutical treatments (NPIs) play a larger role in reducing illness incidence and healthcare system burden.

COVID-19 was discovered on the night before China’s Lunar New Year, and it is linked to the annual Chinese New Year migration (which may involve up to three billion people). Given that Wuhan is a major rail and air traffic center in China, the risk of COVID-19 spread is extremely concerning. This research uses the Gaode Map traffic big data to synthesize the cell phone activities (including geographic position) of cell phone users across the country to quantify the population outflow in the Wuhan area before and after the quarantine on January 23, 2020. Based on the actual data collected from January 1 to March 31, 2020, the following is the actual out-migration intensity from Wuhan to other cities, and the time period studied included the time when the epidemic in question first appeared and the population migration before Lunar New Year (ending on January 24, 2020), and the data were daily, and the indicators of the data included in-migration and out-migration, characterized by time, departure province, departure province code, leaving city, leaving city code, arriving province, arriving province code, arrival city, arrival city code, willing migration index, and actual migration index. The study period of this paper coincides with the Chinese New Year’s Eve, when a massive migration (roughly 3 billion people) for the Spring Festival is approaching.

The main basis of this study is the relative migration intensity of the population rather than the actual population size from Wuhan to each city, so the methodology used in this paper does not need to examine the overall population migration. The reason for this is that the main study in this paper employs the relative actual migration indices from the Gaode population movement statistics. They are represented as a proportion of the total demographic outflow in these indexes. Since Wuhan is the largest city in Hubei Province, the main analysis uses population outflows from Wuhan, the city with the biggest share in COVID-19’s early years. The massive outflow of COVID-19 virus raises a risk of spreading to neighboring places beyond Hubei Province. As a result of the survey, the spread of COVID-19 in Hubei cities outside Wuhan is much less affected by population movement than that in Wuhan. The number of confirmed patients from the Wuhan area was not used in the estimate of the impact of the COVID-19 epidemic. Therefore, any inaccurate statistics or changes in reporting methods in Wuhan do not affect the analysis in this paper.

2.2. Different Types of Wuhan Population Outflow Data. A comparison of mobility in Wuhan before and after the implementation of the closure measures found that the number of in-migrants in the city fell by more than 90% [13]. Although Fang et al. [30] showed a 56.4% decline in outflows from Wuhan, they also reported a 76.6% decline in inflows and a 54.2% decline in the city. Kraemer et al. [31] noted that using 8 days as the standard difference in the mean incubation period, confirmed cases with definite cases arriving in Wuhan before January 31 cases decreased from 515 to 39. Based on the study, migration and confirmed COVID-19 cases have a positive correlation in the initial period, suggesting that the prohibition of population movement is an effective preventive control tool, as an effective preventive control measure [30]. However, millions of people were evacuated from Wuhan before the outbreak was controlled and the COVID-19 virus had spread to other cities [32]. According to studies, the ban on migration caused the first cases to be detected outside the border cities to take 3 to 5 days to appear [13]. Although some success has been achieved in stopping internal transmission [33], the immigration ban has been more effective in limiting the spread across international borders [17]. Based on features like as the epidemiological incubation time and the rate of asymptomatic individuals, however, blockade measures alone cannot entirely prevent local outbreaks of COVID-19.

We plot the outflow population intensity of Wuhan (within Hubei Province) and the outflow population intensity of Wuhan (outside Hubei Province) for selected major cities (as in Figures 1 and 2). In addition to Lunar New Year’s Eve on January 24, which is when most outbound travel will be wrapped up, and Wuhan implementing quarantine procedures on January 23, there are also significant dates. The number of in-migrants, out-migrants, and intracity migrants in Wuhan decreased by 39.56 percent, 12.9 percent, and 5.93 percent, respectively, on January 23, 2020. By contrast, consumer spending fell 49.94 percent more than on January 22, 2020, 22.92 percent, and 65.04 percent on January 24, 2020, compared to January 23, 2020. As of 10:00 on January 23, 2020, Wuhan and two adjacent cities have had a complete stoppage of population outflows after the installation of the newest blockade control system. During the evening of January 24, 2020, additional barriers were placed in 12 more cities in Hubei.

As can be seen in Figures 1 and 2, the outflow from Wuhan to the rest of Hubei (as shown in Figure 1) exceeded the movement to the provinces by more than four times during the previous peak spring holiday travel period. As of 10:00 a.m. on January 23, 2020, the residents of Wuhan have been reduced to only a small number of people flowing into nearby cities (as shown in Figure 1). In Figure 1, the first peak can correspond to the winter vacation of college students in Wuhan (about 1 million), followed by the spring migration out of the city. According to statistics, there were 75,002 verified cases in mainland China as of February 19, 2020, 29,975 cases outside Wuhan, and 2,118 deaths (China Health and Wellness Commission). People can become infected with the virus through contact with patients in the same city or in another city. Infections from the virus in other cities can affect local public health agencies and citizen awareness. Increased protection can slow the spread of the virus.
3. Analysis of the Relevance between Population Migration and Confirmed Cases

3.1. Relevance of Confirmed Cases with Population Exodus from Wuhan to Hubei Province. In the preliminary analysis, daily population outflow data from January 24 to March 3 were used to calculate how the correlation between confirmed cases and population outflows in Wuhan and Hubei Province (excluding Wuhan) increased over time (see Figures 3 and 4).

Population migration in Wuhan is assumed to spread the virus to other regions, causing local outbreaks (e.g., through importation or community spread [14, 34, 35]). Indeed, research has found a clear link between population migration and infection rates in different locations (see Figures 3 and 4). From January 24 to March 3, 2020, there was a substantial correlation between total migration and total movement in Wuhan, which grew with time from \( r = 0.502 \) on January 24, 2020, to \( r = 0.892 \) on February 11, 2020, and subsequently to \( r = 0.911 \) on February 15, 2020, confirming the hypothesis (see Figure 3). In Hubei Province (excluding Wuhan), there was a similar, but substantially weaker, association between outflow and confirmed cases per city (see Figure 3); from \( r = 0.558 \) on January 24, 2020, to \( r = 0.745 \) on February 16, 2020.

Using Figure 3, we can see that the number of new confirmed cases is fairly strongly related to the daily population inflow from Wuhan (cumulatively to March 3, 2020), a time-series displaying the number of cases reported and the total number of cases verified. During the late stage of the outbreak, from \( r = 0.522 \) on January 24 to \( r = 0.952 \) on February 19, the association between the number of victim cases and the population exodus from Wuhan rapidly grew, while the correlation between Wuhan COVID-19 cases and Wuhan COVID-19 cases gradually decreased, which indicates that community infection predominates in the late stages.

Figure 4 depicts the relationship between exodus from Hubei Province (excluding Wuhan) and confirmed cases, which includes the number of new confirmed cases (to March 3, 2020), the cumulative number of confirmed cases (to March 3, 2020), and the population influx from Hubei Province (excluding Wuhan) over time. From Pearson’s \( r = 0.558 \) on January 24, 2020 to \( r = 0.744 \) on February 16, 2020, the association between exodus from Hubei Province and the number of cases grew over time. Figure 4 shows the link between exodus from Hubei Province (excluding Wuhan) and daily instances that have been verified and cases that have been confirmed in the past through time, with correlations ranging from 0.558 on January 24, 2020 to 0.744 on February 16, 2020. The association has weakened in recent days, showing that there are no new instances in more than 90% of cities outside of Hubei.

3.2. Correlation Analysis of Different Time Windows of Population Exodus. After a period, the daily population inflow...
exodus from Wuhan and the number of confirmed cases were correlated using Pearson’s correlation method (Figures 5 and 6). As time went by, the largest correlation was found between the number of diagnosed cases in different time windows and the population outflow from Wuhan through January 24. Because there was little population movement across the country from January 24 to March 3, the variable of population outflow was a lagged variable. The association is true even when other time periods for population mobility are considered, as seen in Figures 5 and 6. In this case, \( t = 3 \) indicates that the correlation is determined by measuring the number of diagnosed cases annually on January 24, 2020, versus the number of daily outflows from Wuhan on January 21, 2020.

Figures 5 and 6 show the Pearson correlation of Wuhan outflow with new diagnosed cases, in addition to the link between the total number of diagnosed cases and the Wuhan exodus. The correlation with new confirmed cases continued to increase from January 20 to the peak on January 24, after which the correlation continued to decrease, while the correlation between Wuhan outflow and cumulative diagnosed cases also started to increase from January 20 and remained stable in the subsequent period. The increase in diagnosed cases could be due to spread of the disease by the individuals travelling from Wuhan to other provinces. However, the decrease would be an effect of the lockdown which ensured containment of the disease. The travel patterns have revealed patterns relevant to the spread of the disease primarily by large pools of migrant workers.

4. A Transmission Risk Model Based on Population Migration

4.1. Transmission Risk Model. Population movement provides substantial benefits for policy formulation, as incorrect predictions can yield adverse consequences: lack of response will lead to disease spread, while overreaction will produce medically, socially, and financially ineffective policies. Figures 3 and 6 show that population migration from Wuhan has a strong correlation with the total number of confirmed cases at the destination, and that this correlation grows over time, peaking on January 24, so the overall population outflow from Wuhan until January 24 is used as an influencing factor in this paper’s model. In addition, this paper also considers factors such as the distance between the destination city and Wuhan, the population size, and the GDP of each city that may also affect the change of confirmed cases in that city.
This paper considered several different machine learning approaches for propagation risk prediction. Machine learning algorithms have been predominantly used for analyzing transmission and diagnosis patterns of the disease. The disease is novel; and hence, the genetic information is unknown. Thus, there was lot of dependency on mathematical and data-driven models for the prevention and control of the same. Also, accurate diagnosis and preventive measure were extremely important for the containment of the disease. Deep learning and machine learning algorithms served to fulfill this purpose [36]. Studies have been conducted wherein deep learning and machine learning models were tested using dataset collected from seven countries which were impacted severely by the disease. The results revealed the superiority of the hybrid deep learning models in predicting the disease efficiently and accurately in comparison to the other state of the art approaches [37–39]. To implement the propagation risk analysis based on population migration, classical ML algorithms such as decision tree algorithm, random forest algorithm, adaboost algorithm, GBT algorithm, ExtraTrees algorithm, CatBoost algorithm, K-nearest neighbor (KNN) algorithm, and LightGBM algorithm are used. The decision tree algorithm incorporates use of multiple algorithms to split a node into its subnodes which increases the homogeneity of the resultant subnodes. The random forest algorithm is a supervised machine learning algorithm that helps in resolving classification and regression problems. The algorithm uses a decision tree-based approach for the various samples and considers the majority of the votes for classification results and averages for regression analysis. The adaboost algorithm is an adaptive boosting technique which uses an ensemble method and decision tree-based approach. The trees used in adaboost are called decision stumps. The gradient boosting is also a machine learning approach which uses regression and classification methods. The prediction model is developed in the form of decision trees wherein if the decision tree is a weak learner, the resultant algorithm is known as gradient boosted trees. The extremely randomized trees or ExtraTrees algorithm uses an ensemble of decision tree-based approach, namely, bootstrap aggregation and random forest. The method uses a large number of unpruned decision trees in order to train the dataset. The predictions are made by computing an averaging the predictions generated by the decision trees while performing regression. Similarly in case of classification, the majority of the votes are considered. The CatBoost algorithm is also an ensemble machine learning algorithm that
uses gradient boosting approach while developing the decision trees. This algorithm is also used for solving classification and regression problems. The KNN algorithm is a machine learning algorithm that implements supervised learning technique. The algorithm classifies new data points based on the similarity with the available data. This algorithm can be used to solve classification and regression problems. The LightGBM algorithm uses a gradient boosting approach based on decision tree to ensure increase in model efficiency and reduction in memory usage. The algorithm uses two techniques, namely, the gradient-based one side sampling and exclusive feature bundling to yield efficient and accurate results. The algorithms mentioned are mostly gradient boosting algorithms that use ensemble decision tree-based approach to generate accurate results ensuring enhanced efficiency. This could be considered as the commonality and reasoning behind selection of these algorithms. The implementation performance of each algorithm is assessed using five model evaluation coefficients: MSE, RMSE, MAE, MAPE, and $R^2$. The values of the specific results are shown in Table 1 below:

<table>
<thead>
<tr>
<th>Models</th>
<th>MSE</th>
<th>RMSE</th>
<th>MAE</th>
<th>MAPE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision tree algorithm</td>
<td>1.879</td>
<td>1.371</td>
<td>0.795</td>
<td>106.726</td>
<td>0.89</td>
</tr>
<tr>
<td>Random forest algorithm</td>
<td>2.968</td>
<td>1.723</td>
<td>0.943</td>
<td>89.342</td>
<td>0.93</td>
</tr>
<tr>
<td>adaBoost algorithm</td>
<td>8.333</td>
<td>2.887</td>
<td>2.466</td>
<td>81.25</td>
<td>0.809</td>
</tr>
<tr>
<td>Gradient boosting tree algorithm</td>
<td>1.025</td>
<td>1.012</td>
<td>0.307</td>
<td>70.207</td>
<td>0.972</td>
</tr>
<tr>
<td>ExtraTrees algorithm</td>
<td>3.053</td>
<td>1.747</td>
<td>1.087</td>
<td>97.853</td>
<td>0.877</td>
</tr>
<tr>
<td>CatBoost algorithm</td>
<td>1.508</td>
<td>1.228</td>
<td>0.704</td>
<td>86.797</td>
<td>0.966</td>
</tr>
<tr>
<td>K-nearest neighbor algorithm</td>
<td>12.974</td>
<td>3.602</td>
<td>1.646</td>
<td>97.724</td>
<td>0.588</td>
</tr>
<tr>
<td>LightGBM algorithm</td>
<td>10.857</td>
<td>3.295</td>
<td>1.374</td>
<td>100.663</td>
<td>0.716</td>
</tr>
</tbody>
</table>

Table 1: Table of model performance evaluation coefficients.

Figure 7: January 25 spread risk index of some cities.

4.2 Transmission Risk Index. Using the predicted cases from the transmission risk model, a daily transmission risk index was constructed for each city regarding the variation between the number of confirmed and predicted cases on any certain day, as follows:

$$I = y_i - y_i'.$$  \hfill (1)
Among them, $I$ indicates dissemination risk index, $y_i$ indicates the number of instances that have been confirmed, and $y'_i$ indicates the estimated number of confirmed cases. A city with a greater transmission risk index has more cases spreading throughout the community, and conversely, a lower transmission risk index may be of interest because they are implementing successful public health measures (or most likely reporting less accurate data). For example, Figures 7–9 show that in cities with high transmission risk index values on January 25, January 26, and January 29, the transmission risk index was indeed associated with an impending quarantine.

Figures 7–9 show the cities with high transmission risk indices on January 25, January 26, and January 29, 2020, respectively. The propagation risk model prediction results can be used to determine which places have significant differences. The transmission risk model describes the extent of imported infection prevalent at any given time, as it forecasts the number of cases in a city based primarily on the outflow of population from Wuhan. This relates to the influx of cases and primary transmission of the virus. To be more specific, wider difference between the verified and forecasted cases highlights the superiority levels of dissemination in the community. The communal dissemination risk index was stronger in the left city. On January 25, 2020, the results of the model calculations revealed that Huanggang had the highest probability of community transmission; the authorities declared a full quarantine of the city on the same day as Wuhan. Jingmen and Xiaogan had the highest transmission risk index on January 26, 2020, and they both adopted lockdown control measures on January 24. It was Wenzhou that had a higher transmission risk index on January 29, 2020, and Wenzhou adopted lockdown control measures on February 2.

Next, using a similar logic as above, we compare actual confirmed and predicted to measure the risk of transmission of the outbreak. Here, we examine how the predicted and actual cumulative confirmed cases change over time in different regions (Figure 10 and 11). The disparity in the growing trend between the number of forecast and confirmed cases could indicate a rise in COVID-19 community transmission. In fact, the transmission risk model identified a list
most of them adopted the corresponding control measures at the same time. On the other hand, cities with actual confirmed cases lower than predicted may have more successful public health measures.

Figure 10 presents the predicted versus actual cumulative cases for selected cities in Hubei Province, where the difference in growth trends between predicted and confirmed cases could predict a higher level of community transmission of COVID-19. After February 13, 2020, there was a discontinuous jump in confirmed cases in some cities, reflecting changes in local government standards for confirmed counts; in these cities (within Hubei Province), the clinically diagnosed cases began to be included in the total confirmed case count.

Figure 11 presents the predicted versus actual cumulative cases for selected cities outside Hubei Province (cities with more confirmed cases), and we can see that the difference between predicted and actual cumulative confirmed cases is large at first, but the model fits better and better as time flows. When people migrate, they take infectious diseases with them, and cities with more outgoing travelers from Wuhan may experience high incidence earlier. Therefore, population migration predicts the trend of an infectious disease to some extent, so that data analysis techniques can be used to control the disease. Based on the general population outflow from Wuhan, sustained local transmission may have occurred in numerous cities in early January, several weeks before the quarantine restrictions were implemented. Since the start of containment in Wuhan, the number of local infections in Beijing, Wenzhou, and Chongqing may have surpassed 500. The outbreaks in Hangzhou, Ningbo, and Zhengzhou began later and were smaller at the time of closure than those in Beijing, Wenzhou, and Chongqing, reflecting the relative quantity of population transfer from Wuhan. Furthermore, there was no evidence that the spread...
in the cities was increased by the Spring Festival period. The higher liquidity in 2020 compared to 2019 prior to the Lunar New Year can be attributed to several causes, incorporating year-to-year variations and presumably COVID-19-related difficulties, such as those linked to the epidemic’s rapid expansion and impending closure measures.

Although the number of cases can be predicted by utilizing data presented to recalculate the model’s assumptions, the analysis in this work focuses on the comparative severity of outbreaks in each location, rather than on individual absolute numbers. This is due primarily to the cumulative population outflow from Wuhan, which is the primary contributor to the relative distribution of cases across geographical locations and time. Another advantage is that if the temporal distribution of population mobility can be reliably represented, the nonsynthetic flaws of the COVID-19 case count are generally inconsequential. Standard infectious illness models are built based on criteria such as population mix, population interval size, and virus characteristics. The logic of the population migration-based transmission risk model developed in this work differs dramatically from traditional infectious disease models. The transmission risk model established in this study can describe the distribution of epidemics concisely by assuming that the risk derives from population mobility.

5. Conclusion
This paper investigates the study of COVID-19 transmission risk model based on population migration, this paper uses detailed cell phone geolocation data to calculate the total population migration, the flow value of GOD migration data is presented in the form of relative index, firstly, the correlation between population migration and confirmed cases was analyzed, and it was concluded that the total population outflow of Wuhan city before January 24 had the highest

Figure 11: Predicted and actual cumulative cases in selected cities outside Hubei Province.
correlation with confirmed cases. Therefore, factors such as the total population exodus from Wuhan in the past January 24, the distance between the destination city and Wuhan, the population size, and GDP of each city were considered to develop a classifier to simulate the impact of population outflow on infection, and the transmitting risk index was determined due to the difference between the expected and observed confirmed data, and it was discovered that areas with a higher transmission risk index had more community transmission, and areas with a lower transmission risk index had less community transmission (data may not be accurately reported or may be inaccurate).

The research in this paper differs significantly from earlier research in that it employs real-time data on population movements, focuses on overall population movements rather than personal tracking, and adopts a unique modeling strategy. The gradient boosting algorithms used in the paper follow ensemble approach and hence makes it easier to interpret the results generated from the model and yield accurate predictions. Also, the possibility of overfitting also gets eliminated. The method used in this paper is applied to all population migration data. Where suitable data are available, the approach can also be deployed on-the-fly to enable for policy proposals, such as the deployment of resources and manpower to certain physiographic areas depending on the expected degree of outbreaks. Based on our findings, it can also provide a dynamic performance indicator that, when compared with immediate infection reports, can determine which region is at greater risk of virus transmission or more effective measures. The paper generates transmission index which is considered as the primary metrics for evaluation of the model. However as part of future study accuracy, precision and various other metrics could be included to compare the proposed work with other state of the art approaches.

Data Availability

The datasets used during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The author declares that he has no conflict of interest.

References


