

1 **Assessment of the outbreak risk, mapping and infestation**  
2 **behavior of COVID-19: Application of the autoregressive and**  
3 **moving average (ARMA) and polynomial models**

4

5 Hamid Reza Pourghasemi<sup>1\*</sup>, Soheila Pouyan<sup>2</sup>, Zakariya Farajzadeh<sup>3</sup>, Nitheshnirmal  
6 Sadhasivam<sup>4</sup>, Bahram Heidari<sup>5\*</sup>, Sedigheh Babaei<sup>1</sup>, John P. Tiefenbacher<sup>6</sup>

7 <sup>1</sup>Department of Natural Resources and Environmental Engineering, College of Agriculture,  
8 Shiraz University, Shiraz, Iran

9 <sup>2</sup>Department of Arid Land and Desert Management, Faculty of Natural Resources and Desert  
10 Studies, Yazd University, Yazd, Iran

11 <sup>3</sup>Department of Agricultural Economics, College of Agriculture, Shiraz University, Shiraz,  
12 Iran

13 <sup>4</sup>Department of Geography, School of Earth Science, Bharathidasan University,  
14 Tiruchirappalli, Tamil Nadu, India

15 <sup>5</sup>Department of Plant Production and Genetics, School of Agriculture, Shiraz University,  
16 Shiraz, Iran.

17 <sup>6</sup>Department of Geography, Texas State University, San Marcos, United States of America

18 \* Corresponding authors

19 **E-mail:** [hr.pourghasemi@shirazu.ac.ir](mailto:hr.pourghasemi@shirazu.ac.ir) (HP)

20 **E-mail:** [bheidari@shirazu.ac.ir](mailto:bheidari@shirazu.ac.ir) (BH)

21

22 **Short title:** Modelling outbreak risk and infestation behavior of coronavirus

23

24

25

26

27

28

29

30

31 **Abstract**

32 Infectious disease outbreaks pose a significant threat to human health worldwide. The  
33 outbreak of pandemic coronavirus disease 2019 (COVID-2019) has caused a global health  
34 emergency. Identification of regions with high risk for COVID-19 outbreak is a major  
35 priority of the governmental organizations and epidemiologists worldwide. The aims of the  
36 present study were to analyze the risk factors of coronavirus outbreak and identify areas with  
37 a high risk of human infection with virus in Fars Province, Iran. A geographic information  
38 system (GIS)-based machine learning algorithm (MLA), support vector machine (SVM), was  
39 used for the assessment of the outbreak risk of COVID-19 in Fars Province, Iran. The daily  
40 observations of infected cases was tested in the third-degree polynomial and the  
41 autoregressive and moving average (ARMA) models to examine the patterns of virus  
42 infestation in the province and in Iran. The results of disease outbreak in Iran were compared  
43 with the data for Iran and the world. Sixteen effective factors including minimum temperature  
44 of coldest month (MTCM), maximum temperature of warmest month (MTWM), precipitation  
45 in wettest month (PWM), precipitation of driest month (PDM), distance from roads, distance  
46 from mosques, distance from hospitals, distance from fuel stations, human footprint, density  
47 of cities, distance from bus stations, distance from banks, distance from bakeries, distance  
48 from attraction sites, distance from automated teller machines (ATMs), and density of  
49 villages – were selected for spatial modelling. The predictive ability of an SVM model was  
50 assessed using the receiver operator characteristic – area under the curve (ROC-AUC)  
51 validation technique. The validation outcome reveals that SVM achieved an AUC value of  
52 0.786 (March 20), 0.799 (March 29), and 86.6 (April 10) a good prediction of change  
53 detection. The growth rate (GR) average for active cases in Fars for a period of 41 days was  
54 1.26, whilst it was 1.13 in country and the world. The results of the third-degree polynomial

55 and ARMA models revealed an increasing trend for GR with an evidence of turning,  
56 demonstrating extensive quarantines has been effective. The general trends of virus  
57 infestation in Iran and Fars Province were similar, although an explosive growth of the  
58 infected cases is expected in the country. The results of this study might assist better  
59 programming COVID-19 disease prevention and control and gaining sorts of predictive  
60 capability would have wide-ranging benefits.

61 **Keywords:** COVID-19, Outbreak risk mapping, Support vector machine, Machine learning  
62 algorithm, ARMA model, Growth rate of deaths.

### 63 **Introduction**

64 In December 2019 several pneumonia infected cases were reported in Wuhan, China [1-2]. In  
65 January 2020, a novel coronavirus (2019-nCoV) that was later formally named COVID-19  
66 was approved in Wuhan [3]. It was announced that the disease is a severe acute respiratory  
67 syndrome coronavirus 2 (SARS-CoV-2). The virus elevated concerns within China as well as  
68 the global community as it was believed to be transmitted from human to human [4]. Initially,  
69 China witnessed the largest outbreak in Hubei and other nearby provinces. The spread in  
70 China was controlled soon thereafter through stringent preventive measures, but other parts  
71 of the world (Europe, the Middle East, and the United States) were increasingly affected by  
72 the outbreak through transmission by infected travellers from China. A similar outbreak soon  
73 followed in other Asian countries [5]. Its global spread to more than 150 countries led to the  
74 declaration in mid-March 2020 that COVID-19 was a pandemic [6]. By April 10, 2020, there  
75 were nearly 1.70 million cases worldwide with 102684 deaths attributed to COVID-19 [7].  
76 Currently, the United States has the largest number of confirmed cases, while Italy has  
77 reported the highest number of casualties [7-8]. Iran with 68,192 recorded cases and 4232  
78 deaths is the most affected country in the Middle East (as of April 10, 2020) and infected

79 cases are expected to surge in the coming days [7, 9]. The outbreak of COVID-19 has  
80 disrupted and depressed the world economy, whereas Iran is among the most severely  
81 affected by massive economic losses, largely compounded by politically motivated sanctions  
82 imposed by other governments [10]. The problem has been exacerbated as no specific  
83 medicine is yet realized for COVID-19 disease treatment, though there are a few pre-existing  
84 drugs that are being tested, so regions are presently concentrating their efforts on maintaining  
85 the infection rate in a level that assists to reduce virus spread [11]. This has led to most states  
86 imposing lockdowns, encouraging social distancing, and restricting the sizes of gatherings to  
87 limit transmission [12]. There is a pressing necessity for scientific communities to aid  
88 governments in their efforts to control and prevent transmission of the virus [13].

89 During previous virus outbreaks stemming from Zika, influenza, West Nile, Dengue,  
90 Chikungunya, Ebola, Marburg, and Nipah, geographic information systems (GISs) have  
91 played significant roles in providing significant insight via risk mapping, spatial forecasting,  
92 monitoring spatial distributions of supplies, and providing spatial logistics for management  
93 [13]. In this current situation, risk mapping is critical and may be used to aid governments'  
94 need for tracking and management of the disease as it spread in places with the highest risk.  
95 Sánchez-Vizcaíno et al. [14] used a multi-criteria decision making (MCDM) model to map  
96 the risk of Rift Valley fever in Spain. Traditional statistical techniques had been also used to  
97 detect the risk of outbreak [14]. Reeves et al. [15] employed an ecological niche modelling  
98 (ENM) technique for mapping the transmission risk of MERS-CoV; the Middle Eastern name  
99 for the coronavirus known as SARS-CoV-2. Similar techniques have been in the  
100 Nyakarahuka et al. [16] study to map Ebola and Marburg viruses risks in Uganda. They  
101 assessed the importance of environmental covariates using the maximum entropy model.

102 More recently, the use of machine learning algorithms (MLAs) for mapping the risk of  
103 transmission of viruses has been increasing which is due to the demonstrated superior (and

104 more accurate) predictive abilities of the MLA models over traditional methods [17]. Jiang et  
105 al. [18] employed three MLAs – backward propagation neural network (BPNN), gradient  
106 boosting machine (GBM), and random forest (RF) – to map the risk of an outbreak of Zika  
107 virus. Tien Bui et al. (2019) compared different MLAs – artificial neural network (ANN) and  
108 support vector machine (SVM) with ensemble models including adaboost, bagging, and  
109 random subspace – for modelling malaria transmission risk. Similarly, GBM, RF, and general  
110 additive modelling (GAM) were used by Carvajal et al. [19] to model the patterns of dengue  
111 transmission in the Philippines. Mohammadinia et al. [20] employed geographically weighted  
112 regression (GWR), generalized linear model (GLM), SVM, and ANN to develop a forecast  
113 map of leptospirosis; GWR and SVM produced highly accurate predictions. The literature  
114 shows that very few studies have tried to use GIS for analysis of COVID-19 outbreak in  
115 human communities. Kamel Boulos and Geraghty [21] described the use of online and  
116 mobile GIS for mapping and tracking COVID-19 whilst Zhou et al. [13] revealed the  
117 challenges of using GIS for SARS-CoV-2 big data sources . To our knowledge, there has  
118 been no study with focus on mapping the outbreak risk of the COVID-19 pandemic. The aims  
119 of the present study were to analyze the risk factors of coronavirus outbreak and test the SVM  
120 model for mapping areas with a high risk of human infection with virus in Fars Province,  
121 Iran. The outcome of the present study lays a foundation for better programming and  
122 understanding the factors that accelerate virus spread for use in disease control plans in  
123 human communities.

124

## 125 **Materials and methods**

### 126 **Study area**

127 The study area is in the southern part of Iran with an area of 122608 square kilometres  
128 located between 27°2' and 31°42' N and between 50°42' and 55°36' E. Fars is the fourth  
129 largest province in Iran (7.7 % of total area) with a population density of 4851274 (based on  
130 in 2016 report). Fars Province is divided into 36 counties, 93 districts, and 112 cities (Fig 1).

131 **Fig 1.** The counties of Fars Province, Iran, and the number of COVID-19 infected case  
132 identified from March 29, 2020.

133

### 134 **Methodology**

135 The multi-phased workflow implemented in this investigation (Fig. 2) is described  
136 comprehensively below.

137 **Fig 2.** The methodological framework followed in this study.

138

### 139 **Preparation of location of COVID-19 active cases**

140 A dataset of active cases of COVID-19 in Fars was prepared to analyse the relationships  
141 between the locations of active cases and the effective factors that may be useful for  
142 predicting outbreak risk. The data utilized in this research was collected on April 10, 2020  
143 from Iranian's Ministry of Health and Medical Education (IMHME).

### 144 **Preparation of effective factors**

145 Choosing the appropriate effective factors to predict the risk of pandemic spread is vital as its  
146 quality affects the validity of the results [17]. Since, there have been no previous studies of  
147 risk for COVID-19 distribution, the selection of effective factors is a quiet challenging task.  
148 Ongoing research on the pandemic has revealed that local and community-wide transmission

149 of the virus largely happens in public places where the most people are likely to come into  
150 contact with largest number of potential carriers of the infection [22]. Wang et al. [23]  
151 indicated that meteorological conditions, such as rapidly warming temperatures in 439 cities  
152 around the world resulted in a decline of COVID-19 cases. Accordingly, in this research, we  
153 selected sixteen most relevant effective factors for the outbreak risk mapping of COVID-19  
154 in Fars Province of Iran, which includes minimum temperature of coldest month (MTCM),  
155 maximum temperature of warmest month (MTWM), precipitation in wettest month (PWM),  
156 precipitation of driest month (PDM), distance from roads, distance from mosques, distance  
157 from hospitals, distance from fuel stations, human footprint, density of cities, distance from  
158 bus stations, distance from banks, distance from bakeries, distance from attraction sites,  
159 distance from automated teller machines (ATMs) and density of villages. All the effective  
160 factors employed in this research are generated using the ArcGIS 10.7.

161 A few studies have established that variation in temperature would impact the transmission of  
162 COVID-19 [23]. It has been also reported that alteration in temperature would have impacted  
163 the SARS outbreak, which was caused by the identical type of coronavirus as SARS-CoV-2  
164 [24]. Recently, Ma et al. [2] disclosed that surge in temperature and humidity conditions have  
165 resulted in the decline of death caused by SARS-CoV-2. Thus, climatic factors such as  
166 temperature and precipitation can have an impact in the outbreak of SARS-CoV-2. The  
167 temperature and precipitation data namely MTWM, MTCM, PDM and PCM of Fars Province  
168 is acquired from world climatic data (<https://www.worldclim.org/>). In this study, the MTWM  
169 of the Fars Province ranges from 27.7<sup>0</sup>C to 41.8<sup>0</sup>C (Fig 3a) whereas MTCM ranges between  
170 -15.3<sup>0</sup>C and 10.4<sup>0</sup>C (Fig 3b). The PWM of the study area varies between 28 mm and 86 mm  
171 (Fig 3c) and also the PDM is presented in Fig 3d.

172 **Fig 3.** Preparation of effective factors of COVID-19 outbreak

173

174 The proximity to various public places including roads, mosques, hospitals, fuel stations, bus  
175 stations, banks, bakeries, attraction sites, and ATMs where people come in close contact to  
176 each other can also be considered as significant factors that influence the distribution of  
177 COVID-19. The distance from roads ranges from 0 to 45 in the study area (Fig 3e) whereas  
178 the distance from mosques varies between 0 and 0.71 (Fig 3f) and the distance from fuel  
179 stations spans 0 to 0.67 (Fig 3g). The distance from bus stations, banks, bakeries, attraction  
180 sites, and ATMs of Fars Province have the minimum value of 0 and maximum value of 1.31,  
181 0.68, 0.97, 0.79, and 0.78 respectively (Fig 3h – 3l). Since, humans are the potential carriers  
182 of the COVID-19, the use of human footprint (HFP) can aid in understanding the terrestrial  
183 biomes on which humans have more influence and access [25]. In this study, HFP of the  
184 study area is acquired from the Global Human Footprint Dataset. The HFP of Fars Province  
185 ranges from 6 to 78 (Fig 3m) where the minimum value represents the places having least  
186 access by humans and the maximum value refers to those regions having highest human  
187 influence and access. The density of population is also considered to be an important factor  
188 for the spread of the disease [26-27]. Gilbert et al., [28] revealed that the number of COVID-  
189 19 cases were proportional to the population density in Africa. Accordingly, in this research,  
190 density of cities and villages were assessed and the outcome displays that density of cities in  
191 Fars Province ranges between 0 and 0.60 (Fig 3n) while the density of villages varies from 0  
192 to 0.58 (Fig 3o). The distance from hospitals ranged from 0 to 1.11 (Fig 3p).

### 193 **Evaluation of variable importance using ridge regression**

194 The association among the location of COVID-19 active cases and effective factors were  
195 evaluated using ridge regression in order to assess the significance of individual effective  
196 factor in predicting the outbreak risk [17]. To our knowledge, no previous study in epidemic  
197 outbreak risk mapping have utilized ridge regression in determining the significance of  
198 effective factors. However, the ridge regression algorithm has been utilized for modelling



199 purposes in various fields [29]. It was first given by Hoerl and Kennard [30] which exploits  
200  $L_2$  norm of regularization for lessening the model complication and controlling overfitting.  
201 Ridge regression was also developed to avoid the excessive instability and collinearity  
202 problem caused by least square estimator [31]. The ‘caret’ package ([https://cran.r-](https://cran.r-project.org/web/packages/caret/caret.pdf)  
203 [project.org/web/packages/caret/caret.pdf](https://cran.r-project.org/web/packages/caret/caret.pdf)) of R 3.5.3 was utilized for assessing the variable  
204 importance using ridge regression.

## 205 **Machine learning algorithm (MLA)**

### 206 **Support vector machine**

207 SVM is an extensively exercised MLA in diverse fields of research that functions on the  
208 principle of statistical learning concept and structural risk minimization given by Vapnik  
209 [32], which is utilized for classification as well as regression intricacies [33-34]. SVM has a  
210 high efficacy in classifying both linearly separable and inseparable data classes [35]. It  
211 utilizes an optimal hyperplane to distinguish linearly divisible data whereas kernel functions  
212 are employed for transforming inseparable data into a higher dimensional space so that it can  
213 be easy categorized [36]. Assume a calibration dataset to be  $(s_m, t_m)$ , where  $m$  is 1, 2, 3, ...,  $x$ ;  
214  $s_m$  refers to the sixteen independent factors;  $t_m$  denotes 0 and 1 that resembles risk and non-  
215 risk classes and  $x$  represents the total amount of calibration data. This algorithm tries to  
216 obtain an optimal hyperplane for classifying the aforementioned classes by utilizing the  
217 distance between them, which can be formulated as follows [37]:

$$218 \quad \frac{1}{2} \|p\|^2 \quad (1)$$

$$219 \quad t_m((p \times s_m) + a) \geq 1 \quad (2)$$

220 where,  $\|p\|$  denotes the rule of normal hyperplane;  $a$  refers to a constant. When Lagrangian  
221 multiplier ( $\lambda_m$ ) and cost function is introduced, the expression can be given as follows [38]:

$$222 \quad l = \frac{1}{2} \|p\|^2 - \sum_{n=1}^x \lambda_n (t_n ((p \times s_m) + a) - 1) \quad (3)$$

223 In case of inseparable dataset, a slack covariate  $\delta_m$  is added into the previous Eq. (2) that is  
224 provided as follows [32]:

$$225 \quad t_m ((p \times s_m) + a) \geq 1 - \delta_m \quad (4)$$

226 Accordingly, the Eq. (3) can be described as follows [32]:

$$227 \quad L = \frac{1}{2} \|p\|^2 - \frac{1}{ux} \sum_{n=1}^x \delta_m \quad (5)$$

228 Moreover, SVM contains four kernel functions (linear, polynomial, radial basis function:  
229 RBF and sigmoid) for making an optimal margin in case of inseparable dataset [32].  
230 Mohammadinia et al. [20] revealed that RBF kernel type produces high prediction accuracy  
231 than other kernel types for epidemic outbreak risk mapping. Thus, in this study, RBF is used  
232 for creating decision boundaries and the kernel function is expressed as follows [32]:

$$233 \quad K(z_a, z_b) = (-v \square z_a - z_b \square), v > 0 \quad (6)$$

234 where,  $K(z_a, z_b)$  refers to kernel function and  $v$  represents its parameter.

### 235 **Analysis of growth rate for active and death cases of COVID-19**

236 In this study, the growth rate (GR) of active and death cases around the world, Iran, and Fars  
237 Province were evaluated using the data acquired from WHO and IMHME between February  
238 26, 2020 and April 10, 2020 for active cases and from March 3, 2020 to April 10, 2020 for  
239 death cases.

### 240 **Validation of outbreak risk map**

241 The cross-checking of calibrated model using untouched testing data is vital for determining  
242 the scientific robustness of the prediction [33]. In this research, we utilized ROC- AUC

243 curve values for the validation of COVID-19 outbreak risk map generated using SVM model.  
244 It is a widely utilized validation technique for analysing the predictive ability of a model [35].  
245 A model is determined to be perfect, very good, good, moderate and poor if the AUC values  
246 were 1.0-0.9, 0.9-0.8, 0.8-0.7, 0.7-0.6 and 0.6-0.5 respectively [39].

247

#### 248 **Models for infection cases trend**

249 The behavior of the variable infection cases was captured by a third-degree polynomial or  
250 cubic specification as follows:

$$251 \quad \text{Infection}(t) = \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3 \quad (7)$$

252 Where  $\text{Infection}(t)$  represents the total infected cases in day  $t$  and  $t$  denotes the days  
253 starting from 19th of February for Iran and one week later for Fars province. Also, other  
254 specifications including quadratic as well as fourth-degree polynomial specifications were  
255 examined and based on the predictions, the cubic form was selected against other  
256 specifications. In the literature, this form of the specification has been applied by Aik et al.  
257 [40] to examine the Salmonellosis incidence in Singapore. We also used an ARMA model to  
258 compare the process generating the variable for Iran and Fars province. This model includes  
259 two processes: Autoregressive (AR) and Moving Average (MA) process. An ARMA model  
260 of order  $(p,q)$  can be written as [41]:

$$261 \quad x(t) = \beta_0 + \sum_{i=1}^p \beta_i x_{t-i} + \sum_{j=1}^q \beta_j \varepsilon_{t-j} \quad (8)$$

262 Where  $x$  is the dependent variable and  $\varepsilon$  is the white noise stochastic error term. In the  
263 applied model,  $x$  shows the total infected cases and  $t$  is the days starting from the first day of  
264 happening infection cases. Benvenuto et al. [42] also applied an ARIMA model to predict the  
265 epidemiological trend of COVID-2019.

266  
267  
268

## **Results**

### **269 Outcome of the variable importance analysis**

270 The analysis of variable importance using ridge regression revealed that distance from bus  
271 stations, distance from hospitals, and distance from bakeries have the highest significance  
272 whereas distance from ATMs, distance from attraction sites, distance from fuel stations,  
273 distance from mosques, distance from road, MTCM, density of cities and density of villages  
274 exhibit moderate importance. The effective factors such as distance from banks, MTWM,  
275 HFP, PWM and PDM were the least influential factors (Fig 4).

276 **Fig 4.** Variable importance of each effective factors (bus: distance from bus stations; hospital:  
277 distance from hospitals; bakery: distance from bakeries; atm: distance from ATMs; attraction:  
278 distance from attraction sites; fuel: distance from fuel stations; mosque: distance from  
279 mosques; road: distance from road; bio6: MTCM; city: density of cities; village: density of  
280 villages; bank: distance from banks; bio13: MTWM; footprint: HFP, bio14: PWM; bio5:  
281 PDM.

282

### **283 COVID-19 outbreak risk map using SVM**

284 The COVID-19 outbreak risk map generated using SVM displays that risk of SARS-CoV-2  
285 ranges from -0.25 to 1.22 (March 29) and -0.35 to 1.21 (April 10) where -0.25 and -0.35  
286 represents the lower risk of SARS-CoV-2 outbreak and 1.22 and 1.21 indicates the regions of  
287 Fars Province which is likely to experience a higher risk of COVID-19 outbreak (Fig 5, a-b).  
288 It can be observed from Fig 5b (April 10) that Shiraz County and its surrounding counties  
289 including Firouzabad, Jahrom, Sarvestan, Arsanjan, Marvdasht, Sepidan, Abadeh,  
290 Khorrambid, Rostam, Larestan and Kazeron of Fars Province has the highest risk of being the

291 epicentre of SARS-CoV-2 outbreak. Apart from which counties like Eghlid, and Fasa also lie  
292 in the high risk zone.

293 **Fig 5.** The COVID-19 outbreak risk map a) on March 29, 2020 and b) on April 10, 2020

294

### 295 **Outcome of growth rate analysis**

296 The results of GR of active cases in world, Iran, and Fars Province are presented in Fig 6.

297 Our results displayed that the highest active cases in world, Iran, and Fars Province was  
298 related to March 11 (GR=1.95), Feb 26 (GR=2.41), and March 15 (GR=4.8), respectively.

299 Also, the outcome stated that GR average of active cases in world, Iran, and Fars Province  
300 reported since March 1 to April 10 was 1.13, 1.13, and 1.25, respectively. Our observations

301 demonstrated that the highest GR of active cases in Fars Province was on March 16

302 (GR=4.80), March 09 (GR=3.20), March 20 (GR=2.40), March 22 (GR=2.10), April 1<sup>st</sup>

303 (GR=2.10), and March 26 (GR=1.90). On the other hand, the analyses indicated that between

304 February 27 and February 29, the GR of active cases was zero in Fars Province, followed by

305 a GR value of 0.3 in March 14, March 19, and March 21, whereas the lowest GR of active

306 cases in world and Iran observed on March 4 (GR=0.89) and March 3 (GR=0.67)

307 respectively.

**Fig 6.** Growth rate of active cases in world, Iran, and Fars Province 308

309

310 Results of death cases in world, Iran, and Fars Province are given in Fig 7.

**Fig 7.** Growth rate of death cases in world, Iran, and Fars Province 311

312

313 In total of 1762 active cases of COVID-19 in Fars Province, 42 died between February 24  
314 and April 10. The highest GR of death cases in Fars Province was reported on March 24  
315 (GR=4.00), March 26 (GR=3.00), March 22 (GR=2.00), March 4 (GR=2.00), and April 5  
316 (GR= 2.00). Our analyses showed that since March 5 to March 11, March 15 to March 21,  
317 March 28 to April 4, and April 5 to April 8, the GR of death cases was equal to zero.  
318 Although the deaths on March 31, April 3, April 7, and April 10 were 3, 2, 4, and 1,  
319 respectively, the daily growth rate is zero. Also, average of the GR in Fars Province during  
320 41 days was 0.49, whereas this rate in world and Iran was observed as 1.15 and 1.10,  
321 respectively. Fig 7 shows that the highest GR of death cases in world and Iran was nearly  
322 equal during March 08 (GR=2.17) and March 03 (GR=2.50). In contrast, the lowest rate of  
323 death case was observed on March 09 (GR=0.87), April 08 (GR=0.87), and March 04  
324 (GR=0.60).

325 Results of active cases in 31 provinces of Iran country by March 25 is presented in Fig 8.  
326 Observations indicate that the number of active cases in the 100,000 population vary from 0.4  
327 to 13.1. This figure also shows that provinces of Bushehr and Fars have the lowest  
328 cumulative rate of active cases, whereas the highest rate was observed in Qom, Semnan,  
329 Mazandaran, Gilan, and Golestan. The Qom Province was the first place in Iran where the  
330 outbreak of COVID-19 was recorded.

331 **Fig 8.** Results of active cases in 31 provinces of Iran country by March 25, 2020

332

333

334 A comparison among age class of death cases in China, Iran, and Fars Province is presented  
335 in Table 1. Percentage of death cases in China was related to February 29, whereas for Iran  
336 and Fars Province it is related to March 14 and March 31, respectively. Following Table 1

337 show that age class > 50 years old lie in the highest class of death rate. So, this age class of  
338 above 50 years is highly sensitive to COVID-19.

Table 1 Comparison of age in death cases of China, Iran, and Fars Province

Country	China	Iran	Fars Province
Age	Death Rate (%)	Death Rate (%)	Death Rate (%)
>50 years old	93.7	84.15	78
10- 50 years old	6.3	15.46	22
<10 years old	0	0.39	0

339

#### 340 **Validation outcome of outbreak risk map**

341 The ROC-AUC curve cross-validation technique is utilized in this research for validating the  
342 COVID-19 outbreak risk map generated by SVM. The model achieved an AUC value of  
343 0.786 and a standard error of 0.031 indicating a good predictive accuracy when cross-verified  
344 using the remaining 30% testing dataset collected on March 20, 2020 (Fig 9 and Table 2).

**Table 2** Area under the curve based on data from March 20, 2020

Area	Standard Error	Asymptotic Significant	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
0.786	0.031	0.000	0.726	0.846

345

346 **Fig 9.** Receiver operator characteristic (ROC) curve based on testing data from March 20,  
347 2020

348 When tested with active case locations on March 29, 2020, the model achieved an increased  
349 AUC value of 0.799 which proves the stable and good forecast precision of the outbreak risk

350 map (Fig 10 and Table 3). Also, change detection on April 10, 2020 show that accuracy of  
 351 the built models is increased to 86.6% (AUC=0.868) (Fig 11 and Table 4).

**Table 3** Area under the curve based on data from March 29, 2020

Area	Standard Error	Asymptotic Significant	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
0.799	0.022	0.000	0.756	0.841

352

**Table 4** Area under the curve based on data from April 10, 2020

Area	Standard Error	Asymptotic Significant	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.868	.015	.000	.838	.898

353

354 **Fig 10.** Receiver operator characteristic (ROC) curve based on data from March 29, 2020

355 **Fig 11.** Receiver operator characteristic (ROC) curve based on data from April 10, 2020

356

### 357 **Comparison of Fars province and Iran infection cases**

358 Two tools have been applied to compare the general trend of infection in Fars province and  
 359 Iran. The first one is a third-degree polynomial model that is presented in Fig 12. Another  
 360 quantitative model is an ARMA presented in Table 5. Fig 12 shows the trend of infection  
 361 cases in Iran and Fars province, where predicted values extraordinarily keep pace with the  
 362 actual values.  $R^2$  values also indicate that estimated models have significant predictive power.  
 363 The infection cases are increasing over the selected horizon.

364 **Fig 12.** Actual cases versus estimated cases in Fars province and Iran



365 The first derivative of the estimated model which turns it to a second-degree polynomial  
 366 equation, represents the daily infection cases. Based on the daily infection model, there is a  
 367 turning point for both Iran and provincial cases. It was found that the turning point for  
 368 provincial daily infection is 75. In other words, after 75 days the decreasing trend in the daily  
 369 infection is expected.

**Table 5** The results of autoregressive and moving average (ARMA) model for COVID-19 infection cases of Fars province and Iran

	Regressor	Coefficient	Standard error	t-statistics	probability	
Fars province	Constant	596.015	478.41	1.24	0.221	
	AR(1)	1.432	0.056	25.33	0.000	
	AR(3)	-0.438	0.056	-7.79	0.000	
	Ma(4)	0.474	0.122	3.89	0.000	
	Adjusted R <sup>2</sup>	0.997				
	Q(1)a	1.653			0.199	
	Q(2)a	1.875			0.392	
	Heteroskedasticity (ARCH)	0.832			0.367	
	Jarque Berra	1.083			0.581	
	Inverted AR roots	0.99				
	Inverted Ma roots	-0.85				
	Iran	Constant	43360.05	32082.85	1.35	0.184
		AR(1)	1.489	0.015	99.48	0.000
		AR(3)	-0.492	0.014	-35.01	0.000
MA(1)		0.848	0.075	4.16	0.000	
Adjusted R <sup>2</sup>		0.999				
Q(1)a		1.169			0.194	
Q(2)a		1.785			0.410	
Heteroskedasticity (ARCH)		1.149			0.289	
Jarque Berra		0.088			0.956	
Inverted AR roots		0.96				
Inverted Ma roots		0.59				

370 a<sup>Q(p)</sup> is the significance level of the Ljung–Box statistics in which the first p of the residual autocorrelations are jointly equal  
 371 to zero.  
 372

373 The corresponding value for Iran is 211 that is much higher than the provincial one. There are  
374 some evidences showing that a turning point in infection is expected. For instance, it has been  
375 reported for SARS incidence [43], HAV [44], ARI [45], and for A (H1N1)v [46]. It is worth  
376 noting that a turning point means that after passing the peak it is expected to show a  
377 decreasing trend. In the 38<sup>th</sup> day of infection, Fars province accounts for around 2.84% of the  
378 total Iranian cases while its population share is more than 6% (Statistical Center of Iran,  
379 2016). Regarding the values obtained for turning points and the infection share, the measures  
380 taken by the provincial government may be considered more effective than those taken in  
381 other provinces as a whole. However, it should be taken into consideration that Fars province  
382 experienced its first infection cases 7 days after Qom and Tehran, provinces that are  
383 considered as starting point for virus outbreak in Iran. This might have given the provincial  
384 governmental body and the households to take measures to cope with the widespread  
385 outbreak. It is worth noting that the comparison of the specified models is more appropriate  
386 to investigate the effectiveness of the measures taken by the corresponding health body rather  
387 than using it to predict the future values.

388 The ARMA time series models for infection variables of the Fars province and Iran are  
389 presented in Table 5. These models may show the generating process of the variables in time  
390 horizon. It is worth noting that in order to have more comparable models, a 38-day time  
391 horizon is selected. This is the period of time that data are available, starting on 19th of  
392 February for Iran and one week later for Fars province. As shown in Table 4, the both series  
393 are generated by an ARMA (2, 1) process. However, the absolute values of the AR terms for  
394 Fars province are lower than those of Iran, indicating a slower process of increasing trend for  
395 Fars province compared to those of Iran. However, regarding the values for AR roots, the  
396 autoregressive (AR) process for both models isn't explosive. Benvenuto et al. [42] also found  
397 that COVID-2019 spread tends to reveal slightly decreasing spread. In addition,

398 Heteroscedasticity (ARCH) were found to be insignificant in both models, indicating that the  
399 infection cases tend to show insignificant fluctuations. This is the fact that is not easily  
400 captured in the trends shown in Fig 12. Generally speaking, the diagnostic statistics indicate  
401 that the estimated models are acceptable since Q-statistics reveal that the residuals are not  
402 significantly correlated and the Jarque Berra statistic support the normality of residuals at  
403 conventional significance level. Also, ARCH effect was not significant, indicating a low  
404 volatility in the infection cases trend. In addition, all AR and MA roots were found to lie  
405 inside the unit circle, indicating that ARMA process is (covariance) stationary and invertible.

#### 406 **Discussion**

407 There is a great necessity for new robust scientific outcomes that could aid in containing and  
408 preventing the COVID-19 pandemic from spreading. The spatial mapping of COVID-19  
409 outbreak risk can aid governments and policy-makers in implementing strict measures in  
410 certain regions of a city or a country where the risk of outbreak is very high. It is therefore  
411 crucial to identify the regions that would have high outbreak risk through predictive  
412 modelling with the help of machine learning algorithms (MLAs). In recent times, MLAs have  
413 demonstrated promising results in forecasting the epidemic outbreak risk [17]. In this  
414 research, the SVM model showing good forecast accuracy was used for mapping the outbreak  
415 risk of COVID-19. Similarly, Mohammadinia et al. [20] revealed that GWR and SVM had  
416 the highest precision in mapping the occurrence of leptospirosis. Ding et al. [47] employed  
417 three MLAs including SVM, RF and GBM for mapping the transmission risk assessment of  
418 mosquito-borne diseases and disclosed that all three MLAs acquired excellent validation  
419 outcome. Machado et al. [48] also applied RF, SVM and GBM in modelling the porcine  
420 epidemic diarrhoea virus and demonstrated 90% specificity values in case of SVM. Tien Bui  
421 et al. [17] stated that SVM achieved an AUC value of 0.968 in mapping the susceptibility to  
422 malaria. The ability to classify inseparable data classes is the greatest benefit of SVM model

423 [49]. It is among the most precise and robust MLAs [50]. SVM can be useful and has higher  
424 prediction accuracy when it comes to handling a small dataset. However, Huang and Zhao  
425 [51] demonstrated that SVM also yields excellent precision in predictive modelling when a  
426 large dataset is utilized. The algorithm has a very low probability to overfit and is not  
427 disproportionately impacted by noisy data [49]. Behzad et al. [52] revealed that SVM had  
428 huge capacity in simplification and had enduring forecast accuracy. It should be also noted  
429 that the predictive exactness of SVM model largely depends on the choice of kernel function  
430 [50]. Among the four kernel functions of SVM, RBF has been proved to generate high  
431 accuracy models [49]. SVM includes diverse kinds of categorization functions which are  
432 responsible for assessing overfitting and simplifying data that needs a minor tuning of model  
433 parameters [53]. The significance of each effective factor employed in this research is  
434 assessed using ridge regression. Since, there is no previous study in COVID-19 that outlines  
435 the proper effective factors. The outcome of this research can be very helpful for scientists in  
436 experimenting the same and additional effective factors for COVID-19 outbreak risk  
437 mapping. The proximity factors including distance from bus stations, distance from hospitals,  
438 distance from bakeries were most influential in forecasting the COVID-19 outbreak risk  
439 whereas other proximity factors such as distance from ATMs, distance from attraction sites,  
440 distance from fuel stations, distance from mosques and distance from roads had the moderate  
441 influence which is followed by MTCM, density of cities and density of villages. It should be  
442 noted that climatic factors including MTWM, PWM and PDM had the least significance in  
443 mapping the outbreak risk. From this, it can be concluded that precipitation factors PWM and  
444 PDM are not associated with the transmission of COVID-19 in Fars Province whereas in case  
445 of temperature factors MTCM had moderate influence in mapping COVID-19 outbreak risk  
446 but MTWM exhibited a least significance. This outcome reveals that proximity factors had  
447 high influence in the transmission of SARS-CoV-2. In addition, the study conducted

448 disclosed that increase in temperature will not decline the SARS-CoV-2 cases, although it has  
449 been also revealed that increase in temperature and absolute humidity could decrease the  
450 death of patients affected by COVID-19 [54]. A third-degree polynomial and ARMA models  
451 were applied to examine the behaviour of infection in Fars province and Iran. The general  
452 trend of infection in Iran and Fars province are similar while more explosive behavior for  
453 Iran's cases is expected. The methodology and effective factors used in this research can be  
454 adapted in studies investigated in other parts of the world for preventing and controlling the  
455 outbreak risk of COVID-19.

456

## 457 **Conclusions**

458 Mapping of SARS-CoV-2 outbreak risk can aid decision makers in drafting effective policies  
459 to minimize the spread of the disease. In this research, GIS based SVM was used for mapping  
460 the COVID-19 outbreak risk in Fars Province of Iran. Sixteen effective factors including  
461 MTCM, MTWM, PWM, PDM, distance from roads, distance from mosques, distance from  
462 hospitals, distance from fuel stations, human footprint, density of cities, distance from bus  
463 stations, distance from banks, distance from bakeries, distance from attraction sites, distance  
464 from automated teller machines (ATMs) and density of villages were selected along with the  
465 locations of active cases of SARS-CoV-2. The results of ridge regression revealed that  
466 distance from bus stations, distance from hospitals, and distance from bakeries had the  
467 highest significance and the outcome was utilized in mapping the outbreak risk of the  
468 pandemic with the help of SVM. The generated model had good predictive accuracy of 0.786  
469 and 0.799 when verified with the locations of active cases during March 20 and March 29,  
470 2020. The Iranian government should take restrict preventive measures for controlling the  
471 outbreak of SARS-CoV-2 in Shiraz as a tourism destination and the counties having high

472 risk. Based on the results of polynomial and an ARMA model, the infection behavior is not  
473 expected to reveal an explosive process, however; the general trend of infection will last for  
474 several months especially in the Iran as a whole. A more slowly trend is expected in Fars  
475 Province, demonstrating extensive home quarantine and travel and movement restrictions  
476 were good strategies for disease control in Fars province. The main policy implication is that  
477 the infection cases, to some extent, may be controlled using more effective measures.  
478 Although, the estimated models may be used to predict the infection in following days,  
479 however; this contribution is less significant than the other implications derived from them.  
480 Generally speaking, it is expected to encounter a decreasing trend, however; this may be  
481 reversed if the ongoing attempts are slowed down, pointing out the need to keep the measures  
482 like quarantine or even to try more restricting attempts.

483

#### 484 **Competing interests**

485 The authors declare that they have no competing interests.

#### 486 **Funding**

487 Shiraz University, Iran, Grant No. 96GRD1M271143.

#### 488 **Availability of data and materials**

489 All data and materials used in this work were publicly available.

#### 490 **Ethics approval and consent to participate**

491 The ethical approval or individual consent was not applicable

#### 492 **Authors' contributions**

493

494 HRP, SP, BH, ZF, NS, MHT, BH, SB, and JPT contributed to study design, the literature  
495 search, data collection, data analysis, software working, and writing of this article. All  
496 authors read and approved the final draft of the manuscript.

497  
498  
499  
500  
501  
502

**Consent for publication**

Not applicable.

503 **References**

- 504 1. Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Feng, Z. Early transmission  
505 dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *New England*  
506 *Journal of Medicine*. 2020. <https://doi.org/10.1056/nejmoa2001316>
- 507 2. Ma, Y., Zhao, Y., Liu, J., He, X., Wang, B., Fu, S., Luo, B. Effects of temperature  
508 variation and humidity on the death of COVID-19 in Wuhan, China. *Science of the Total*  
509 *Environment*. 2020; 138226. <https://doi.org/10.1016/j.scitotenv.2020.138226>
- 510 3. Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Cao, B. Clinical features of patients  
511 infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*. 2020; 395(10223), 497–  
512 506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)
- 513 4. Wang, C., Horby, P. W., Hayden, F. G., Gao, G. F. A novel coronavirus outbreak of global  
514 health concern. *The Lancet*. 2020; 395, 470–473. [https://doi.org/10.1016/S0140-](https://doi.org/10.1016/S0140-6736(20)30185-9)  
515 [6736\(20\)30185-9](https://doi.org/10.1016/S0140-6736(20)30185-9)
- 516 5. Sohrabi, C., Alsafi, Z., O’Neill, N., Khan, M., Kerwan, A., Al-Jabir, A., Agha, R. World  
517 Health Organization declares global emergency: A review of the 2019 novel coronavirus  
518 (COVID-19). *International Journal of Surgery*. 2020; 76, 71–76.  
519 <https://doi.org/10.1016/j.ijisu.2020.02.034>
- 520 6. WHO, 2020a. WHO characterizes COVID-19 as a pandemic, 2020 (3).
- 521 7. WHO, 2020b. Coronavirus disease 2019 (COVID-19) Situation Report –70.

- 522 8. Remuzzi, A., Remuzzi, G. COVID-19 and Italy: what next? *The Lancet*. 2020;  
523 [https://doi.org/10.1016/S0140-6736\(20\)30627-9](https://doi.org/10.1016/S0140-6736(20)30627-9)
- 524 9. Arab-Mazar, Z., Sah, R., Rabaan, A. A., Dhama, K., Rodriguez-Morales, A. J. Mapping  
525 the incidence of the COVID-19 hotspot in Iran – Implications for Travellers. *Travel Medicine*  
526 *and Infectious Disease*. 2020; 101630. <https://doi.org/10.1016/j.tmaid.2020.101630>
- 527 10. Takian, A., Raofi, A., Kazempour-Ardebili, S. COVID-19 battle during the toughest  
528 sanctions against Iran. *Lancet* (London, England). 2020; (20), 30668.  
529 [https://doi.org/10.1016/S0140-6736\(20\)30668-1](https://doi.org/10.1016/S0140-6736(20)30668-1)
- 530 11. Singh, A. K., Singh, A., Shaikh, A., Singh, R., Misra, A. Chloroquine and  
531 hydroxychloroquine in the treatment of COVID-19 with or without diabetes: A systematic  
532 search and a narrative review with a special reference to India and other developing  
533 countries. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*. 2020;  
534 <https://doi.org/10.1016/j.dsx.2020.03.011>
- 535 12. McCloskey, B., Zumla, A., Ippolito, G., Blumberg, L., Arbon, P., Cicero, A., Borodina,  
536 M. Mass gathering events and reducing further global spread of COVID-19: a political and  
537 public health dilemma. *The Lancet*. 2020; [https://doi.org/10.1016/S0140-6736\(20\)30681-4](https://doi.org/10.1016/S0140-6736(20)30681-4)
- 538 13. Zhou, C., Su, F., Pei, T., Zhang, A., Du, Y., Luo, B., Xiao, H. COVID-19: Challenges to  
539 GIS with Big Data. *Geography and Sustainability*. 2020;  
540 <https://doi.org/10.1016/j.geosus.2020.03.005>
- 541 14. Sánchez-Vizcaíno, F., Martínez-López, B., Sánchez-Vizcaíno, J. M. Identification of  
542 suitable areas for the occurrence of Rift Valley fever outbreaks in Spain using a multiple  
543 criteria decision framework. *Veterinary Microbiology*. 2013; 165(1–2), 71–78.  
544 <https://doi.org/10.1016/j.vetmic.2013.03.016>



- 545 15. Reeves, T., Samy, A. M., Peterson, A. T. MERS-CoV geography and ecology in the  
546 Middle East: Analyses of reported camel exposures and a preliminary risk map. BMC  
547 Research Notes. 2015; 8(1), 1–7. <https://doi.org/10.1186/s13104-015-1789-1>
- 548 16. Nyakarahuka, L., Ayebare, S., Mosomtai, G., Kankya, C., Lutwama, J., Mwiine, F. N.,  
549 Skjerve, E. Ecological Niche Modeling for Filoviruses: A Risk Map for Ebola and Marburg  
550 Virus Disease Outbreaks in Uganda. PLoS Currents. 2017; 9.  
551 <https://doi.org/10.1371/currents.outbreaks.07992a87522e1f229c7cb023270a2af1>
- 552 17. Tien Bui, Q. T., Nguyen, Q. H., Pham, V. M., Pham, M. H., Tran, A. T. Understanding  
553 spatial variations of malaria in Vietnam using remotely sensed data integrated into GIS and  
554 machine learning classifiers. Geocarto International. 2019; 34(12), 1300–1314.  
555 <https://doi.org/10.1080/10106049.2018.1478890>
- 556 18. Jiang, D., Hao, M., Ding, F., Fu, J., & Li, M. Mapping the transmission risk of Zika virus  
557 using machine learning models. Acta Tropica. 2018; 185, 391–399.  
558 <https://doi.org/10.1016/j.actatropica.2018.06.021>
- 559 19. Carvajal, T. M., Viacrusis, K. M., Hernandez, L. F. T., Ho, H. T., Amalin, D. M.,  
560 Watanabe, K. Machine learning methods reveal the temporal pattern of dengue incidence  
561 using meteorological factors in metropolitan Manila, Philippines. BMC Infectious Diseases.  
562 2018; 18(1), 183. <https://doi.org/10.1186/s12879-018-3066-0>
- 563 20. Mohammadinia, A., Saeidian, B., Pradhan, B., Ghaemi, Z. Prediction mapping of human  
564 leptospirosis using ANN, GWR, SVM and GLM approaches. BMC Infectious Diseases.  
565 2019; 19(1), 971. <https://doi.org/10.1186/s12879-019-4580-4>
- 566 21. Kamel Boulos, M. N., Geraghty, E. M. Geographical tracking and mapping of  
567 coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-  
568 CoV-2) epidemic and associated events around the world: how 21st century GIS technologies

- 569 are supporting the global fight against outbreaks and epidemics. *International Journal of*  
570 *Health Geographics*. 2020; 19, 8. <https://doi.org/10.1186/s12942-020-00202-8>
- 571 22. Chen, S., Yang, J., Yang, W., Wang, C., Bärnighausen, T. COVID-19 control in China  
572 during mass population movements at New Year. *The Lancet*. 2020; 395, 764–766.  
573 [https://doi.org/10.1016/S0140-6736\(20\)30421-9](https://doi.org/10.1016/S0140-6736(20)30421-9)
- 574 23. Wang, M., Jiang, A., Gong, L., Luo, L., Guo, W., Li, C., Li, H. Temperature significant  
575 change COVID-19 Transmission in 429 cities. *MedRxiv*. 2020; 20025791.  
576 <https://doi.org/10.1101/2020.02.22.20025791>
- 577 24. Tan, J., Mu, L., Huang, J., Yu, S., Chen, B., Yin, J. An initial investigation of the  
578 association between the SARS outbreak and weather: With the view of the environmental  
579 temperature and its variation. *Journal of Epidemiology and Community Health*. 2005; 59(3),  
580 186–192. <https://doi.org/10.1136/jech.2004.020180>
- 581 25. Correa Ayram, C. A., Mendoza, M. E., Etter, A., Pérez Salicrup, D. R. Anthropogenic  
582 impact on habitat connectivity: A multidimensional human footprint index evaluated in a  
583 highly biodiverse landscape of Mexico. *Ecological Indicators*. 2017; 72, 895–909.  
584 <https://doi.org/10.1016/j.ecolind.2016.09.007>
- 585 26. Tarwater, P. M., Martin, C. F. Effects of population density on the spread of disease.  
586 *Complexity*. 2001; 6(6), 29–36. <https://doi.org/10.1002/cplx.10003>
- 587 27. Schmidt, W. P., Suzuki, M., Thiem, V., White, R. G., Tsuzuki, A., Yoshida, L. M.,  
588 Ariyoshi, K. Population density, water supply, and the risk of dengue fever in Vietnam:  
589 Cohort study and spatial analysis. *PLoS Medicine*. 2011; 8(8).  
590 <https://doi.org/10.1371/journal.pmed.1001082>

- 591 28. Gilbert, M., Pullano, G., Pinotti, F., Valdano, E., Poletto, C., Boëlle, P. Y., Colizza, V.  
592 Preparedness and vulnerability of African countries against importations of COVID-19: a  
593 modelling study. *The Lancet*. 2020; 395(10227), 871–877. <https://doi.org/10.1016/S0140->  
594 [6736\(20\)30411-6](https://doi.org/10.1016/S0140-6736(20)30411-6)
- 595 29. Pourghasemi, H. R., Kariminejad, N., Amiri, M., Edalat, M., Zarafshar, M., Blaschke, T.,  
596 Cerda, A. Assessing and mapping multi-hazard risk susceptibility using a machine learning  
597 technique. *Scientific Reports*. 2020; 10(1), 3203. <https://doi.org/10.1038/s41598-020-60191-3>
- 598 30. Hoerl, A. E., Kennard, R. W. Ridge Regression: Biased Estimation for Nonorthogonal  
599 Problems. *Technometrics*. 1970; 12(1), 55–67.  
600 <https://doi.org/10.1080/00401706.1970.10488634>
- 601 31. Tikhonov, A. N., Goncharsky, A. V., Stepanov, V. V., Yagola, A. G., Tikhonov, A. N.,  
602 Goncharsky, A. V., Yagola, A. G. Regularization methods. In *Numerical Methods for the*  
603 *Solution of Ill-Posed Problems*. 1995; 7–63. [https://doi.org/10.1007/978-94-015-8480-7\\_2](https://doi.org/10.1007/978-94-015-8480-7_2)
- 604 32. Vapnik, V.N., An overview of statistical learning theory. *IEEE Trans. Neural Network*.  
605 1999; 10 (5): 988–999. <https://doi.org/10.1109/72.788640>.
- 606 33. Gayen, A., Pourghasemi, H. R., Saha, S., Keesstra, S., Bai, S. Gully erosion susceptibility  
607 assessment and management of hazard-prone areas in India using different machine learning  
608 algorithms. *Science of the Total Environment*. 2019; 668, 124–138.  
609 <https://doi.org/10.1016/j.scitotenv.2019.02.436>
- 610 34. Yousefi, S., Sadhasivam, N., Pourghasemi, H. R., Ghaffari Nazarlou, H., Golkar, F.,  
611 Tavangar, S., Santosh, M. Groundwater spring potential assessment using new ensemble data  
612 mining techniques. *Measurement*. 2020; 157, 107652.  
613 <https://doi.org/10.1016/j.measurement.2020.107652>

- 614 35. Pourghasemi, H. R., Sadhasivam, N., Kariminejad, N., Collins, A. Gully erosion spatial  
615 modelling: Role of machine learning algorithms in selection of the best controlling factors  
616 and modelling process. *Geoscience Frontiers*. 2020; <https://doi.org/10.1016/j.gsf.2020.03.005>
- 617 36. Garosi, Y., Sheklabadi, M., Conoscenti, C., Pourghasemi, H. R., Van Oost, K. Assessing  
618 the performance of GIS- based machine learning models with different accuracy measures for  
619 determining susceptibility to gully erosion. *Science of the Total Environment*. 2019; 664,  
620 1117–1132. <https://doi.org/10.1016/j.scitotenv.2019.02.093>
- 621 37. Yao, X., Tham, L. G., Dai, F. C. Landslide susceptibility mapping based on Support  
622 Vector Machine: A case study on natural slopes of Hong Kong, China. *Geomorphology*.  
623 2008; 101(4), 572–582. <https://doi.org/10.1016/j.geomorph.2008.02.011>
- 624 38. Chen, W., Pourghasemi, H. R., Naghibi, S. A. A comparative study of landslide  
625 susceptibility maps produced using support vector machine with different kernel functions  
626 and entropy data mining models in China. *Bulletin of Engineering Geology and the*  
627 *Environment*. 2018; 77(2), 647–664. <https://doi.org/10.1007/s10064-017-1010-y>
- 628 39. Dodangeh, E., Choubin, B., Eigdir, A. N., Nabipour, N., Panahi, M., Shamshirband, S.,  
629 Mosavi, A. Integrated machine learning methods with resampling algorithms for flood  
630 susceptibility prediction. *Science of the Total Environment*. 2020; 705, 135983.  
631 <https://doi.org/10.1016/j.scitotenv.2019.135983>
- 632 40. Aik, J., Heywood, A. E., Newall, A. T., Ng, L.C., Kirk, M. D., Turner, R. Climate  
633 variability and salmonellosis in Singapore – A time series analysis. *Science of the Total*  
634 *Environment*. 2018; 639, 1261-1267.
- 635 41. Enders, W. *Applied Econometric Times Series*. John Wiley & Sons. 2004.

- 636 42. Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., Ciccozzi, M. Application of  
637 the ARIMA model on the COVID-2019 epidemic dataset. *Data in Brief*. 2020; 29:105340.
- 638 43. Wong, G. Has SARS infected the property market? Evidence from Hong Kong. *Journal*  
639 *of Urban Economics*. 2008; 63, 74-95.
- 640 44. Alberts, C. J., Boyd, A., Bruisten, S. M., Heijman, T., Hogewoning, A., van Rooijen, M.,  
641 Siedenburg, E., Sonder, G. J. B. Hepatitis A incidence, seroprevalence, and vaccination  
642 decision among MSM in Amsterdam, the Netherlands. *Vaccine*. 2019; 37, 2849–2856.
- 643 45. Leonenko, V. N., Ivanov, S. V., Novoselova, Y. K. A computational approach to  
644 investigate patterns of acute respiratory illness dynamics in the regions with distinct seasonal  
645 climate transitions. *Procedia Computer Science*. 2016; 80, 2402–2412.
- 646 47. Ding, F., Fu, J., Jiang, D., Hao, M., Lin, G. Mapping the spatial distribution of *Aedes*  
647 *aegypti* and *Aedes albopictus*. *Acta Tropica*. 2018; 178, 155–162.  
648 <https://doi.org/10.1016/j.actatropica.2017.11.020>
- 649 48. Machado, G., Vilalta, C., Recamonde-Mendoza, M., Corzo, C., Torremorell, M., Perez,  
650 A., VanderWaal, K. Identifying outbreaks of Porcine Epidemic Diarrhea virus through  
651 animal movements and spatial neighborhoods. *Scientific Reports*. 2019; 9(1), 1–12.  
652 <https://doi.org/10.1038/s41598-018-36934-8>
- 653 49. Gigović, L., Pourghasemi, H. R., Drobňjak, S., Bai, S. Testing a New Ensemble Model  
654 Based on SVM and Random Forest in Forest Fire Susceptibility Assessment and Its Mapping  
655 in Serbia’s Tara National Park. *Forests*. 2019; 10(5), 408. <https://doi.org/10.3390/f10050408>
- 656 50. Abdollahi, S., Pourghasemi, H. R., Ghanbarian, G. A., Safaeian, R. Prioritization of  
657 effective factors in the occurrence of land subsidence and its susceptibility mapping using an

- 658 SVM model and their different kernel functions. *Bulletin of Engineering Geology and the*  
659 *Environment*. 2019; 78(6), 4017–4034. <https://doi.org/10.1007/s10064-018-1403-6>
- 660 51. Huang, Y., Zhao, L. Review on landslide susceptibility mapping using support vector  
661 machines. *Catena*. 2018; 165: 520–529. <https://doi.org/10.1016/j.catena.2018.03.003>
- 662 52. Behzad, M., Asghari, K., Coppola, E. A. Comparative Study of SVMs and ANNs in  
663 Aquifer Water Level Prediction. *Journal of Computing in Civil Engineering*. 2010; 24(5):  
664 408–413. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000043](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000043)
- 665 53. Joachims, T. Text categorization with support vector machines: Learning with many  
666 Relevant Features. 1998. <https://doi.org/10.1007/bfb0026683>
- 667 54. Zhu, Y., Xie, J. Association between ambient temperature and COVID-19 infection in  
668 122 cities from China. *Science of the Total Environment*, 2020; 138201.  
669 <https://doi.org/10.1016/j.scitotenv.2020.138201>
- 670

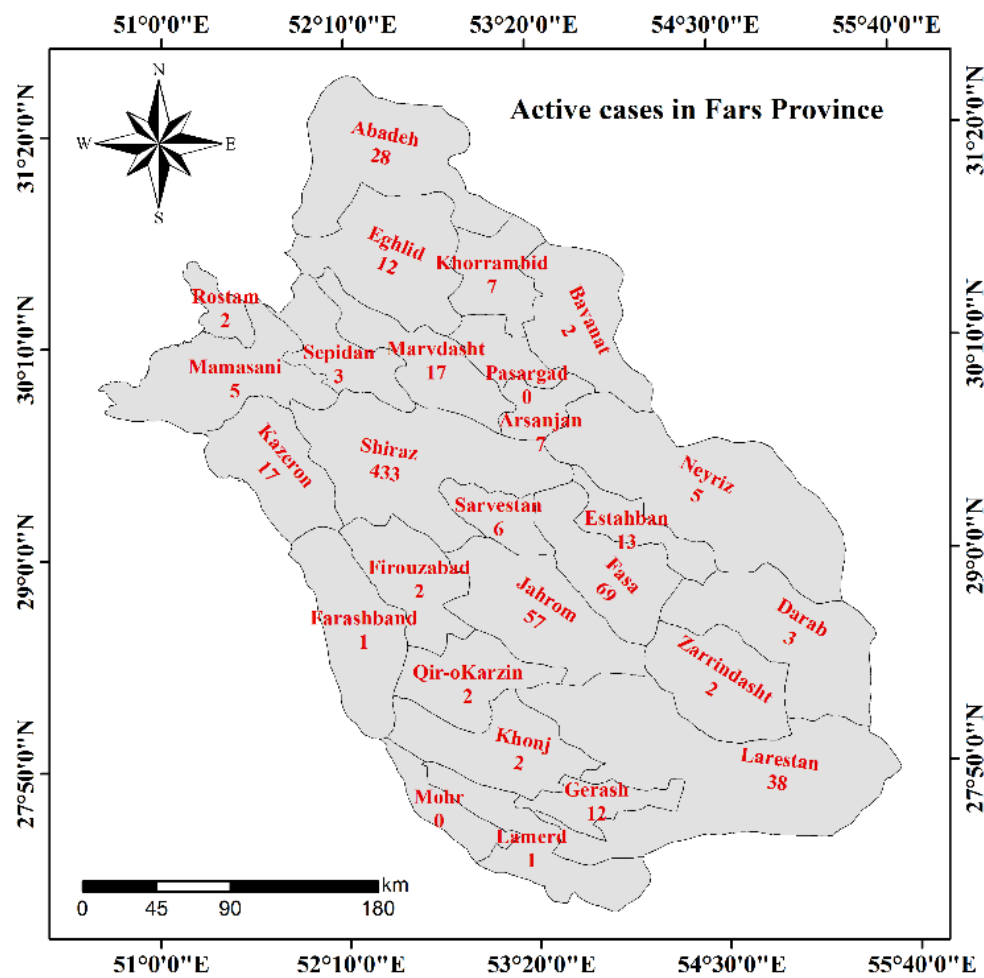


Fig. 1

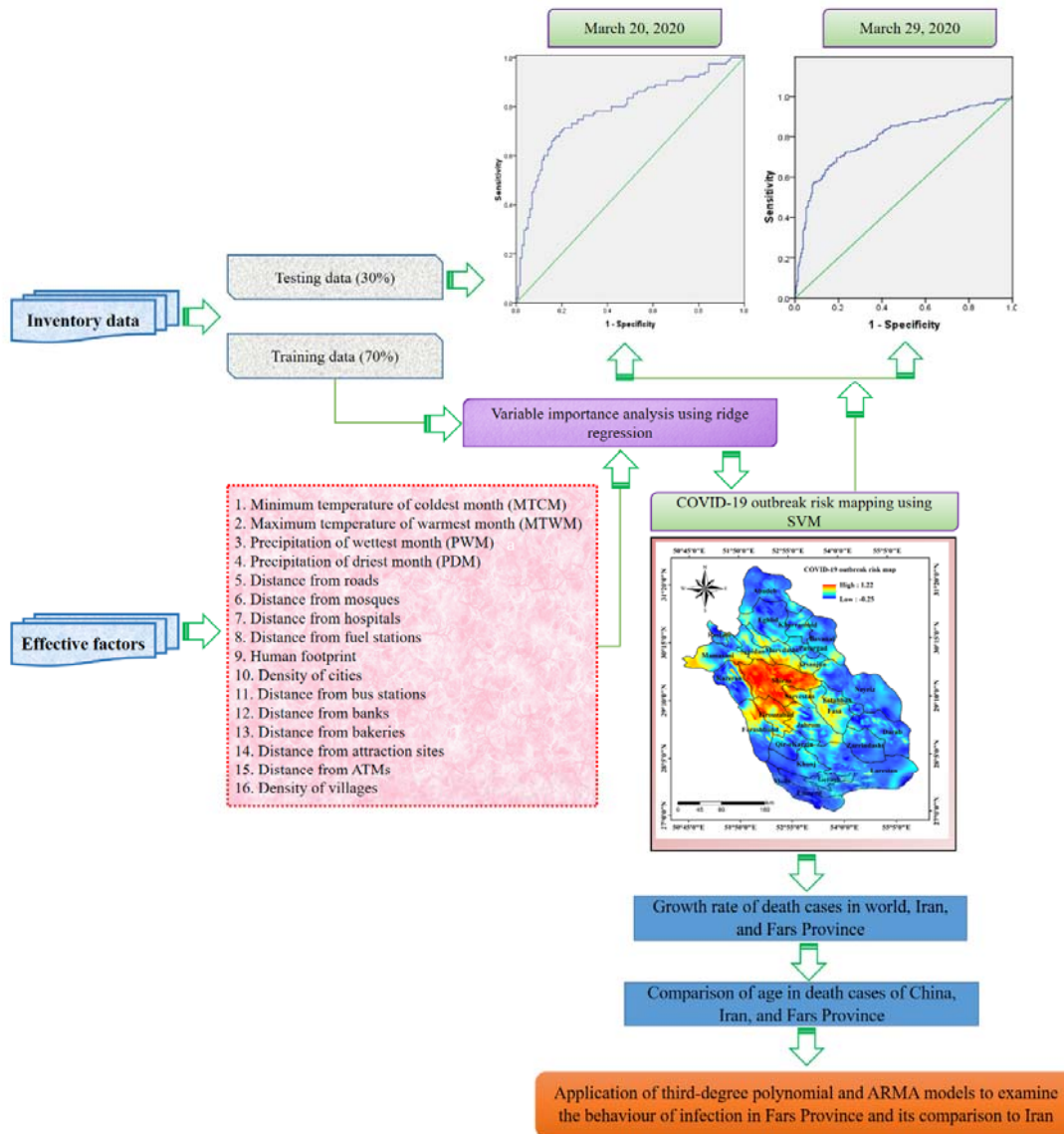
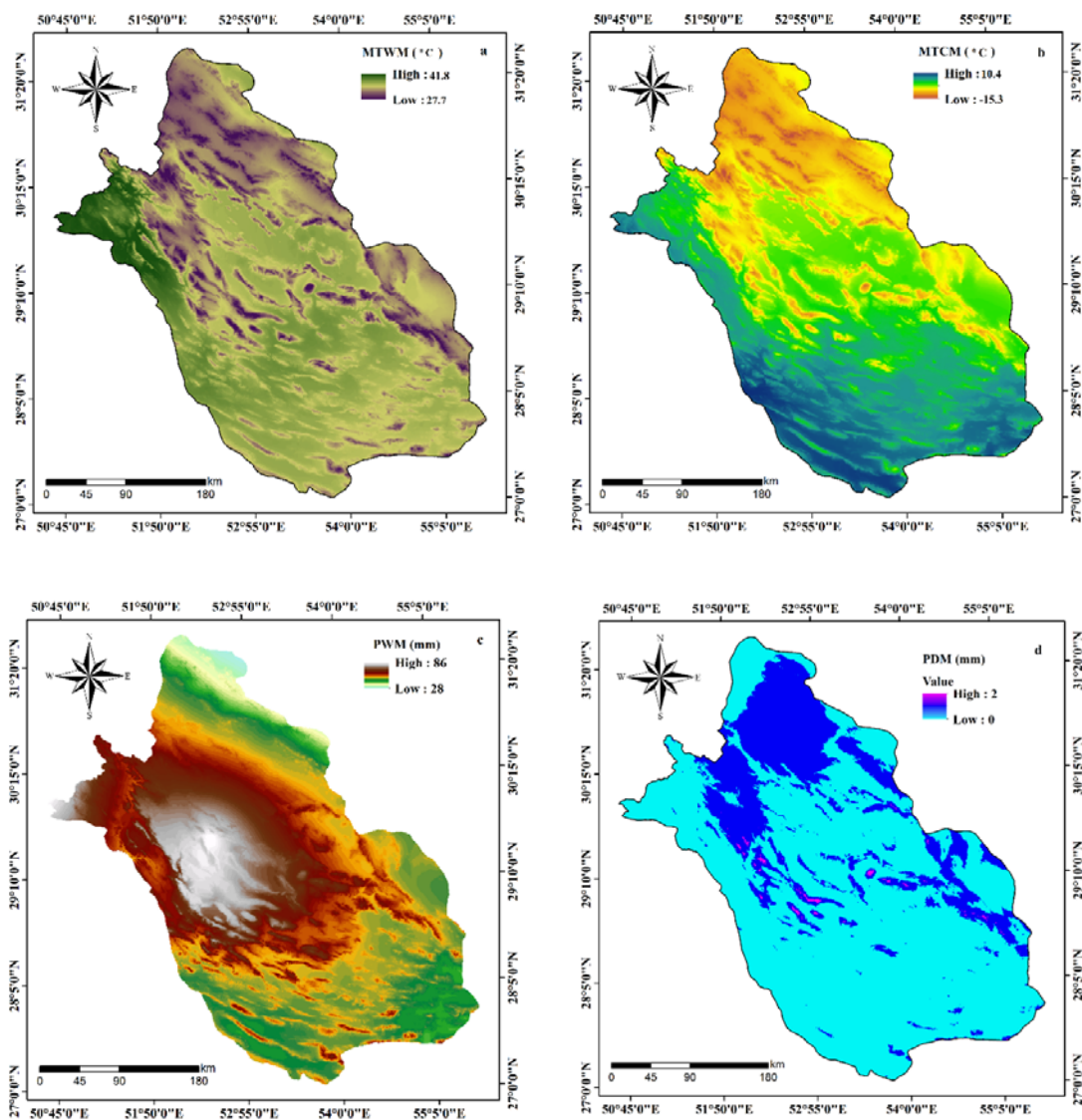
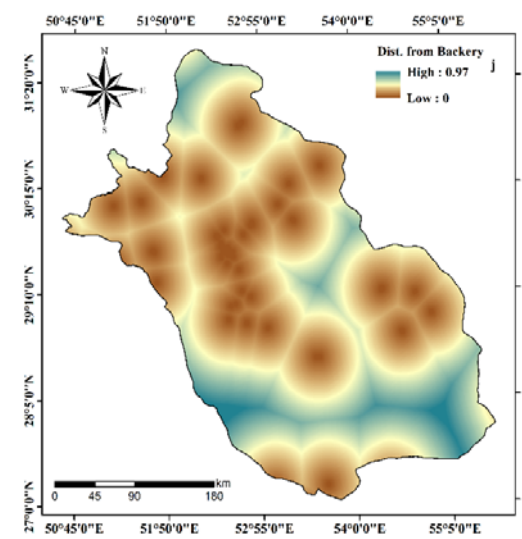
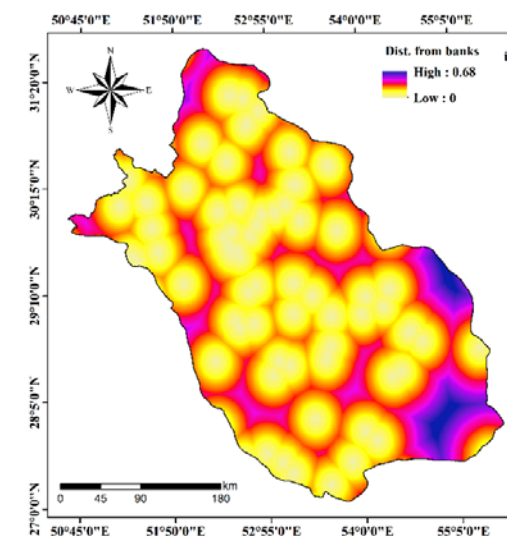
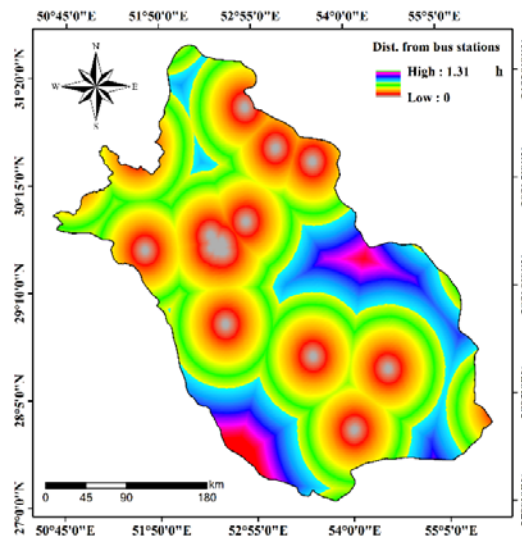
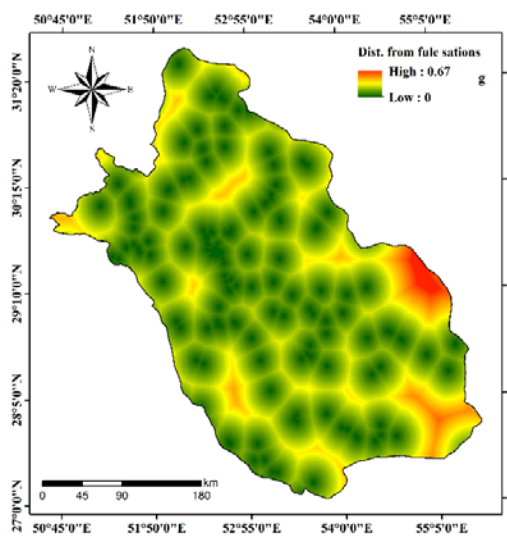
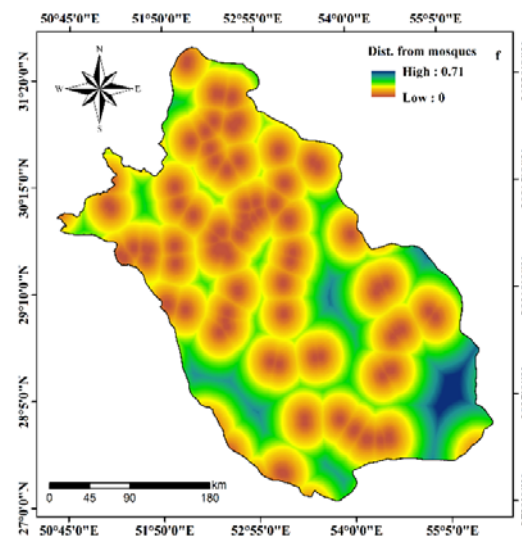
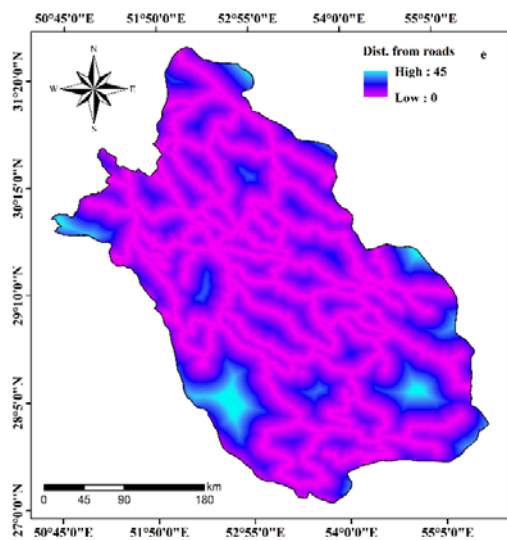


Fig. 2







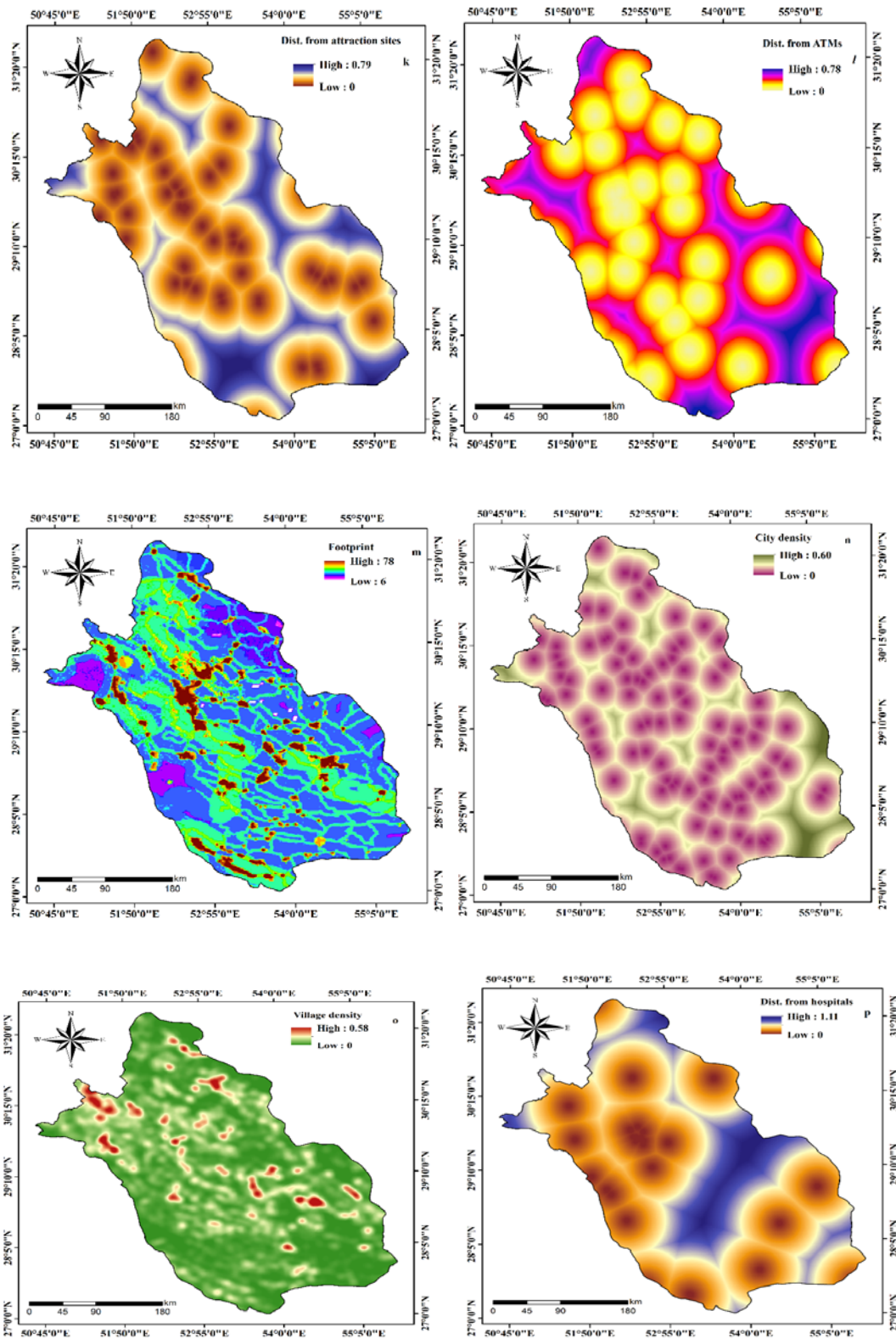


Fig. 3

### Variable Importance using Ridge Regression

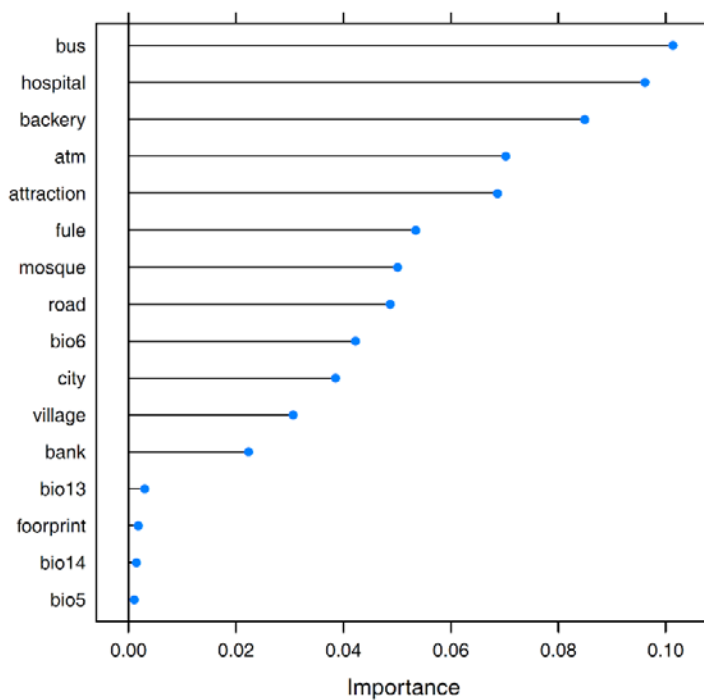


Fig. 4

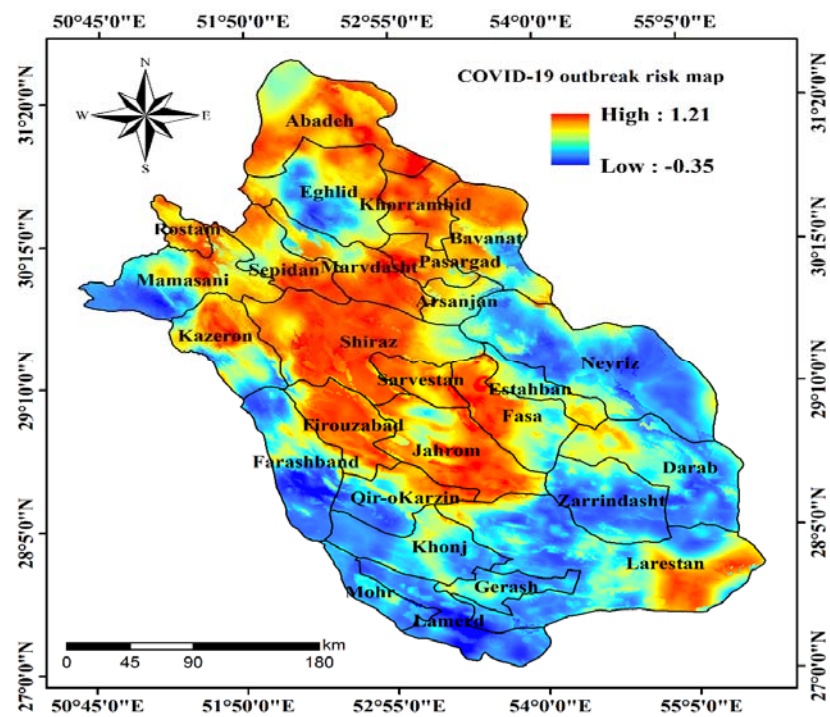
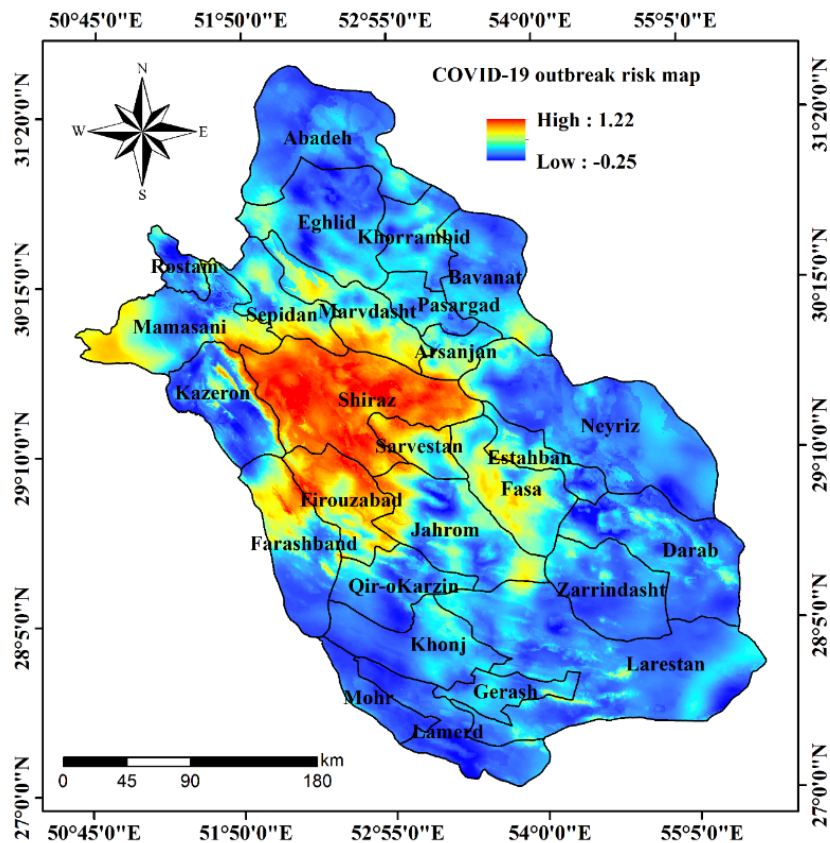


Fig. 5

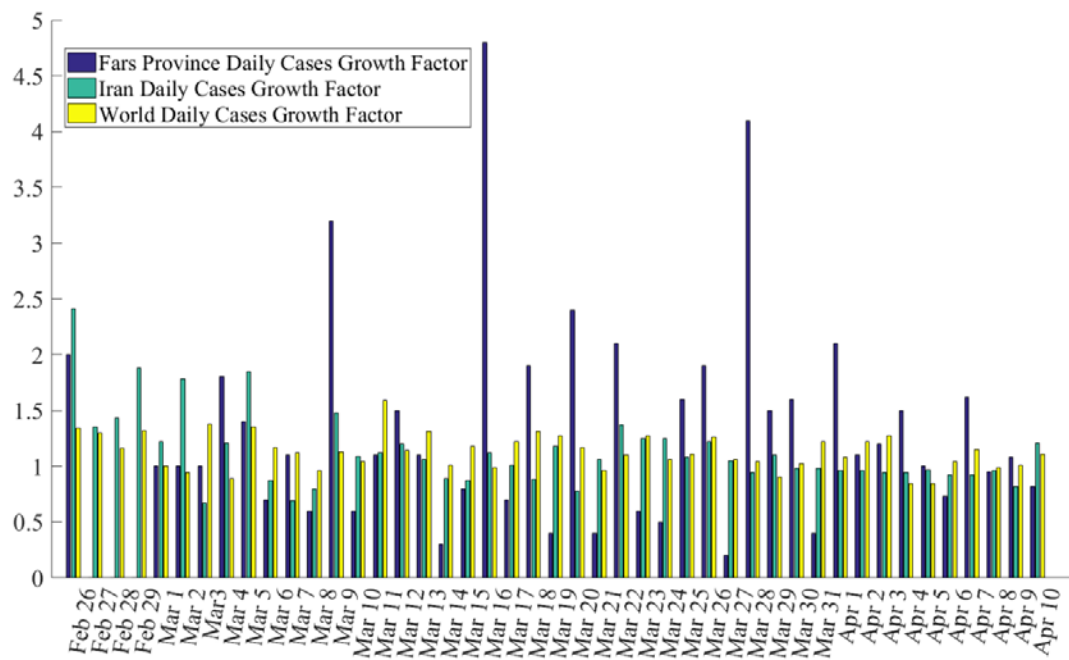


Fig. 6

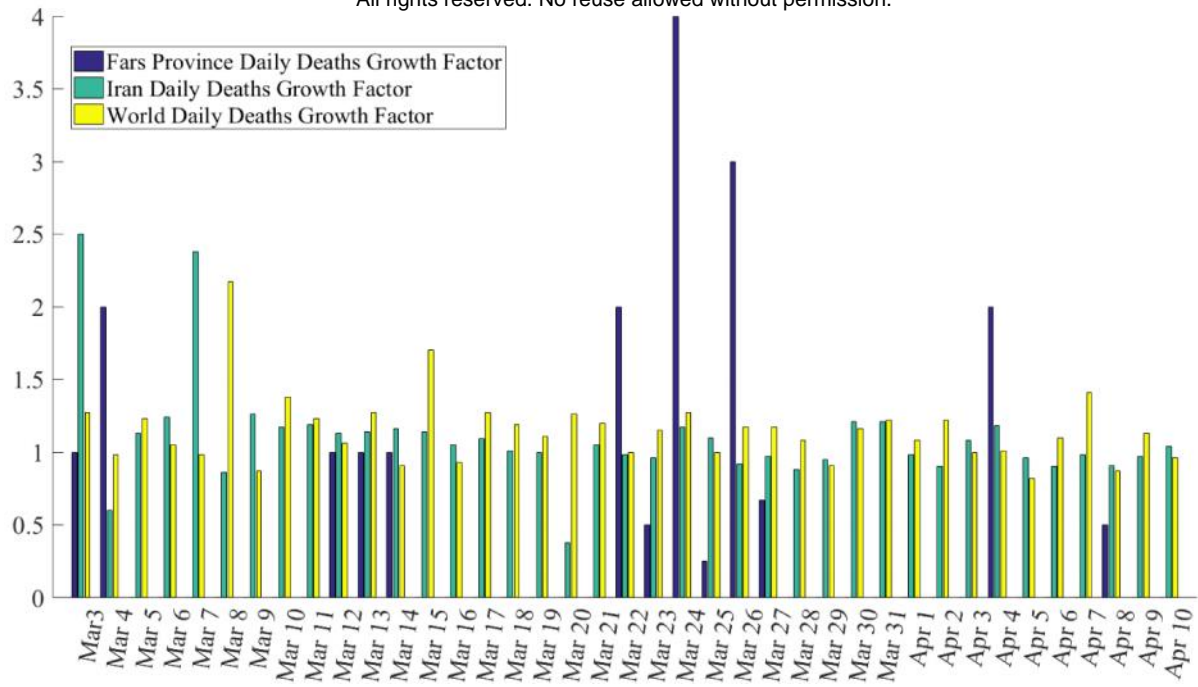


Fig. 7

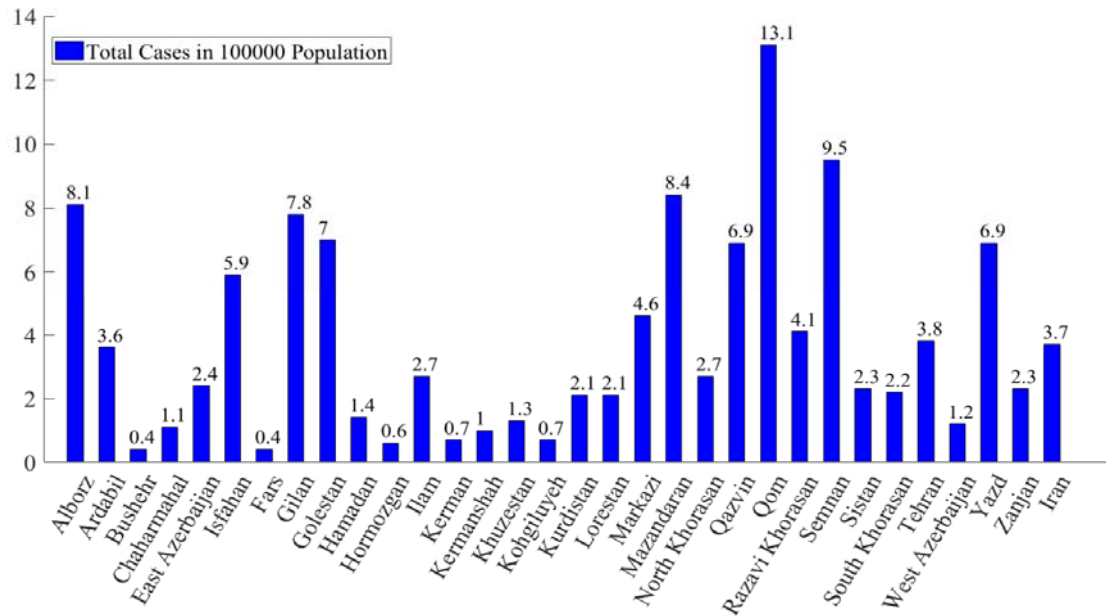


Fig. 8



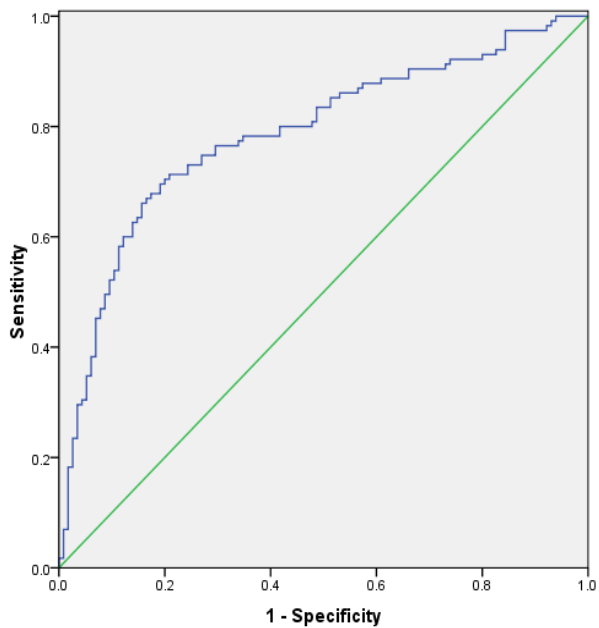


Fig. 9

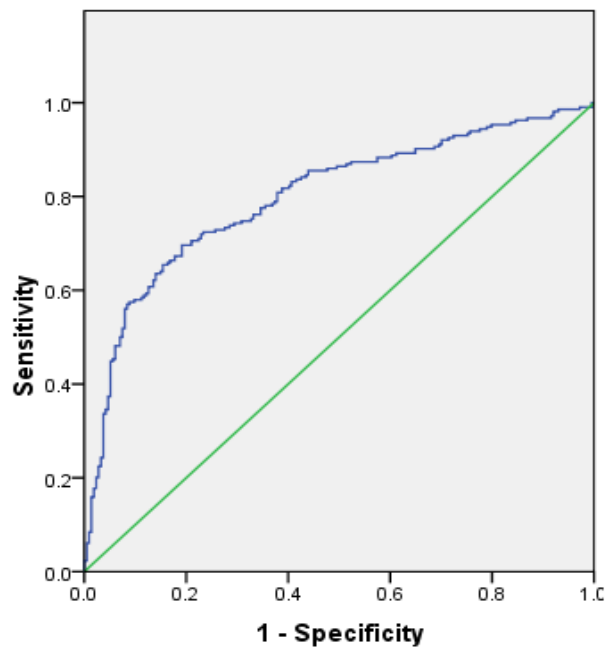


Fig. 10

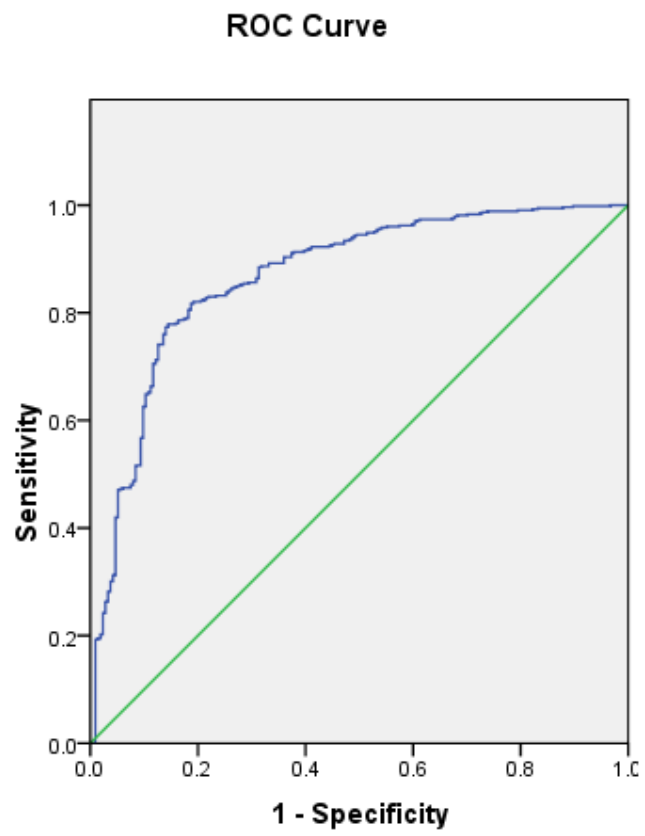


Fig. 11

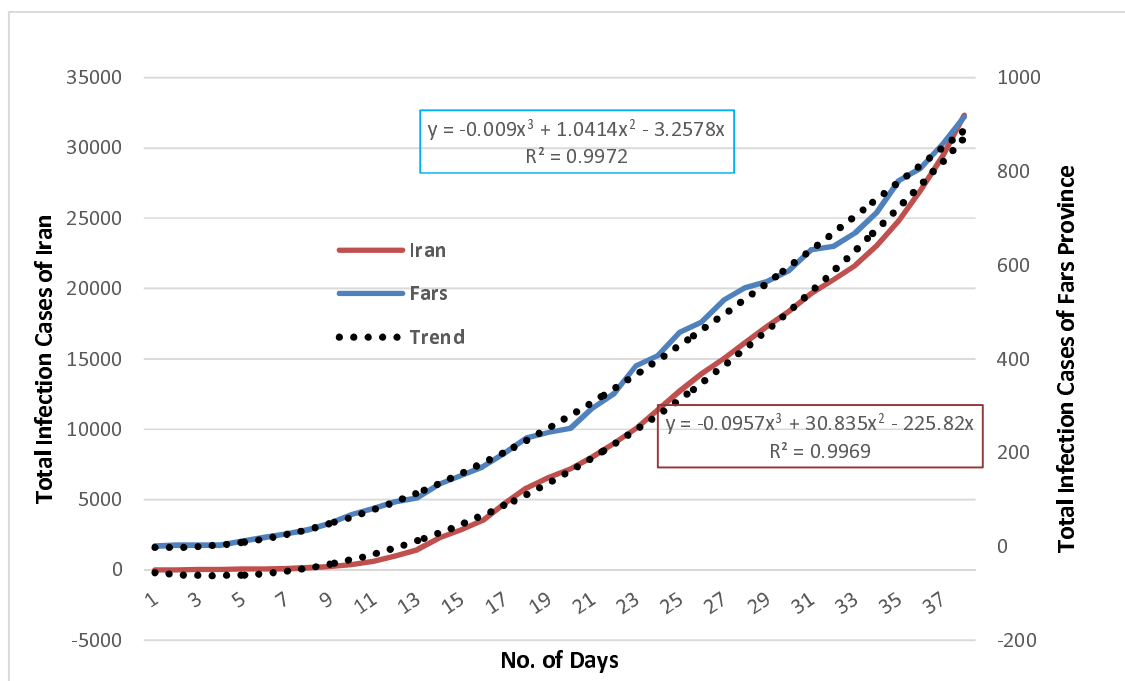


Fig. 12