A machine learning based forecast model for the COVID-19 pandemic and investigation of the impact of government intervention on COVID-19 transmission in China

Xingcheng Lu
   The Hong Kong University of Science and Technology

Dehao Yuan
   The Hong Kong University of Science and Technology

Wanying Chen
   The Hong Kong University of Science and Technology

Jimmy Fung (✉ majfung@ust.hk)
   The Hong Kong University of Science and Technology

1 **A machine learning based forecast model for the COVID-19 pandemic and investigation of**

2 **the impact of government intervention on COVID-19 transmission in China**

3 Xingcheng Lu [a], Dehao Yuan [b], Wanying Chen [a], Jimmy C.H. Fung [a,b,*]

4 [a] Division of Environment and Sustainability, Hong Kong University of Science and Technology, Clear Water Bay,

5 Hong Kong SAR, China

6 [b] Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong

7 SAR, China

8 correspondence to: Jimmy C.H. Fung (majfung@ust.hk).

9 **Abstract**

10 The coronavirus disease 2019 (COVID-19) pandemic has killed over 0.3 million people, disrupted people's

11 normal lives, and severely restricted economic activities globally. In this work, a model for the next-day

12 COVID-19 prediction in China was built based on the ensemble back-propagation neural network machine

13 learning technique, Baidu migration index, internal travel flow index, and confirmed cases from the

14 previous days. The 10-fold cross-validation results showed that the model performs well in estimating the

15 next-day confirmed cases with a correlation coefficient of 0.97. To investigate the impacts of government

16 interventions on the spread of this new coronavirus infection, the Baidu migration index and internal travel

17 flow index multiplied by a factor of two were input into the trained machine learning model, and the results

18 showed that the confirmed cases in the analyzed cities would increase dramatically. The correlation between

19 the daily new confirmed cases and some meteorological factors were also analyzed, and the results revealed

20 that these factors are not dominant in influencing the spread of this disease. Overall, the results of this work

21 suggest that besides early diagnosis and medical treatment, a city lockdown policy is one of the most

22 effective methods in suppressing the rapid spread of COVID-19.

23 **Keywords:** COVID-19, Baidu Index, Machine Learning, Prediction, Intervention

24

25

26

27

28

29

Introduction

In December 2019, several cases of unidentified viral pneumonia were observed in Wuhan, China (Lu et al., 2020). Later, the patients were identified to have been affected by a new coronavirus. As the disease spread with an unprecedented speed over China, on January 30, 2020, the World Health Organization (WHO) officially declared the outbreak of "2019 novel coronavirus (2019-nCoV)" as a "public health emergency of international concern" (Sohrabi et al., 2020). The coronavirus disease 2019 (COVID-19) has been detrimental not only to China but the whole world (Wang et al., 2020), and as of Aug 4, 2020, over 18.2 million people globally have been reported to be infected by this serious disease (Worldmeter, 2020). Around 0.692 million deaths were confirmed to be due to 2019-nCoV infection, and the mortality rate of this disease has reached 3.8%. Unfortunately, no miracle drug or vaccine has yet been found or developed after 7 months since this disease was first identified. As the most crucial black swan event of 2020, the COVID-19 pandemic has severely restricted the global economy and has sharply reduced goods consumption in some countries (Fernandes, 2020). Most social and commercial activities have been banned in many Western countries, such as Italy, Spain, and the U.S.A. (Nicola et al., 2020). This pandemic has dragged China's GDP growth into the worst recession since 1990 (Kishan, 2020); in the first quarter of 2020, China's economy shrank 6.8% compared with the previous year (Tan and Cheng, 2020).

Because of this significant disruption caused by the COVID-19 pandemic, it is crucial to identify the key factors affecting the infectiousness of 2019-nCoV and develop a reliable model for near-real-time forecasting (Petropoulos and Makridakis, 2020, Wynants et al., 2020). Since the wider outbreak of this disease in January, substantial efforts have been devoted globally to unveil its mysteries (Fanelli and Piazza, 2020, Wu et al., 2020, Yang et al., 2020). Using the Glob al Epidemic and Mobility Model, a study reported that the travel restrictions in Wuhan on January 23, 2020, had marked effects on reducing the spread of the new coronavirus infection globally, with an 80% importation reduction until mid-February (Chinazzi et al., 2020). Based on a dynamic transmission model, public health interventions were found to be effective in reducing the risk of 2019-nCoV spread (Su et al., 2020). Based on the regression method, a study found that, without the lockdown in Wuhan, the infected cases in other Hubei cities, excluding Wuhan, would have increased by 52.6% on February 29 (Fang et al., 2020). Studies have also reported that airborne transmission might play a role in the spread of this novel coronavirus (Liu et al., 2020). The RNA of 2019-nCoV could be detected in an air sample obtained from a patients' toilet at the Fangcang Hospital in Wuhan. However, in the analyzed patient rooms, the concentration of airborne 2019-nCoV was quite low or undetectable. Meteorological factors have also been reported to affect the transmission of this disease in the environment. Based on the reported cases in China from January 20 to February 29, 2020, 2019-nCoV infection was found to be negatively correlated with the environmental temperature (Shi et al., 2020).

However, due to the limitation of the sample sizes and interference from other confounding factors, the impacts of environmental factors (e.g., aerosol concentration and temperature) on the 2019-nCoV transmission have not yet been confirmed.

To date, only close-quarters interaction with infected person has been confirmed as a major factor for 2019-nCoV transmission (Pung et al., 2020). Thus, population dynamics should be an important factor in constructing a COVID-19 pandemic forecast model. Baidu, established in 2000, is a Chinese technology company specializing in its search engine, Internet services, and artificial intelligence (https://www.baidu.com/). Baidu Migration, a big data visualization project developed by Baidu Inc., is utilized to display the population migration of Chinese cities. Baidu Migration is based on the LBS (Location-Based Service) open platform and Baidu Tianyan to calculate and analyze the track and characteristics of population migration in China through the inbound and outbound migration of inter-city and the intensity of intracity traffic. Baidu Migration provides historical and near-real-time indexes to quantify the cross-city population migration and within-city population movement in the major cities of China. In this work, a prediction model for the next-day 2019-nCoV-infected cases was built based on the Baidu migration data and ensemble back-propagation neural network (BPNN) machine learning technique. The city lockdown policy implemented in China is already known to be a highly important factor in controlling the spread of 2019-nCoV in China (Zhang et al., 2020). Some previous researches have perceived the fruitfulness of travel restriction in controlling the spread of COVID-19 (Fanelli and Piazza, 2020, Ray et al., 2020, Sardar et al., 2020). As distinguished from previous studies, this paper combines the BPNN model with Baidu Migration LBS data and considers the impact of intracity and inter-city travel on the spread of the disease, so it has specific promotion and generalization ability. This work further investigated the effectiveness of this policy by enlarging the Baidu migration index and input it into the trained BPNN model.

87　**Methodology**

88　**Baidu Migration Index**

89　The Baidu migration index can be downloaded at https://qianxi.baidu.com/. This index is based on the

90　Baidu Maps LBS open platform, which tracks and collects the location information of the users via mobile

91　phones. The data include the migration index in a single day from cities other than the specific city (based

92　on the cross-city travel population number), the percentage of migrant population coming from each of the

93　other cities, and the internal travel flow index within the specific city (based on the population travel within

94　the city and the number of local residents). According to previous studies, the mean incubation period of

95　2019-nCoV is approximately 5-6 days (Backer et al., 2020, Linton et al., 2020). Therefore, in this work,

96　the Baidu migration data of the previous 4-7 days' averages have been used for the prediction of the

97　confirmed cases in the current day. A previous epidemic estimation study also applied this index to analyze

98　the COVID-19 outbreak in China (Li et al., 2020). Before inputting the data into the BPNN model, the

99　movement data (migration from other cities, internal movement, and the percentage of migrant population

100　from other cities) together with the averages of the reported confirmed cases in the previous 4-7 days were

101　processed using Equations (1) and (2) for the "external_risk" and "internal_risk" calculations:

102　　$external\_risk(t, C) = \frac{1}{4} \sum_{i=t-7}^{i=t-4} QR(t_i, C) \cdot \sum_{a \in all-other-cities} XZ(t_i, a) \cdot QX(t_i, a, C)$　　(1)

103　　$internal\_risk(t, C) = \frac{1}{4} \sum_{i=t-7}^{i=t-4} LD(t_i, C) \cdot XZ(t_i, C)$　　　　　　　　　(2)

104　where QR (t, C) is the migration index (dependent on the total migrant population) from other cities to the

105　targeted city C on day t, XZ (t, a) represents the number of new confirmed cases in the city a on day t, QX

106　(t, a, C) is the percentage of the migrant population in city C that came from city a on day t, LD (t, C) is the

107　internal travel flow index for city C on day t, and XZ (t, C) is the number of new reported confirmed cases

108　in the targeted city C on day t. Besides Wuhan, many Chinese cities began to report COVID-19 cases after

109　January 23, 2020 (Leung et al., 2020). The growth rate of the disease in most of the cities approached zero

110　after February 24. In early March, imported infection cases from foreign countries were observed in some

111　Chinese cities. Therefore, the data used for analysis in this work ranged from January 23 to March 5, 2020.

112　The data from the previous 4-7 days were needed as input for the prediction. Hence, the prediction period

113　was from January 30 to March 5, 2020, giving a total of 36 days. The top 100 cities with the highest number

114　of accumulated confirmed cases (except Wuhan) till early March were obtained. The names of these 100

115　cities, averages of the migration index and internal travel flow index, and the numbers of the confirmed

116　case by March 5th 2020 are shown in Table S1. The origin of the novel coronavirus has not yet been

117　determined. Snake, bat, and pangolin are all possible hosts of this new virus (Ji et al., 2020, Lam et al.,

118   2020, Zhou et al., 2020). Thus, besides interactions with patients, there might be some other causative

119   factors for the spread of this infection in Wuhan, the city of origin, such as contact with possible animal

120   hosts. Therefore, the analysis for Wuhan was not included in this work. But if people migrated from Wuhan

121   to other cities, the 'external_risk' contributed by Wuhan was considered.

122   **BPNN machine learning technique**

123   **Backward Propagation Neural Network (BPNN)** BPNN is a machine learning model which can regress

124   non-linear relations between input features and output targets. The framework of BPNN is a composition

125   of several layer transformations. Each layer transformation is the composition of a nonlinear activation

126   function and a linear transformation. The known linear activation is user-defined and is usually chosen to

127   be hyperbolic tangent function or ReLU function (return itself if input > 0 and 0 if input < 0). The linear

128   transformation can be written as the matrix multiplication, where the entries of the matrices can be adjusted

129   to minimize the prediction loss. The algorithm that is used to learn the learnable parameters is gradient

130   descent. Each parameter is adjusted by its gradient with respect to the loss, multiplied by a user-defined

131   constant (called learning rate). The number of learnable parameters is determined by the number of layers,

132   and the neuron numbers of each layer. The greater number of learnable parameters a network equips, the

133   more complicated function the network can model. The detail of the BPNN is shown in the following

134   formulae.

135   Step 1: Compute the network output.

136
$$\hat{Y} = W_{n_{out} \times n_n} \times [layer_n \circ layer_{n-1} \circ \cdots \circ layer_1 \circ layer_0(X)] + b_{n_{out} \times 1}$$

137   where $layer_k(Z) = g(W_{n_k \times n_{k-1}} \times Z + b_{n_k \times 1})$, g is a non-linear activation function.

138   Step 2: Compute the prediction loss.

139
$$loss = \sum |\hat{Y} - Y|^2$$

140   Step 3: For every weight and bias, update them according to the loss. The α is the user-defined constant,

141   called learning rate.

142
$$W = W - \alpha \cdot \frac{\partial loss}{\partial W}$$

143
$$b = b - \alpha \cdot \frac{\partial loss}{\partial b}$$

144

145 **BPNN Ensemble** Due to the random initialization of the learnable parameters, and the random split of
146 training/test sets, the single BPNN training may harbor relatively large uncertainty. Ensemble method,
147 which refers to the aggregation of several BPNNs, can effectively improve the stability of the BPNN
148 predictions. The usual ways of conducting BPNN ensemble are computing the simple or weighted average
149 of different BPNNs, splitting the training/test sets in a more systematic way (in our case, 10-fold cross
150 validation, which will be introduced later). Since we want a more accurate estimation of the confirmed
151 cases trend, we conducted the BPNN ensemble to reduce the error due to the randomness of BPNNs.

152 An ensemble BPNN was used to model the relationship between the external_risk/internal_risk on the
153 fourth to the seventh preceding days and the number of new confirmed cases on the current day. The number
154 of days since January 23, 2020 was also input into the BPNN for training. Before being input into the BPNN,
155 external_risk, internal_risk and the number of days since January 23, 2020 were normalized to values
156 between 0 and 1. The structure of the ensemble BPNN and processing method is displayed in Fig. 1. Briefly,
157 each BPNN component consists of two layers with five and three neurons, respectively. The data from 90
158 cities were extracted randomly and input into five BPNNs for the training step (Step-1). The data from the
159 other 10 cities were input into the built-up model for the verification step (Step-2). The 10-fold cross-
160 validation results were then obtained after conducting this process 10 times (Step-3). The 10-fold cross-
161 validation ensures each city to have a chance to be tested. This can eliminate the errors caused by the
162 random split of training/test sets. As the weights in the BPNNs were assigned randomly, the results may
163 not be consistent in different rounds of training. Hence, the above three steps were processed 100 times to
164 estimate the uncertainty of the results caused by the random weight initialization in the BPNN.

165 **Results**

166 **Prediction performance of the BPNN**

167 A comparison of the estimated and observed numbers of COVID-19 cases (10-fold cross-validation) by
168 March 5, 2020 for 100 analyzed Chinese cities is shown in Fig. 2 (a). Please note that the verifications
169 presented in Figure 2 are based on the hold-out testsets in the 10-fold cross validation. From the total case
170 number perspective, the BPNN performed well, generally yielding reasonable estimations before March 5
171 for most cities (R = 0.97). However, the total number of confirmed cases in some cities was underestimated
172 by the BPNN. For example, in Ezhou, 1,394 confirmed cases were observed by March 5, but the BPNN
173 underestimated this number by 51%, with a prediction of 688 cases. In Xiaogan, the numbers of observed
174 and estimated confirmed cases were 3,518 and 2,478 respectively, an underprediction of 29.6%. In Suizhou,
175 the BPNN underestimated the number of confirmed cases by 232, a discrepancy of 17.8%. The migration
176 index and within-city population movement index can be taken to represent the overall degree of mobility

177   in a city. However, they do not contain information on gatherings of large groups. Even if only one infected

178   person had attended such a gathering, this might have led to an outbreak within a cluster of people. This

179   may be one of the main reasons for the underestimation of the confirmed COVID-19 cases in these three

180   cities in Hubei province.

181   Fig. 2(b) shows the cumulative numbers of estimated and observed confirmed COVID-19 cases in the 100

182   analyzed cities from January 30 to March 5, 2020. The BPNN was trained 100 times. The green line in the

183   figure is the median value of the estimated cumulative case numbers, with the shaded area representing two

184   standard deviations. In general, the BPNN accurately predicted the trend and magnitude of the confirmed

185   cases from January 30 to March 5. However, a relatively large discrepancy appeared from February 10 to

186   February 17. For example, the BPNN underestimated the number of confirmed COVID-19 cases on

187   February 15 by 2.8% (721 cases). The discrepancy between the prediction and observation narrowed after

188   February 25, and the percentage difference dropped below 1% for March 5. This demonstrates that this

189   ensemble BPNN can accurately capture the relationship between internal_risk/external_risk on the fourth

190   to the seventh preceding days and the number of new confirmed cases on the current day. This in turn

191   indicates that population movement is a determinant of the transmission of this new virus. However, this

192   ensemble BPNN model cannot be used to predict the importation of cases into China from other countries,

193   because no inbound flight information was included in the training input.

194   Impacts of population movement

195   Based on information released by the Ministry of Transport of the People's Republic of China, population

196   movement during the 2020 Chinese Spring Festival (January 10 to February 18, 2020) decreased by 50%

197   relative to 2019 (http://www.mot.gov.cn/jiaotongyaowen/202002/t20200220_3334989.html, in Chinese).

198   Hence, the Baidu migration index and within-city travel flow index were multiplied by a factor of two and

199   input into the trained ensemble BPNN to estimate the impacts that greater population movement would

200   have had on COVID-19 transmission. As shown in Fig. 3, after the Baidu migration index and the within-

201   city travel flow index were multiplied by two, the number of estimated COVID-19 cases was much higher

202   than the official confirmed count. The rate of increase of reported COVID-19 infections began to drop on

203   February 12 and approached zero after February 21. However, if the migration index and within-city travel

204   flow index had been double their actual values, the infection rate would not have begun to slow until

205   February 20, and the estimated number of infected cases would have continued to rise even after March 5.

206   On March 5, the number of estimated COVID-19 cases in this counterfactual scenario reached 127,275

207   (two standard deviations: 57,961–237,467), which is 4.6 times larger than the actual reported number of

208   confirmed cases (27,494) for the 100 analyzed cities. These results indicate that the city lockdown policy

209   after January 23 in a number of Chinese cities was highly effective in controlling the wide spread of the

210 novel coronavirus. Although the lockdown policy interrupted most economic activities, it proved
211 worthwhile, as many residents' lives were saved and substantial medical costs were avoided.

212 The impact of government intervention on COVID-19 infection cases in six major Chinese cities, Beijing
213 (39.90°N, 116.41°E; capital city), Shanghai (31.23°N, 121.47°E; eastern China), Guangzhou (23.13°N,
214 113.26°E; southern China), Chongqing (29.43°N, 106.91°E; southwestern China), Chengdu (30.57°N,
215 104.07°E; southwestern China) and Nanjing (32.06°N, 118.80°E; eastern China), is presented in Fig. 4.
216 Without the implementation of the city lockdown policy, the number of confirmed COVID-19 cases in
217 these six cities would have shot up to 2,237 (946–4,554), 1,904 (445–4,490), 825 (306–1,782), 3,708
218 (1,914–5,592), 1,768 (304-4,442) and 418 (174-690), respectively, by March 5, 2020. Considering the
219 actual numbers of confirmed cases, the government interventions for these cities were highly effective, and
220 the rate of new infections approached zero after February 12. However, it was estimated that without these
221 interventions, the number of COVID-19 cases would have continued to increase after March 5. This
222 demonstrates that reducing personal social contact (e.g., staying at home and maintaining social distance)
223 was one of the main factors responsible for the successful reduction of COVID-19 transmission in China
224 after mid-February. It should be noted that a relatively large uncertainty exists in the values estimated for
225 the six cities here. Besides migration intensity, coronavirus transmission is also controlled by group
226 gatherings and disinfection measures in the community, but neither of these factors was input into the
227 BPNN due to a lack of data. However, our results do imply that the number of COVID-19-infected cases
228 would have increased drastically without appropriate control of population movement.

229 **Discussion**

230 Based on our results, population movement is a key determinant of the transmission of COVID-19. The
231 ensemble BPNN designed in this work can be combined with the Baidu migration index and within-city
232 travel flow index to predict near-future COVID-19 infection. For the regions that do not have the Baidu
233 migration index, population movement data can be used as the alternative for the infection prediction.
234 However, several other factors also play important roles in suppressing the transmission of the disease, such
235 as self-quarantine, personal testing for the SARS-CoV-2 virus, face mask wearing and disinfection
236 measures in the community. Fig. 5 shows the Baidu search indices for "face mask" and "disinfectant" in
237 the six major cities listed above. Both indices were relatively high for January 23 to February 5, which
238 implies that Chinese citizens became strongly motivated to protect themselves from this novel disease by
239 wearing face masks and using disinfectant when the lockdown of Wuhan city was announced on January
240 23. As the reported number of confirmed COVID-19 cases began to drop after February 10, the search
241 indices for "face mask" and "disinfectant" decreased gradually as well. Some recent works have pointed
242 out that face mask use can effectively curtail the community transmission of COVID-19 (Cheng et al., 2020,

243   Eikenberry et al., 2020). Without face mask wearing and effective disinfection, it is likely that the number

244   of confirmed COVID-19 cases would have risen to a higher level under the same degree of population

245   movement. Therefore, to make the ensemble BPNN proposed in this work more universal and better able

246   to predict COVID-19 transmission in other countries, variables that describe face mask wearing and

247   disinfection measures need to be included.

248   It is still unclear whether meteorological factors, such as temperature, relative humidity (RH) and wind

249   speed, influence COVID-19 transmission. Fig. 6 shows the correlation between the number of daily

250   confirmed cases (on day t) and the averages on the fourth to the seventh preceding days for related

251   meteorological factors (from day $t - 7$ to day $t - 4$). The correlation coefficients between the number of

252   new daily confirmed cases and the day $t - 7$ to day $t - 4$ averages of temperature, RH, wind speed and RH

253   × temperature were −0.044, 0.055, −0.036 and −0.043, respectively. No clear correlation was found

254   between the number of confirmed cases and any of these meteorological factors, which is similar to the

255   conclusions made by a previous work (Yao et al., 2020).

256   As the numbers of new daily confirmed cases are mainly governed by population migration and inter-city

257   travel intensity, these numbers were normalized by "internal_risk × external_risk" to remove the impact of

258   population contact on COVID-19 transmission, as shown in Equation (3):

259
$$\text{NCase} = \frac{XZ(t, C)}{\frac{1}{4}\sum_{i=t-7}^{i=t-4}\left(internal_risk(t_i, C) * external_risk(t_i, C)\right)} \quad (3)$$

260   As shown in Table 1, the correlation coefficients between the meteorological factors and the normalized

261   number of new daily confirmed cases ranged from −0.0642 to −0.00727. No clear relationships were found

262   between COVID-19 transmission and the meteorological factors even after removing the risk factor of

263   personal contact. Two other extreme conditions were also considered. First, if the number of confirmed

264   cases in the target city was much higher than that in other cities, only internal_risk should be taken into

265   account. Second, if the number of confirmed cases in the target city was much smaller than that in other

266   cities, only external_risk should be considered. However, when the number of new confirmed cases was

267   also normalized by internal_risk and external_risk, respectively, still no clear relationship was found

268   between COVID-19 transmission and the meteorological factors, as shown in Table 1. This is consistent

269   with the conditions and spread of the disease in other countries. For example, 100-500 new cases of

270   COVID-19 were reported during June to July everyday in Singapore, a tropical region, where the

271   temperature is generally over 30 °C. However, this does not necessarily mean that the transmission of the

272   novel coronavirus is not influenced by temperature and RH. Influenza viruses are known to be affected by

273 temperature and humidity. Meteorological factors may simply have less influence on the transmission of
274 this novel coronavirus than face mask wearing and government intervention.

275 **Conclusions**

276 Since the COVID-19 outbreak in late January, this disease has now spread globally and has substantially
277 disrupted people's lives and economic activities. In this work, an ensemble BPNN machine learning model
278 combining with Baidu migration index was developed to predict near future COVID-19 cases in the major
279 cities over China. The forecast performance is satisfactory and can be used to predict the transmission of
280 COVID-19 in other regions. Our results indicate that restricting people's movement is one of the main
281 factors contributing to the control of this disease. Thus, government interventions are likely the main reason
282 for the successful control over the pandemic by mid-February in most Chinese cities (except Wuhan). In
283 some other countries, such as Italy, although initially the numbers of infected cases are high, the daily
284 confirmed cases began decreasing gradually once the lockdown restrictions were implemented (Sebastiani
285 et al., 2020). Therefore, besides early diagnosis and medical treatment, restricting the population movement
286 and maintaining social distance appear to be the best methods to suppress the spread of this highly
287 contagious virus when no effective vaccine is yet available. Some previous studies have pointed out that
288 meteorological factors can also influence 2019-nCoV transmission. However, based on our analysis, these
289 factors are not dominant in controlling the transmission rate of this disease. That being said, although
290 meteorological factors are not dominant in influencing the COVID-19 pandemic, we cannot rule out the
291 possibility that these factors still influence the spread of this coronavirus to a limited extent. More studies
292 are warranted to further understand the meteorological impacts on this new disease.

296
297
298
299
300
301
302
303

304 **References:**

305 Backer JA, Klinkenberg D, Wallinga J. Incubation period of 2019 novel coronavirus (2019-nCoV)
306 infections among travellers from Wuhan, China, 20–28 January 2020. Eurosurveillance
307 2020;25(5):2000062.

308 Cheng VC, Wong S-C, Chuang VW, So SY, Chen JH, Sridhar S, et al. The role of community-wide wearing
309 of face mask for control of coronavirus disease 2019 (COVID-19) epidemic due to SARS-CoV-2. Journal
310 of Infection 2020.

311 Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions
312 on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 2020;368(6489):395-400.

313 Eikenberry SE, Mancuso M, Iboi E, Phan T, Eikenberry K, Kuang Y, et al. To mask or not to mask:
314 Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic.
315 Infectious Disease Modelling 2020.

316 Fanelli D, Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos,
317 Solitons & Fractals 2020;134:109761.

318 Fang H, Wang L, Yang Y. Human mobility restrictions and the spread of the novel coronavirus (2019-ncov)
319 in china. National Bureau of Economic Research; 2020.

320 Fernandes N. Economic effects of coronavirus outbreak (COVID-19) on the world economy. Available at
321 SSRN 3557504 2020.

322 Ji W, Wang W, Zhao X, Zai J, Li X. Cross-species transmission of the newly identified coronavirus 2019-
323 nCoV. Journal of medical virology 2020;92(4):433-40.

324 Kishan H. China's first-quarter economic hit from coronavirus looking more severe: Reuters poll; 2020.
325 Available from: https://www.reuters.com/article/us-china-economy-poll/china-first-quarter-economic-hit-
326 from-coronavirus-looking-more-severe-reuters-poll-idUSKBN20T00L. [Accessed 4 Aug 2020].

327 Lam TT-Y, Jia N, Zhang Y-W, Shum MH-H, Jiang J-F, Zhu H-C, et al. Identifying SARS-CoV-2-related
328 coronaviruses in Malayan pangolins. Nature 2020:1-4.

329 Leung K, Wu JT, Liu D, Leung GM. First-wave COVID-19 transmissibility and severity in China outside
330 Hubei after control measures, and second-wave scenario planning: a modelling impact assessment. The
331 Lancet 2020.

11

332 Li C, Chen LJ, Chen X, Zhang M, Pang CP, Chen H. Retrospective analysis of the possibility of predicting

333 the COVID-19 outbreak from Internet searches and social media data, China, 2020. Eurosurveillance

334 2020;25(10):2000199.

335 Linton NM, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov AR, Jung S-m, et al. Incubation period and

336 other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical

337 analysis of publicly available case data. Journal of clinical medicine 2020;9(2):538.

338 Liu Y, Ning Z, Chen Y, Guo M, Liu Y, Gali NK, et al. Aerodynamic characteristics and RNA concentration

339 of SARS-CoV-2 aerosol in Wuhan hospitals during COVID-19 outbreak. BioRxiv 2020.

340 Lu H, Stratton CW, Tang YW. Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery

341 and the miracle. Journal of medical virology 2020;92(4):401-2.

342 Nicola M, Alsafi Z, Sohrabi C, Kerwan A, Al-Jabir A, Iosifidis C, et al. The socio-economic implications

343 of the coronavirus pandemic (COVID-19): A review. International journal of surgery (London, England)

344 2020;78:185.

345 Petropoulos F, Makridakis S. Forecasting the novel coronavirus COVID-19. PloS one

346 2020;15(3):e0231236.

347 Pung R, Chiew CJ, Young BE, Chin S, Chen MI, Clapham HE, et al. Investigation of three clusters of

348 COVID-19 in Singapore: implications for surveillance and response measures. The Lancet 2020.

349 Ray D, Salvatore M, Bhattacharyya R, Wang L, Du J, Mohammed S, et al. Predictions, role of interventions

350 and effects of a historic national lockdown in India's response to the COVID-19 pandemic: data science

351 call to arms. Harvard data science review 2020;2020(Suppl 1).

352 Sardar T, Nadim SS, Rana S, Chattopadhyay J. Assessment of Lockdown Effect in Some States and Overall

353 India: A Predictive Mathematical Study on COVID-19 Outbreak. Chaos, Solitons & Fractals 2020:110078.

354 Sebastiani G, Massa M, Riboli E. Covid-19 epidemic in Italy: evolution, projections and impact of

355 government measures. European journal of epidemiology 2020;35(4):341.

356 Shi P, Dong Y, Yan H, Li X, Zhao C, Liu W, et al. The impact of temperature and absolute humidity on

357 the coronavirus disease 2019 (COVID-19) outbreak-evidence from China. MedRxiv 2020.

358 Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares

359 global emergency: A review of the 2019 novel coronavirus (COVID-19). International Journal of Surgery

360 2020.

361   Su L, Hong N, Zhou X, He J, Ma Y, Jiang H, et al. Evaluation of the secondary transmission pattern and
362   epidemic prediction of COVID-19 in the four metropolitan areas of China. Frontiers in Medicine 2020;7.

363   Tan H, Cheng E. China says its economy shrank by 6.8% in the first quarter as the country battled
364   coronavirus; 2020. Available from: https://www.cnbc.com/2020/04/17/china-economy-beijing-contracted-
365   in-q1-2020-gdp-amid-coronavirus.html. [Accessed 8 Aug 2020].

366   Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. The
367   Lancet 2020;395(10223):470-3.

368   Worldmeter. COVID-19 CORONAVIRUS PANDEMIC; 2020. Available from:
369   https://www.worldometers.info/coronavirus/. [Accessed 4 Aug 2020].

370   Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread
371   of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. The Lancet
372   2020;395(10225):689-97.

373   Wynants L, Van Calster B, Bonten MM, Collins GS, Debray TP, De Vos M, et al. Prediction models for
374   diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. bmj 2020;369.

375   Yang Z, Zeng Z, Wang K, Wong S-S, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the
376   epidemics trend of COVID-19 in China under public health interventions. Journal of Thoracic Disease
377   2020;12(3):165.

378   Yao Y, Pan J, Liu Z, Meng X, Wang W, Kan H, et al. No Association of COVID-19 transmission with
379   temperature or UV radiation in Chinese cities. European Respiratory Journal 2020;55(5).

380   Zhang C, Chen C, Shen W, Tang F, Lei H, Xie Y, et al. Impact of population movement on the spread of
381   2019-nCoV in China. Emerging Microbes & Infections 2020;9(1):988-90.

382   Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a
383   new coronavirus of probable bat origin. nature 2020;579(7798):270-3.

384

385

386

387

388

389   Table 1: Correlations between the daily confirmed cases normalized by the risk factors of the previous 4-7

390   days' average and meteorological factors.

Normalized by	Temperature	RH	Wind Speed	Temperature $\times$ RH
internal_risk $\times$ external_risk	-0.0605	-0.0642	-0.0073	-0.0614
internal_risk	-0.0435	-0.0099	-0.0156	-0.0428
external_risk	-0.0561	-0.0510	-0.0043	-0.0546

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405
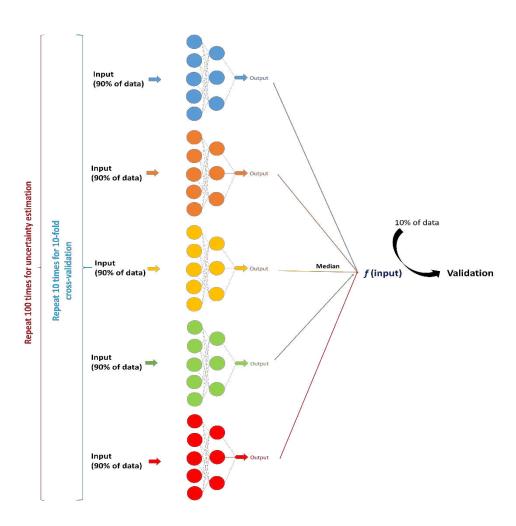
406   Fig. 1: Structure of the ensemble BPNN.

407

408
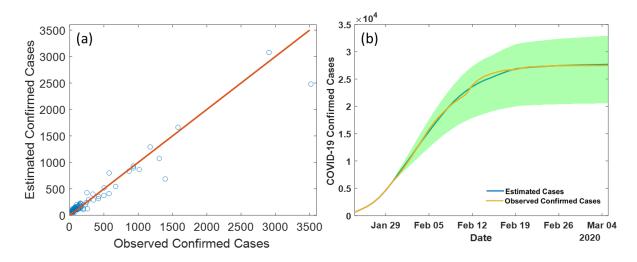
Fig. 2: (a) Comparison between the total observed confirmed cases and estimated confirmed cases on March 5, 2020; (b) comparison between the observed and estimated accumulated cases. The green shaded area is within two standard deviations of running the ensemble BPNN 100 times.

16

428

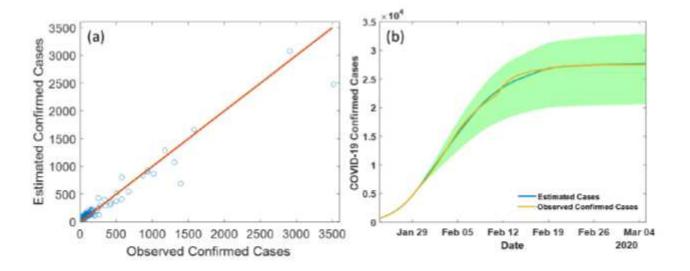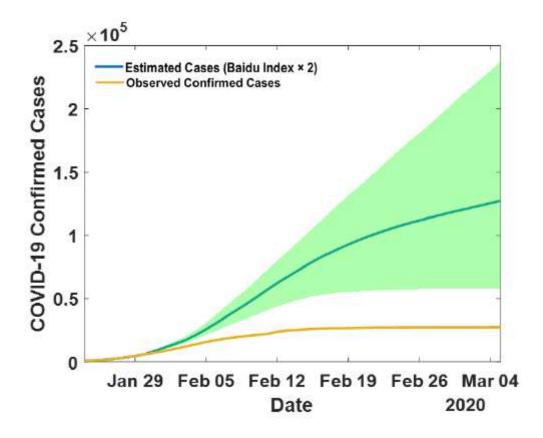Fig. 3: Estimated number of COVID-19 cases after doubling the Baidu indices for the 100 analyzed cities.

429

430

431

432

433

434

435
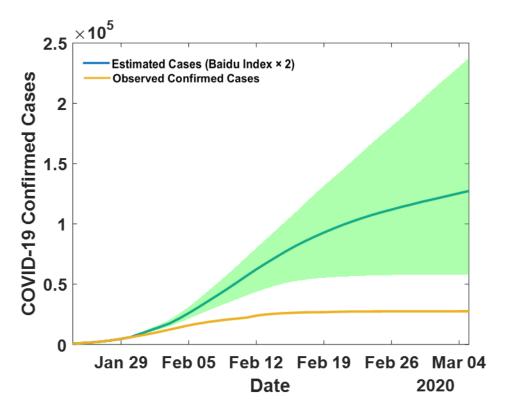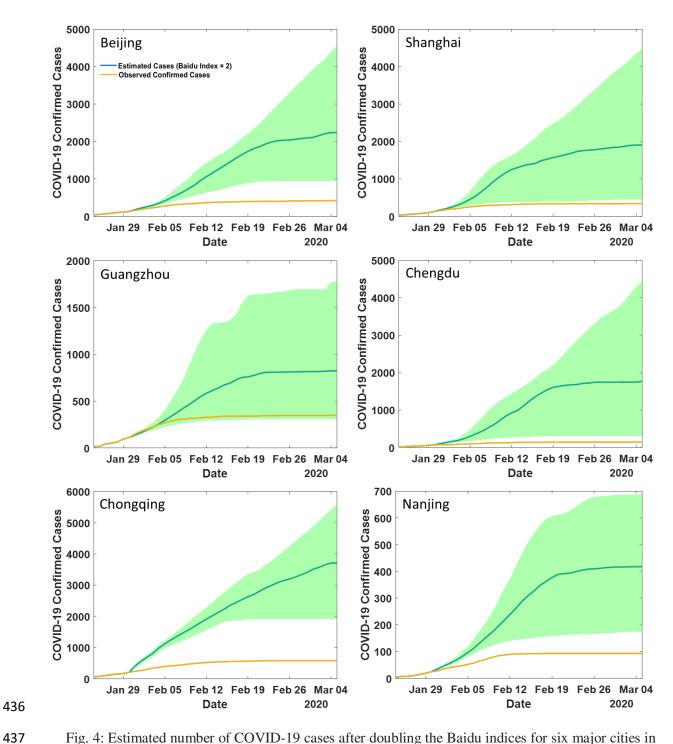
Fig. 4: Estimated number of COVID-19 cases after doubling the Baidu indices for six major cities in China.

Fig. 5: Baidu search indices for "face mask" and "disinfectant" for six major cities in China.

460
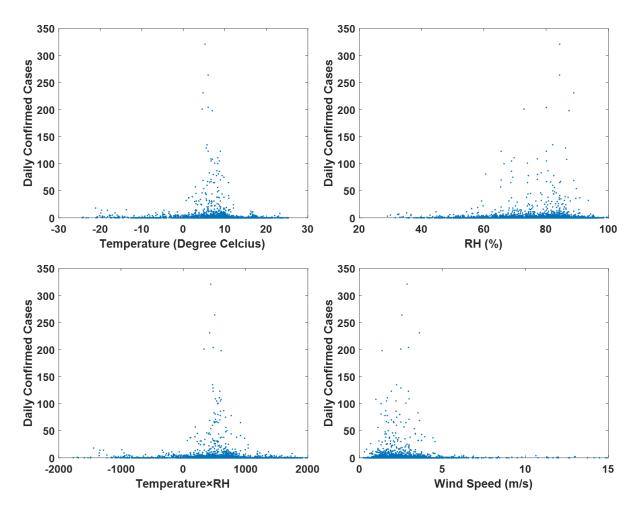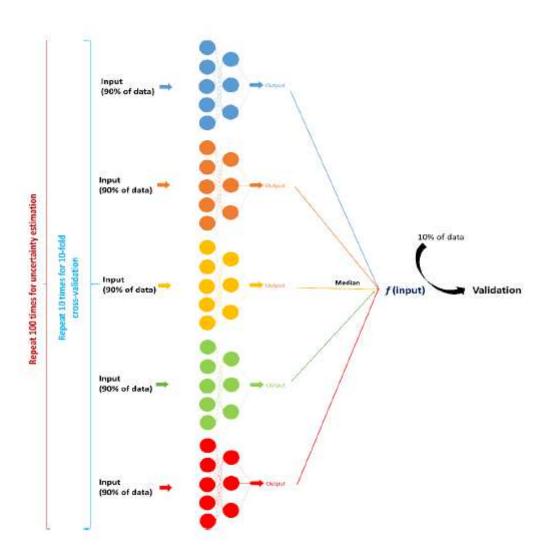
Fig. 6: Relationships between the daily confirmed cases and meteorological factors (previous 4-7 days' averages).

461

462

463

Figures

Figure 1

Structure of the ensemble BPNN.

Figure 2

(a) Comparison between the total observed confirmed cases and estimated confirmed cases on March 5, 2020; (b) comparison between the observed and estimated accumulated cases. The green shaded area is within two standard deviations of running the ensemble BPNN 100 times.
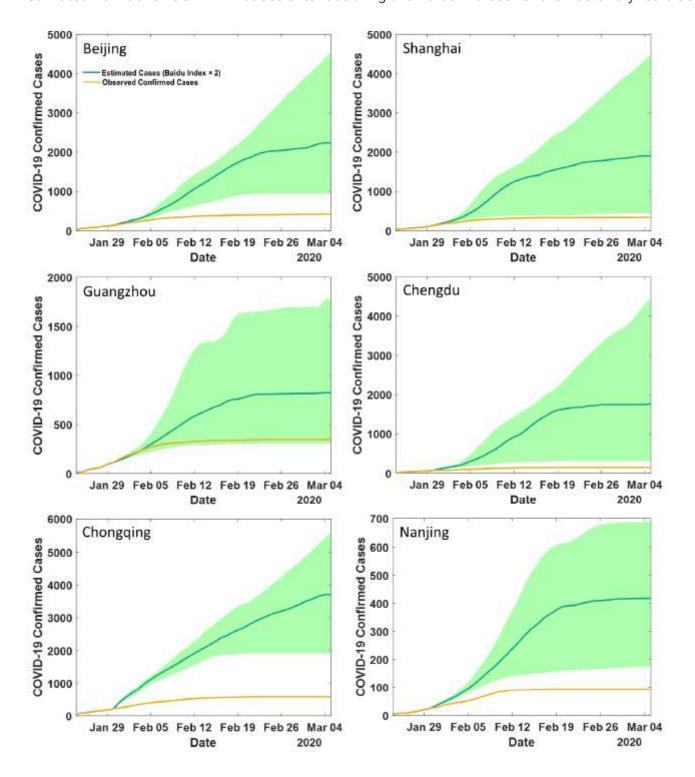

Figure 3

Estimated number of COVID-19 cases after doubling the Baidu indices for the 100 analyzed cities.

Figure 4

Estimated number of COVID-19 cases after doubling the Baidu indices for six major cities in China.
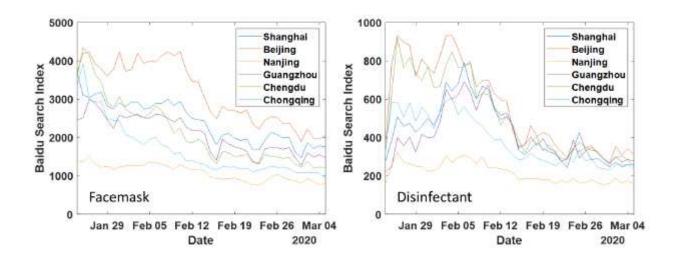
Figure 5

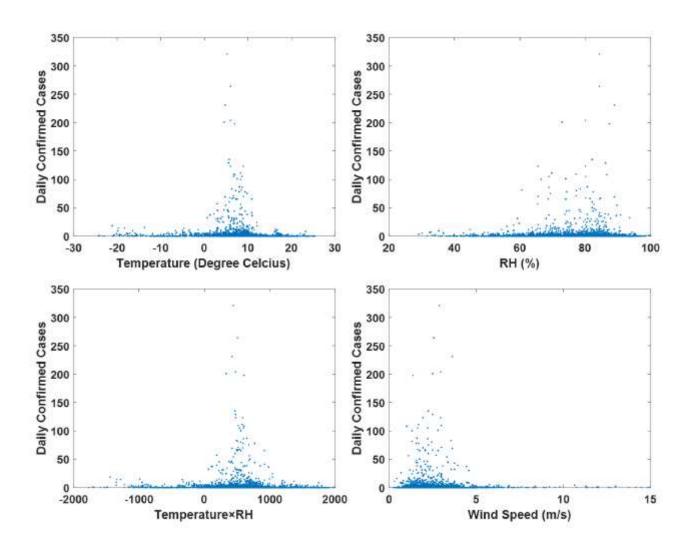Baidu search indices for "face mask" and "disinfectant" for six major cities in China.

Figure 6

Relationships between the daily confirmed cases and meteorological factors (previous 4-7 days' averages).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalMaterial.pdf