



# COVID-19 prediction using AI analytics for South Korea

Adwitiya Sinha<sup>1</sup> · Megha Rathi<sup>1</sup>

Received: 13 July 2020 / Accepted: 12 March 2021 / Published online: 8 April 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

The severe spread of the COVID-19 pandemic has created a situation of public health emergency and global awareness. In our research, we analyzed the demographical factors affecting the global pandemic spread along with the features that lead to death due to the infection. Modeling results stipulate that the mortality rate increase as the age increase and it is found that most of the death cases belong to the age group 60–80. Cluster-based analysis of age groups is also conducted to analyze the maximum targeted age-groups. An association between positive COVID-19 cases and deceased cases are also presented, with the impact on male and female death cases due to corona. Additionally, we have also presented an artificial intelligence-based statistical approach to predict the survival chances of corona infected people in South Korea with the analysis of the impact on the exploratory factors, including age-groups, gender, temporal evolution, etc. To analyze the coronavirus cases, we applied machine learning with hyperparameters tuning and deep learning models with an autoencoder-based approach for estimating the influence of the disparate features on the spread of the disease and predict the survival possibilities of the quarantined patients in isolation. The model calibrated in the study is based on positive corona infection cases and presents the analysis over different aspects that proven to be impactful to analyze the temporal trends in the current situation along with the exploration of deceased cases due to coronavirus. Analysis delineates key points in the outbreak spreading, indicating that the models driven by machine intelligence and deep learning can be effective in providing a quantitative view of the epidemical outbreak.

**Keywords** Coronavirus pandemic · Survival prediction · Artificial intelligence · Machine learning · Deep learning · Logistic regression · Autoencoders · COVID-19 · South Korea

## 1 Introduction

The sudden outbreak of coronavirus in the year 2020 had shaken the whole world with both types of COVID-19 carriers, including symptomatic and asymptomatic. This pandemic was also recorded in South Korea that resulted in many coronavirus patients in Seoul's Shincheonji Church [1]. The outburst of the pandemic cases in South Korea was found to be initially centered in Daegu, which is considered the fourth largest city in South Korea. The surge of the deadly virus from Daegu affected the whole country and COVID-19 cases began to increase at several hospitals, mainly in Seoul and the other

major cities. The Korean government took quick action to control the potential spread of coronavirus. In fact, in January 2020, hospitals and medical centers were closed that were found to be harboring patients with the coronavirus. All major public hospitals such as Taejon, Ulsan, and Daegu, were closed and patients were quickly migrated to other care units. The patients who had to be transferred to these hospitals were only admitted in areas that were approved by the healthcare authorities [2]. Only specialized hospitals were designated to treat corona patients. Hence, gradually patients from all over the country were transferred to such locations to restrict the spread [3]. This greatly helped in the restricted outbreak of the pandemic by restricting the movement of corona casualties. This helped the Korean medical authorities to serve other patients who could become targets of the virus in the future. This initiative greatly helped in preparing for possible vulnerable cases. Moreover, the country was parallelly challenged by sudden cases of measles outbreaks among children. The infected children were transported to other special care units to receive instant medical attention. The authorities efficiently sectioned hospitals based on diseases that were likely to have a high co-morbidity rate with corona. These

---

✉ Adwitiya Sinha  
mailto:adwitiya@gmail.com

Megha Rathi  
megharathi.cs@gmail.com

<sup>1</sup> Department of Computer Science & Engineering and Information Technology, Jaypee Institute of Information Technology, Noida, Sector-62, Noida, Uttar Pradesh, India

steps reduced the chances of the virus spread, especially in cities such as Seoul and Daegu, which could otherwise become more acute, and more victims would have to be admitted to hospitals. Therefore, despite the victims being transferred to other areas, the cases of a severe illness caused by the coronavirus in South Korea have been reduced drastically and the outbreak was controlled very quickly. According to statistics, the incidence of child deaths due to the virus was almost eliminated. Moreover, it has been found that the incidences of deaths due to the coronavirus decreased substantially.

Our research focuses on studying the impact of measures taken to control the pandemic and to develop an Artificial Intelligence (AI) approach, and perform survival prediction of corona-isolated patients. The main motivation of our work is to include all possible state-of-art methodologies from machine learning and deep learning domains in our experimentation, to arrive on a conclusive remark regarding model performance with improved accuracy while making the predictions of survival chances of the coronavirus infected patients in South Korea. Our research will help other countries of the world to analyze and plan ways for confronting the present pandemic situation.

## 2 Related work

An epidemic outbreak of COVID-19 in China is spread globally, with more than 3.35 million positive cases worldwide. Research also endorsed that initial symptoms of corona include common flu, fever, and pneumonia. A recent work by the authors [4] provides insight on comorbidities in the SARS-COV-2 (Severe acute respiratory syndrome coronavirus-2) infected individual and the probability of disease in critical patients as compared to non-critical ones. According to the study, experimental results showed the most common feature was fever, then cough, followed by drowsiness and dyspnea. The most frequent disease association were hypertension, diabetes, heart-related issue, and respiratory illness. Medical predictors of critical and non-critical patients are identified in the study [5]. To successfully rank health supplies for critical patients' identification of clinical predictors is required. Real-Time analysis is done using the data collected from Jin Yin Tan and Tongji Hospital on a total of 150 patients. Out of 150, 68 were dead i.e. 45 %, 82 were discharged i.e. 55 %. The criteria for discharge are to check medical conditions for the past 3 to 4 days like no fever, better respiratory functioning, and twice repeated negative reports of the corona. The age, existing diseases, infection in the body, and increased inflaming indicators in the blood are some major predictors of COVID-19. Also, it has been found that COVID-19 death cases might be due to "cytokine storm syndrome" or fulminant myocarditis. From the latest research work [6] it has been discovered that chronic heart disease patients may have a

higher probability of getting infected with the new virus (COVID-19). Also, COVID-19 increases the likelihood of damaging the heart. Comorbidity of cardiovascular disease with COVID-19 is presented in the study. Individuals with existing heart disease are found to be in a severe risk zone for getting infected with COVID-19 [7]. The probability of death is also high due to COVID-19 with exiting cardiovascular disease. Another significant study discovers the relationship between an individual with heart injury and death ratio in individuals with COVID-19 [8]. Data for the analysis was taken from one of the hospitals in Wuhan, China, and the response of patients was analyzed with and without heart injury. It has been concluded from the study that heart injury is a general symptom among hospitalized corona positive patients and is related to higher mortality risk. To evaluate the linkage between existing heart disease and its impact on COVID-19 is the main emphasis of this research work [9]. Myocardial injury is notably linked with an incurable result of COVID-19. Improper functioning of cardiac is just because of myocardial injury. COVID-19 is the latest virus oriented communicable disease that has a notable connection with cardiac disease [10]. COVID-19 has negatively impacted the heart of the patient and is found to be responsible for myocarditis, venous thromboembolism, myocardial injury, and arrhythmias. In yet another novel work [11] impact of platelet count is outlined on the severity of the disease. COVID-19 is the latest virus with a lack of treatment therapies and vaccines to handle it. Critical and non-critical patients can be discriminated against based on platelet count. Survivability can be predicted based on platelet count i.e. patient with a high platelet count has a higher survivability rate. Additionally, it has been investigated from this study that thrombocytopenia is also responsible for the severity of the disease. Low platelet count may increase the death risk of COVID-19 patients. The research presents Chest CT as the primary testing approach for the detection of the COVID-19 virus [12]. Results showed that the correct diagnosis rate from chest CT was 88 % for the suspected COVID-19 patients. In the retrospective research work, chest CTs of 121 patients were analyzed for medical investigation on COVID-19 [13]. Association to time between symptom and early-stage CT scan were examined. Remarkable information concluded on imaging related to the virus were circumferential and two-sided ground-glass and amalgamated pulmonary opacities. The main emphasis of another contribution of image processing techniques for COVID-19 is to define CT results across different points of time [14] all over the disease time. The combined effort of image processing along with medical approaches could ease the early detection of pneumonia due to COVID-19. Applied informatics and advanced computational techniques are very significant in diagnosing and treating COVID-19 patients. In the recent study [15] deep learning models were applied for the classification of COVID-19 CT images. The study analyzed multiple CNN (convolutional

neural networks) to group images into two categories: 1) No-infection, 2) Infection with viral pneumonia. Results are very effective which achieves an AUC (Area under the curve) of 0.996, the specificity of 92.2 %, and sensitivity of 98.2 %. Authors in the study [16] used CNN for the diagnosis of COVID-19 positive cases from chest CT images. A crucial step in the fight against the virus is the early detection of positive cases so that one can isolate them to stop the further spread in the community. Motivated by this, the authors used CT images of individuals and then applied deep learning techniques for the early screening of positive cases. COVID-Net a deep convolutional neural network is developed for the diagnosis of COVID-19 patients from chest CT images. A deep learning framework is developed to accurately predict COVID-19 patients in another novel work [17]. Using CT images AUC value obtained from the model was 0.96. Medical reports showed that most of the corona patients suffered from a lung infection. Medical imaging techniques are proven to be very effective for diagnosis due to its cost-effectiveness and less time for imaging [18]. One of the core sub-domains of artificial intelligence known as deep learning is widely used by the radiologist for further investigation and accurate diagnosis. Authors in this work developed a more reliable, fast COVID-19 detection model based on deep anomaly detection techniques. A dataset of chest CTs from the Github data repository was used and accuracy achieved was recorded 96 %, specificity as 70.65 % and sensitivity was 70.65 %. Summarized in Table 1 is the contribution of smart techniques for COVID-19 prognosis and diagnosis.

A recent study presents a deep learning segmentation model for the identification of infectious area and their volume of impact [29]. 91.6 % agreement achieved by the target developed model and the mean estimation error was 0.3 %. Notable research work was done by the authors for the quick recognition of target drug treatment for COVID-19 patients [16].

Advanced computational techniques were applied for the fast recognition of potent drugs after the specific 3D-composition of chief virus proteins are resolved. A remarkable contribution in the field of discovering drugs for COVID-19 is presented in the study [19]. Current research work targets to experiment with anti-HCV drugs for fighting with COVID-19. Modeling, docking, and sequence analysis are used to develop a model for COVID-19 RdRp (RNA dependent RNA polymerase). Key findings from the study suggest that IDX-184, Remidisvir, Sofosbuvir, and Ribavirin are powerful drugs for COVID-19.

From the study [30] it has been found that favipiravir, arbidol, chloroquine, and remdesivir are under testing to prove safety and potency for treating COVID-19 patients. In the latest study pinpoint to use of chloroquine phosphate an old medicine to treat malaria is also effective against COVID-19 [20]. In the study [31], the authors have applied a sparse encoder based artificial neural network (ANN) to predict heart disease with better performance and stability with faster convergence. Another latest research in COVID-19 used a set of modified autoencoders to perform time-series prediction of daily confirmed cases in Brazil [32]. In our research, we have attempted to incorporate autoencoders, along with machine learning models, for predicting the survival chances of isolated COVID-19 patients in South Korea.

Granular computing is another widely utilized technique. In the paper [33] for enhancing the prediction accuracy fuzzy information granulation (FIG) is merged with a deep neural network for accurate traffic prediction. The fuzzy approach is used in another recent work [34] for generating more accurate results. The main emphasis is to develop a novel fuzzy approach using ridge regression for time series prediction. In yet another novel work [35] a new approach Fuzzy Learning Vector Quantization is developed which is the amalgamation of Learning Vector Quantization Neural Network and Fuzzy

**Table 1** Drugs for treating COVID-19

S.No.	Authors	Target Drug/Therapy
1.	Elfiky 2020 [19]	IDX-184, Remidisvir, Sofosbuvir, and Ribavirin
2.	Gao et al. 2020 [20]	Chloroquine phosphate
3.	Touret & de Lamballerie 2020 [21]	Chloroquine
4.	Cortegiani et al. 2020 [22]	Chloroquine
5.	Kalil 2020 [23]	Favipiravir, chloroquine, brincidofovir, hydroxychloroquine, monoclonal antibodies, antisense RNA, and convalescent plasma
6.	Gautret et al. 2020 [24]	Chloroquine and hydroxychloroquine
7.	Baron et al. 2020 [25]	Teicoplanin
8.	Colson et al. 2020 [26]	Chloroquine and hydroxychloroquine
9.	Cai et al. 2020 [27]	Favipiravir (Antiviral therapy)
10.	Wu et al. 2020 [28]	TH17 responses in patients with SARS-CoV-2 and JAK2 inhibitor Fedratinib for reducing mortality of patients with TH17 type immune profiles

system. The hybridization proved to be beneficial in terms of enhanced accuracy and reduced training time. Limitations of the Prism technique along with its variant is presented in this study [36]. Results were also conducted to verify and validate the performance of an improved version of Prism and it shows that complexity is reduced with improvement in accuracy which is highly desirable in dealing with real-world data. Another novel hybrid approach is developed using nature-inspired computing for ameliorating the overall prediction accuracy in the setting of granular computing [37]. In the research work [38] authors developed an approach that includes three steps. The first phase is comprised of learning quasi-identifiers, the second step is comprised of granulation calculation and the last step includes masking. The approach implemented used k-anonymity to de-identify private information systems. The research work [39] is an amendment in the traditional fuzzy inference system wherein overall computational model performs type-1, interval type-2, and general type-2 fuzzy computation by adopting genetic algorithm as an optimization technique. A contemporary intuitionistic fuzzy time series model is created for prediction purpose [40]. In the proposed model, the intuitionistic fuzzy c-means technique is used for fuzzification and pi-sigma neural network is utilized for defining fuzzy relations.

### 3 Methods & model description

The prediction of survival chances of coronavirus infected population is conducted in two phases, involving the model analysis and model prediction using machine learning and deep neural networks. The machine learning techniques include Logistic Regression and Support Vector Machine (SVM) for classifying recovered and deceased cases based on the initial set of  $n$  input features denoted as,  $X = \{x_1, x_2, \dots, x_n\}$ , where  $x_i \in \mathbb{R}^n$ . The logistic estimation function is sigmoidal and is expressed as follows:

$$f_{\theta} = \frac{1}{1 + e^{-h(X, \mathbb{C})}} \quad (1)$$

Here,  $\mathbb{C} = (c_0, c_1, c_2, \dots, c_n)$  refers to the set of coefficients used in the sigmoidal function,  $h(\cdot)$  which is expressed as:

$$h(X, \mathbb{C}) = c_0 + c_1x_1 + c_2x_2 + \dots + c_nx_n \quad (2)$$

The likelihood function ( $\Theta$ ) in logistic regression is the objective function to maximize the degree of discreteness in the process of classifying the individual samples having binary output values. This is computed using Eq. (3).

$$\Theta = \underbrace{\arg \max}_{\mathbb{C}} : \log \left\{ \prod_{i=1}^n f_{\theta}^{y_i} (1 - f_{\theta})^{(1-y_i)} \right\} \quad (3)$$

Here,  $y_i \in Y$  refers to the output variable to be predicted and is denoted by,  $Y = \{y_1, y_2, \dots, y_n\}$ . To compare the performance of logistic regression, another machine learning algorithm, SVM is employed for estimating the survival chances of corona infected patients. For the given sample population of size  $n$ , and the training set  $T = \{(x_i, y_i) | x_i \in X, y_i \in Y, i = 1, \dots, n\}$ , the SVM classifier is computed as:

$$\mathcal{P}_{W,B}(X) = Y(W^T \cdot X + B) \quad (4)$$

In the above Eq. (4),  $W = \{w_1, w_2, \dots, w_n\}$  are the weights, and  $B = \{b_1, b_2, \dots, b_n\}$  denotes the intercept. Also,  $Y = \sum_{i=1}^n x_i w_i + b_i$  is the dependent term which is predicted as a function of the independent set of variables,  $X$ . The deep learning model  $X$  having  $l$  layers is considered to be the tensor of information, which is being processed through dense interactions of  $l - 2$  hidden layers for computing the final estimate of the output variable. In our model, six hidden layers with 30 neurons were used for aligning the training loss with validation loss during the model building phase. Also, the Rectified Linear Unit (ReLU) activation with function  $\max(0, h(X, W))$  is adopted to deal with the issue of vanishing gradient in the deep neural network model. The deep learning approach has been further extended to develop an autoencoder neural network. Being an unsupervised learning algorithm, autoencoders are well equipped to apply backpropagation so that the predicted trend of the target variable aligns with the inherent pattern in the input set. The process of building an autoencoder is two-folded, which involves the construction of an encoder and a reconstruction of the input using a decoder. The function for encoder ( $\mathcal{E}$ ) and decoder ( $\mathcal{D}$ ) is expressed as follows:

$$\mathcal{E} = f_{\mathcal{E}}(\mathcal{X}, \mathcal{W}_{\mathcal{E}}, \mathcal{B}_{\mathcal{E}}) = f_{\mathcal{E}}(\mathcal{W}_{\mathcal{E}}\mathcal{X} + \mathcal{B}_{\mathcal{E}}) \quad (5)$$

$$\mathcal{D} = f_{\mathcal{D}}(\mathcal{E}) = f_{\mathcal{D}}(f_{\mathcal{E}}(\mathcal{X})) \quad (6)$$

$$\mathcal{D} = f_{\mathcal{D}}(f_{\mathcal{E}}(\mathcal{W}_{\mathcal{E}}\mathcal{X} + \mathcal{B}_{\mathcal{E}})) = f_{\mathcal{D}}(\mathcal{W}_{\mathcal{D}}^T \mathcal{X} + \mathcal{B}_{\mathcal{D}}) \quad (7)$$

The outcome of the target parameter from the autoencoder can be analyzed using t-Distributed Stochastic Neighbor Embedding (t-SNE). It is a non-linear probabilistic technique that is highly capable of interpreting the complex polynomial relationship among the  $n$  input features. Though, computationally expensive, t-SNE offers an efficient probabilistic solution to map multi-dimensional input features to a lower-dimensional space to reveal patterns in the predicted outcome.

T-SNE uses the symmetric version of SNE that initiates with finding distances between the pair  $x_i$  and  $x_j$ , and

converting them to conditional probabilities.  $x_i$  will choose  $x_j$  as a neighbor if their respective conditional probability  $p(j|i)$  lies within the threshold distance in the Gaussian distribution centered at  $x_i$ . In this case, the conditional probability is computed as follows:

$$p(j|i) = \frac{e^{-\|x_i-x_j\|^2/2\sigma_i^2}}{\sum_{i \neq k} (e^{-\|x_i-x_k\|^2/2\sigma_i^2})} \tag{8}$$

This probabilistic measure is compared with the low-dimensional counterpart  $q(j|i)$ . Finally, the t-SNE plots attempt to visualize the outcome in a manner that minimizes the sum of the difference in conditional probabilities  $p(j|i)$  and  $q(j|i)$ , using the symmetric version of the Kullback-Leibler divergence measure.

### 4 Model analysis

The corona disease dataset is taken from Kaggle, which contains information about the health data of South Korea from December 2019 to March 2020 [Kaggle, 2020].

The dataset from *South Korea* was taken for experimentation of COVID-19 analysis, for the following reasons:

- a. *Availability*: Our research was conducted in the early stage of COVID-19 spread when the availability of data related to coronavirus was largely restricted in content as well as quality. However, the data from South Korea on Kaggle was found to be in a rich format.
- b. *Holistic*: The South Korea COVID-19 dataset was most exhaustively framed, spanning over maximum possible influencing factors, including the gender-wise, web-search, age-related, and region-based data. Moreover, it also included the three major categories of patients,

namely confirmed cases, deaths, and isolated (quarantined cases). This helped us in training & testing our model with confirmed and death cases; and real-world prediction of the quarantined cases.

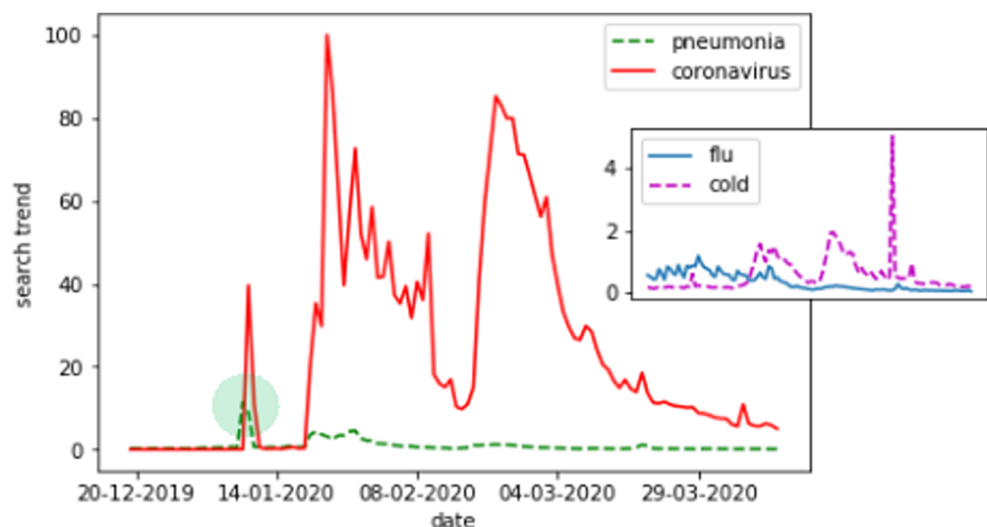
- c. *Authenticity*: The dataset was claimed to have real and verified instances from the Korea Centers for Disease Control & Prevention (KCDC). The structured dataset was validated against the reports received from local governments of South Korea.

Our study spans over conducting age-wise and gender-wise analysis, the temporal evolution of COVID-19, as well as building a machine learning model for prediction of survival chances of isolated patients. Our initial study depicts the percentage of internet search made by Koreans concerning different diseases, during the initial phase of corona spread.

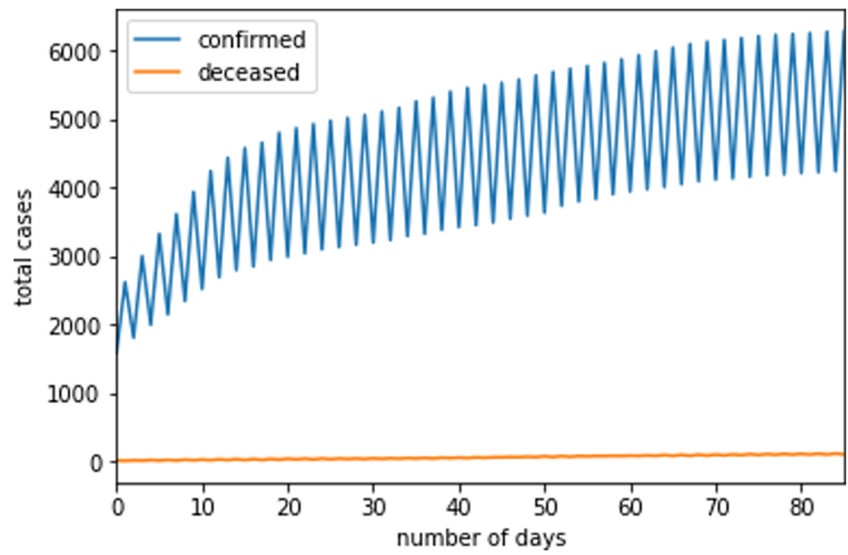
Figure 1 clearly shows that during the inception of COVID in December 2019, people were less aware of the pandemic, but experience breathing-related problems. Therefore, the internet search trend shows that Korean people widely searched on Pneumonia at the beginning, which was later followed by a wide number of searches on coronavirus. The figure also depicts the search trend on other related diseases, including the flu and cold during the same period. This shows that pneumonia, flu, and cold are comorbidities in common with coronavirus. In another study made in Fig. 2, the growth rate of the total confirmed and deceased cases could be seen increasing with an increase in the number of days. Figures 3 and 4 show the distribution of numerical data of age groups for confirmed and deceased corona cases concerning their quartiles. It is apparent from Fig. 3 that people lying in the age group 20 were the most contaminated with the deadly virus, thereby becoming active careers of the disease.

However, the target for COVID was those people who were in the 70 s and 80 s age group (Fig. 4). Figures 6a-e and 7a-e show a recurring funnel plot for capturing the

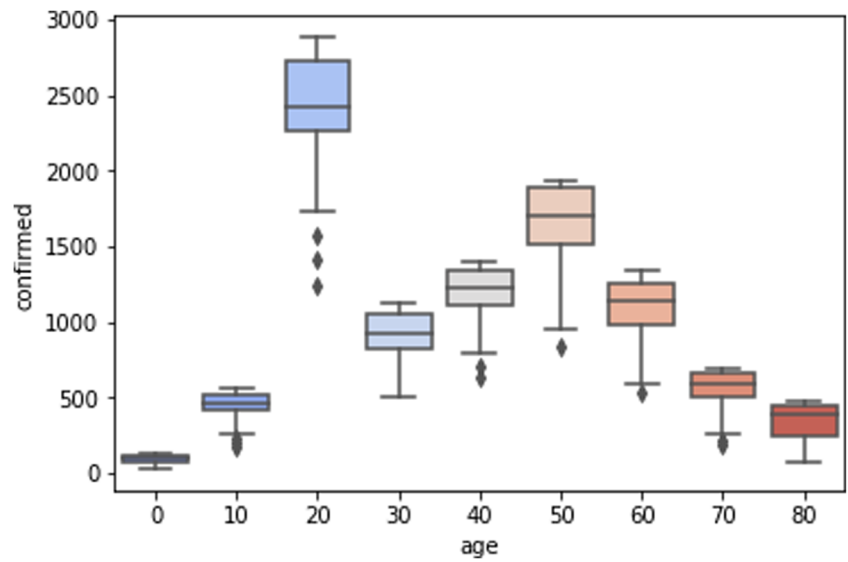
Fig. 1 Search trend analysis of related diseases in Korea



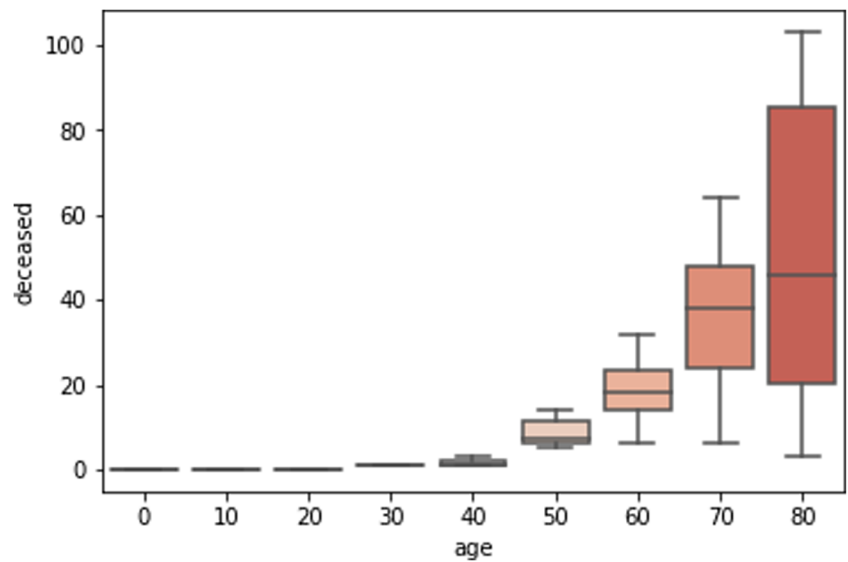
**Fig. 2** Analysis of confirmed & death cases of Corona outbreak



**Fig. 3** Analysis of confirmed cases over different age groups



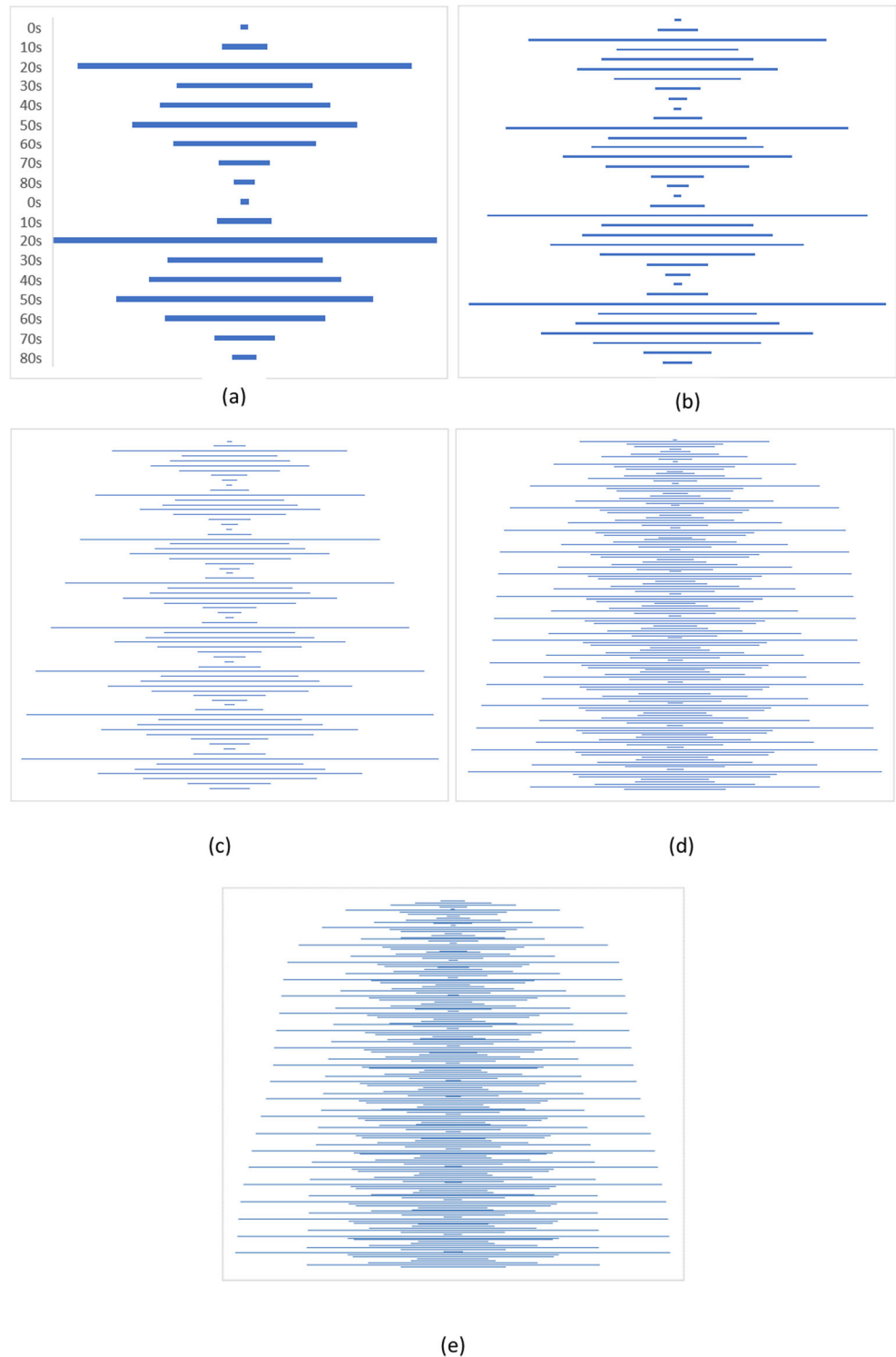
**Fig. 4** Analysis of deceased cases over different age groups



temporal evolution of the confirmed and deceased cases of corona outbreak for 2, 4, 8, and 16 days grouped, to study the inherent evolution pattern. The actual number of cases is represented by the length of the line in the funnel chart. Through Fig. 5a-d, it is apparent that the pattern of disease evolves into a wide funnel with multiple layers. Each layer

corresponds to the number of COVID cases per age group. The distinctly multi-layered funnel plot interprets the fact that the deadly coronavirus initially attacked Korean people belonging to all age groups, especially between the 20 s to 50 s, as shown in Fig. 5e. The temporal evolution over time shows an exponential pattern in the disease spread. However,

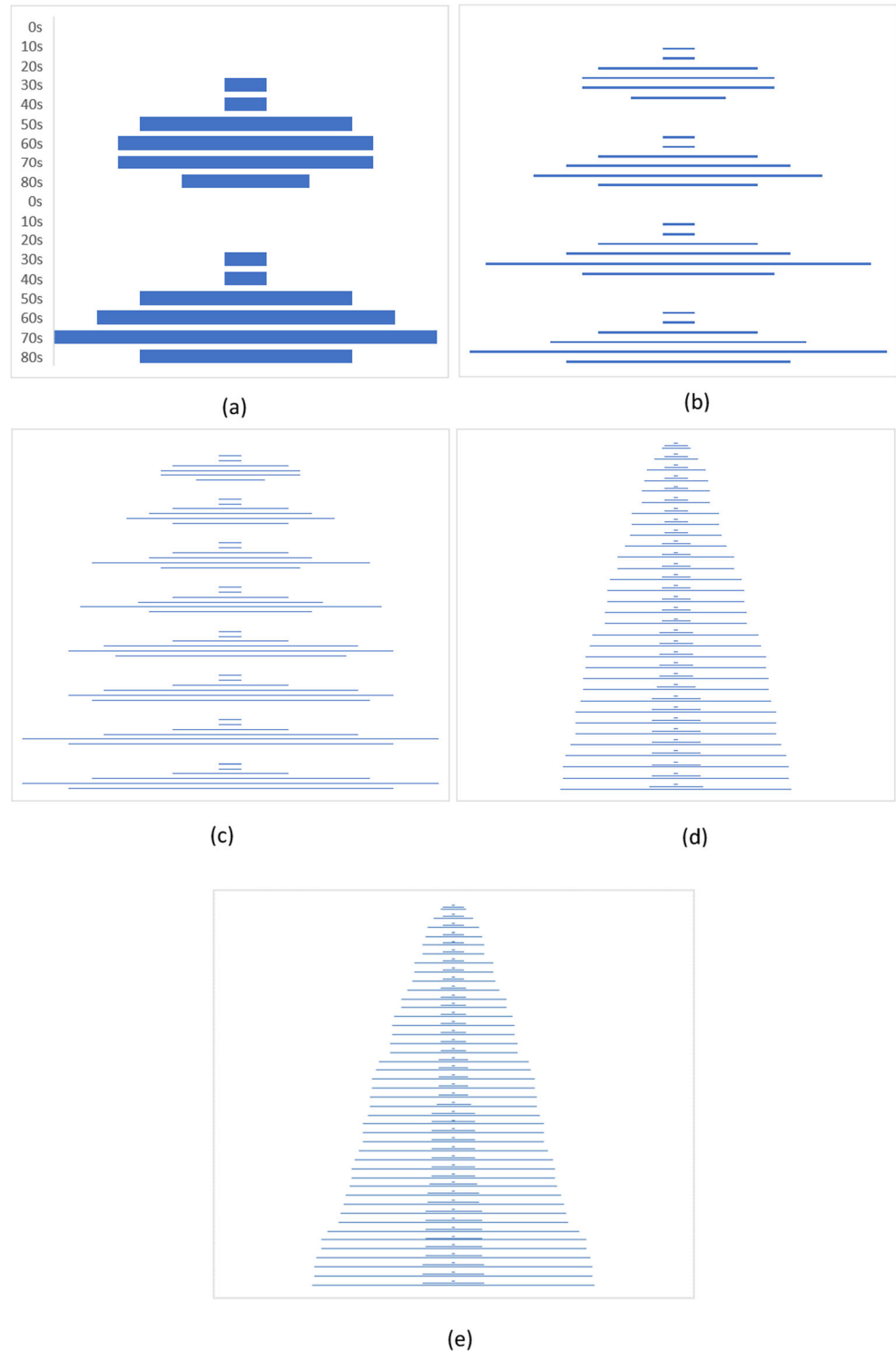
**Fig. 5** Temporal evolution of confirmed cases for (a) 2-days (b) 4-days (c) 8-days (d) 16-days (e) Overall



the deceased plots in Fig. 6a-d show a rather narrow funnel with only two distinct layers, corresponding to the 70 s and 80 s age-groups. This implies that the older age-group became the target and deceased due to the viral infection (Fig. 6e). Unlike Fig. 4, where the 80 s age-group seems to be the most targeted; Fig. 6 highlights the 80 s age-group with a line of

relatively smaller length, especially as compared to the people in the 70 s age-group. This happened due to the reason that the population in South Korea belonging to the 80 s and above is very less as compared to the total population. Hence, even if most of the people in the 80 s eventually become the target for corona; but owing to its lesser population, the funnel

**Fig. 6** Temporal evolution of deceased cases for (a) 2-days (b) 4-days (c) 8-days (d) 16-days (e) Overall





representation in Fig. 7 highlights the 80 s group with relatively lesser deaths.

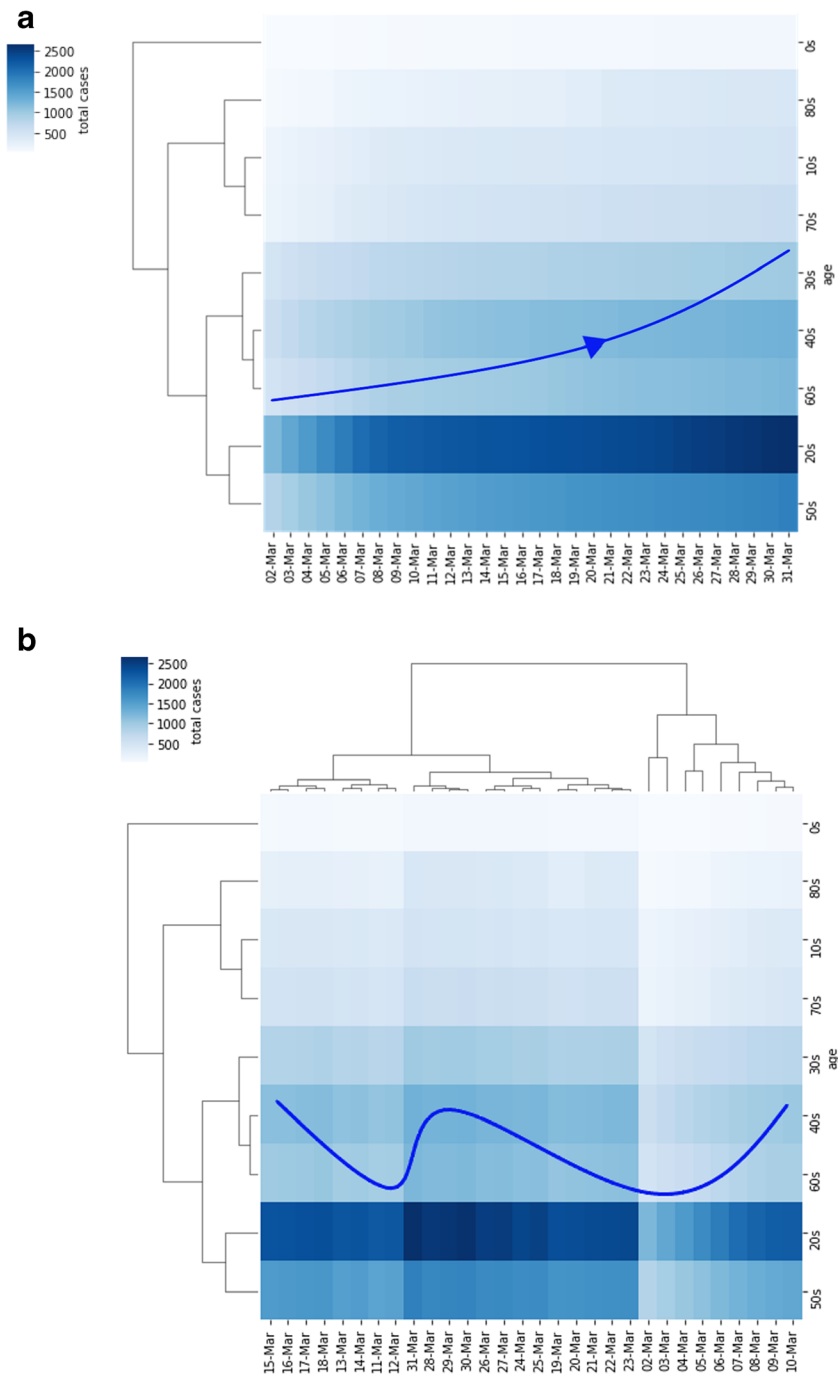
Figure 7a shows a hierarchically clustered map of detected corona patients over the age-group. It can be seen that there is a rise in all the age groups of corona cases detected as confirmed. Further, Fig. 7b shows the cluster map of the age-group based on the detected cases relative to the clustered duration of corona spread. The joint clustering over age group and time is performed using the Bary-Curtis metric, also known as Sorensen distance. As compared to Euclidean, Minkowski, Jaccard, and

Chebyshev distance metrics, the cluster representation with the Sorensen metric was found to be most appropriate. There are three distinct phases apparent from the diagram that shows rise and fall in the detected cases in South Korea in March 2020.

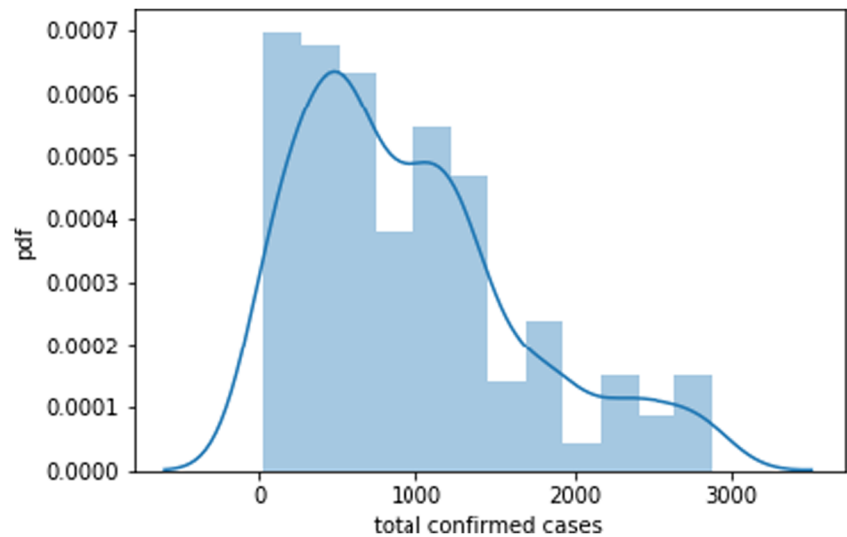
### 5 Statistical analysis of pandemic outbreak

This section reveals the statistical relationship between the deceased and confirmed cases of the corona. In Figs. 8 and

**Fig. 7** a Cluster map of detected corona cases congregated over of age-groups. b Map of age-groups clustered over age and time (in days)



**Fig. 8** Density distribution of confirmed detected cases of COVID-19



9, the density distribution is illustrated for the patients detected as confirmed and deceased in the course of treatment, using Kernel Density Estimate (KDE). In the case of confirmed corona patients in Fig. 8, a wider area under the KDE plot validates the explosive outbreak of corona. However, the density distribution of the deceased cases shows a single peak at the initial phase, which depicts that deaths were more frequent at the beginning of the COVID outbreak (Fig. 9). Moreover, since there is a single in the deceased case, it also explains the reason behind the quick recovery of the Koreans.

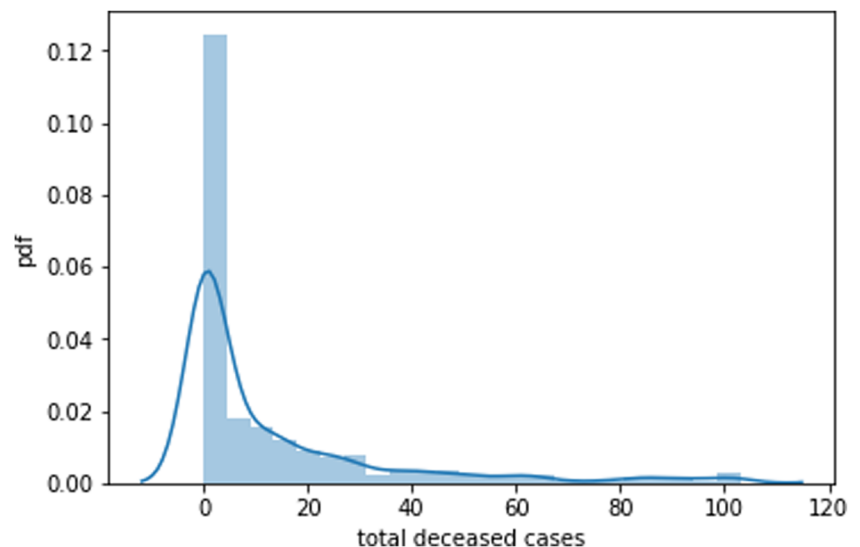
The gender-based analysis as shown in Fig. 10, shows that the trendline of deceased male patients increases rapidly as compared to female patients in Korea. Also, the sparsity of dots over the deceased trend of women portrays the fact that females remained more immune to the coronavirus than Korean males. Figure 11 highlights province-wide total

corona cases in South Korea from December 2019 to April 2020. The study identifies Daegu as the corona hotspot that was responsible for the outbreak of the pandemic in Korea.

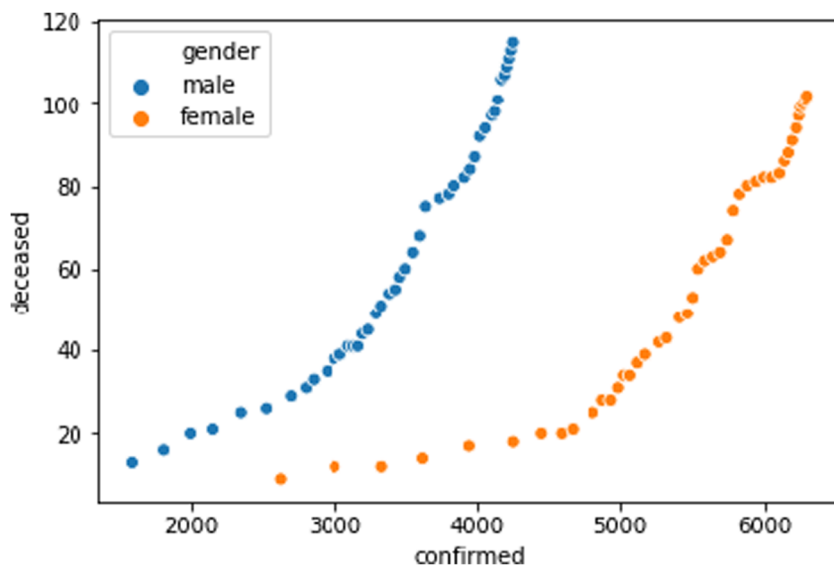
## 6 AI-based model prediction

Machine learning and deep learning approaches are used to estimate the expected chances of survival for 1533 isolated patients in South Korea, based on age, gender, number of days of treatment, and number of previous ailments. The training and testing of the Artificial Intelligence (AI) based models was conducted on 5165 coronavirus instances and was validated over 1533 quarantined patients. The proposed AI-based approach is highlighted in Fig. 12. The model is trained with 70 % data from the total instances, involving confirmed as

**Fig. 9** Density distribution of death cases of COVID-19



**Fig. 10** Gender-based analysis of confirmed cases detected over corona deaths



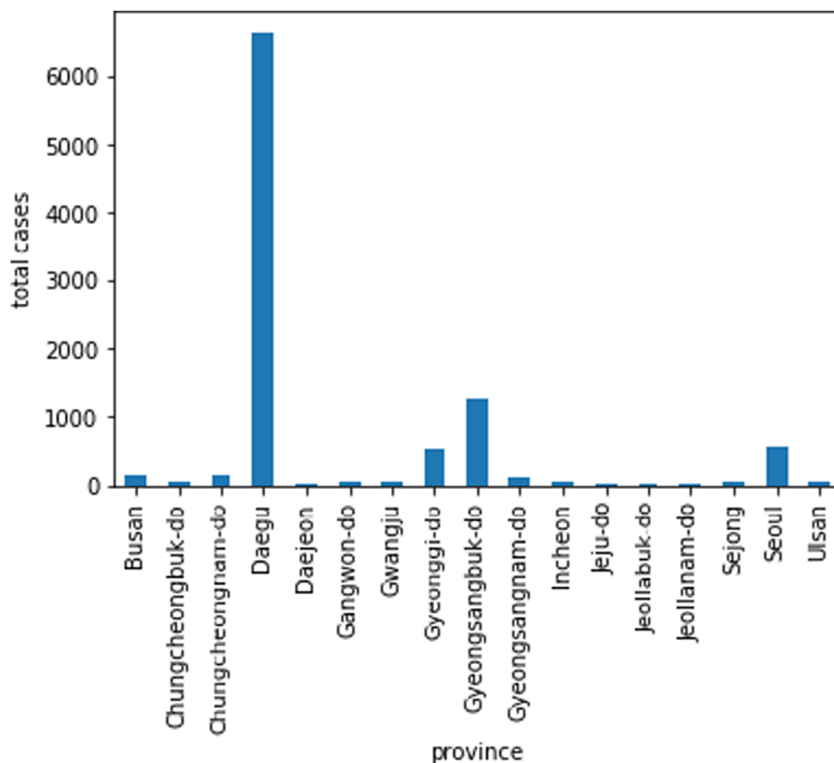
well as deceased cases; and the remaining 30 % of the cases were used for model testing.

**6.1 Hyperparameters tuning**

In the process of building an efficient leaning model, hyperparameters tuning is considered as one of the most important aspects. It refers to the pre-setting of some of the higher-level parameters, before model training, according to

the dataset characteristics and capacity of the AI-based approach. Some of the hyperparameters used for training our model, include *C* -regularization, gamma ( $\gamma$ ), kernel, and degree parameter. The value of *C* trades-off between the width of the hyperplane margin and model learning accuracy. An unusual decrease in *C* value may reduce the training accuracy, hence set to 1.0 in our case. The  $\gamma$  parameter defines the influence of the training dataset. A higher value makes the model completely biased over the training points; while a

**Fig. 11** Province-based analysis of corona infected cases in Korea



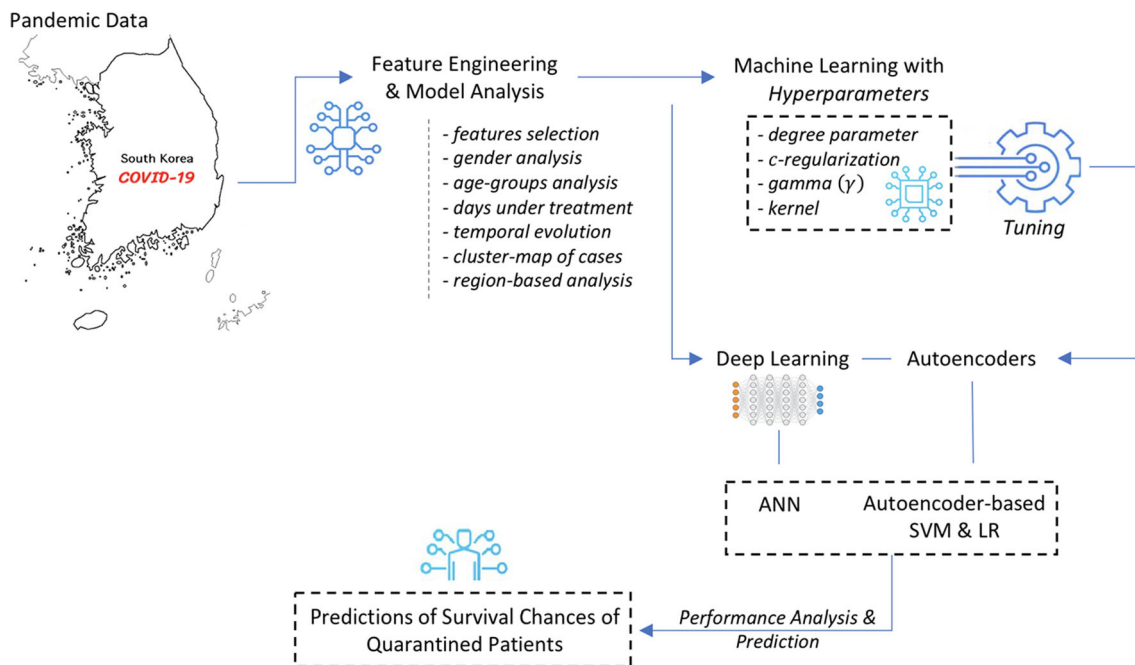


Fig. 12 A working prototype of AI-based approaches with model parameters

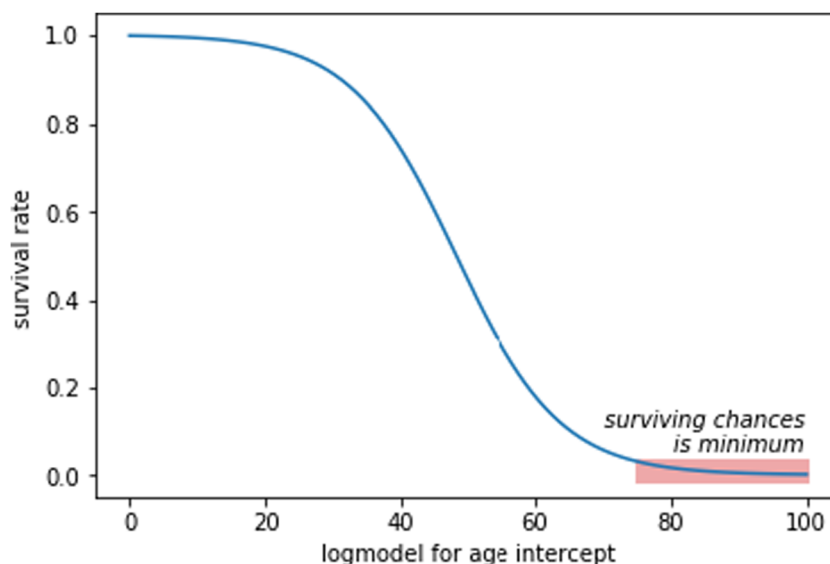
lower value may ignore the complexity of the dataset. For both,  $C$  and  $\gamma$ , the permissible value may range between  $10^{-3}$  to  $10^3$ . In our case, automated tuning of the gamma parameter was performed and the optimal value was found to be  $\gamma = 0.01$ . Setting the kernel is yet another important task, as it reduces the feature space mapping complexity of the machine learning model. The radial basis function was found to be the best option and hence was used as the kernel. The degree hyperparameter with a higher degree would result in a more flexible decision boundary. In our experimentation, this parameter gave optimized results when set to 3. In general, a

degree of 1 result in a linear kernel having lesser learning accuracy.

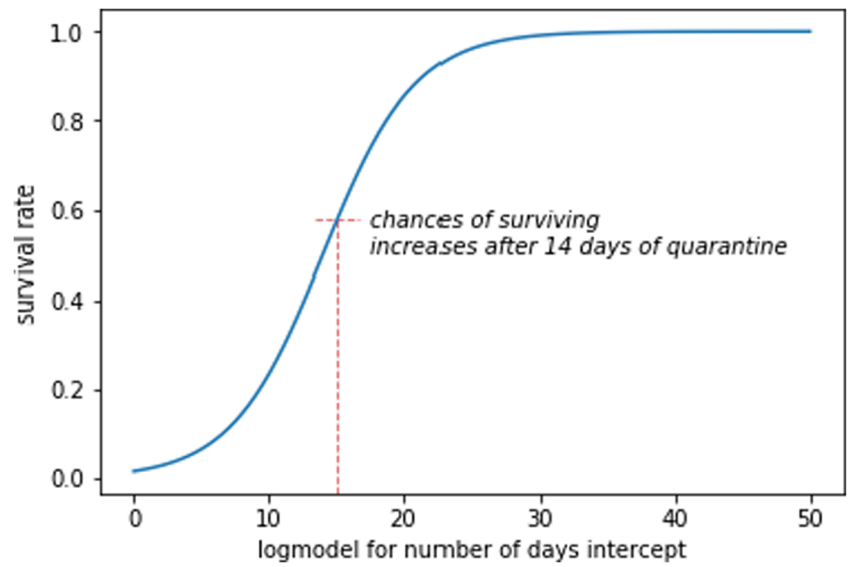
### 6.2 Machine learning model for COVID-19

Initially, machine learning models, including logistic regression and Support Vector Machine (SVM) were applied to build a model for predicting survival chances. Logistic Regression (LR) is employed to perform the prediction with 4000 iterations. The LR model gave an average macro accuracy of 91%. Macro-average accuracy can be defined as the

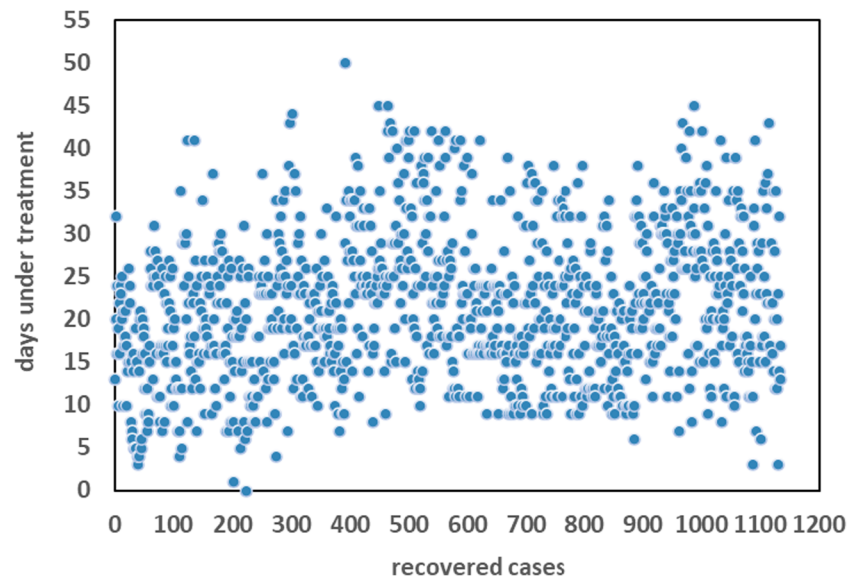
Fig. 13 Log-transformed intercept of patient age



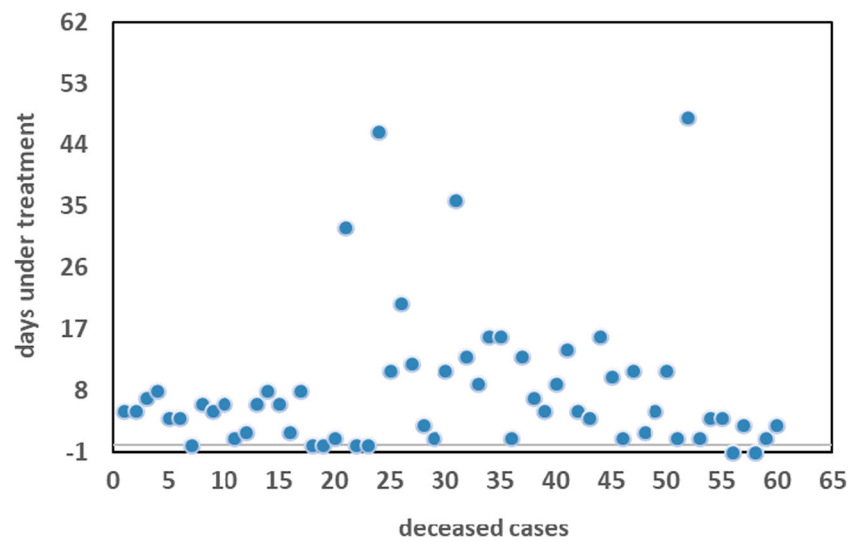
**Fig. 14** Log-transformed intercept of days under coronavirus treatment



**Fig. 15** Number of days under treatment for recovered cases



**Fig. 16** Number of days under treatment for deceased cases

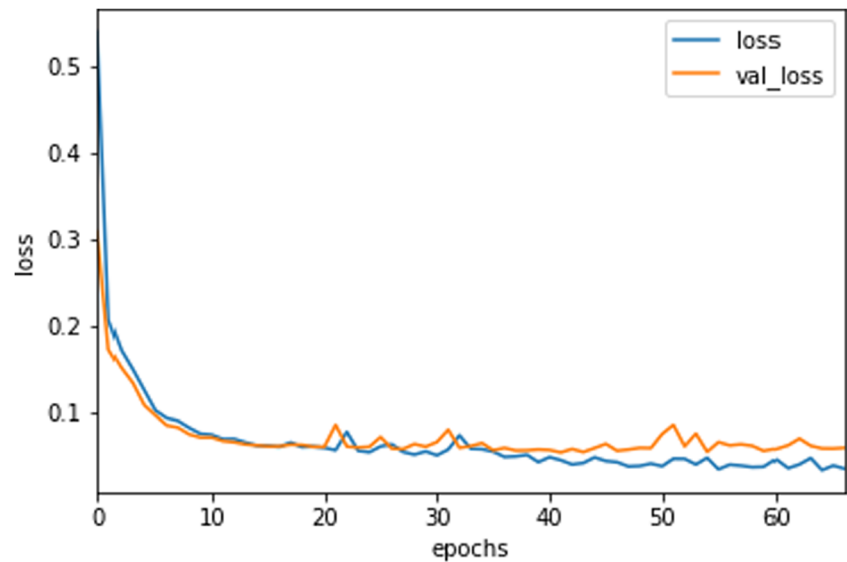


average of model precision and recall. The macro-average treats all the classes of data equally by computing the model accuracy distinctly for each class. Further, SVM is applied with parameter optimization using the grid search cross-validation technique, and the accuracy was improved to 97 %.

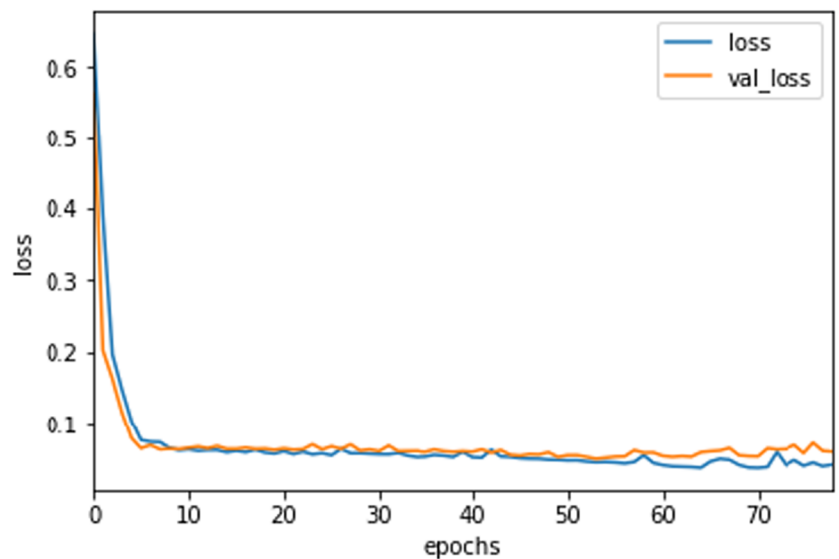
Figures 13 and 14 shows the log model representation of age and number of days under treatment for COVID-19 patients. In Fig. 13, the log-transformed intercept of age experimentally reveals that in general older patients have a lesser survival rate. From Fig. 14, it is evident that 14 days of medical care can eventually lead to an increase in the survival rate.

This interpretation is directly related to the experimental outcome in Figs. 15 and 16, which shows the impact of days of treatment taken, over the survival chances of people in South Korea. It is evident from Fig. 15 that, on average the patients who underwent hospitalization for more than 14 days, gradually developed immunity to fight against the effects of coronavirus. Moreover, Fig. 16 supports the revelation that the chances of death were fairly high on the first 14 days of treatment. This also explains the reason behind the standard recommendation of 2 weeks of home or hospital quarantine (depending upon the criticality of health condition), before any verdict could be given for the survival prospects of a patient.

**Fig. 17** ANN model validation & training loss with (a) *Adam* optimizer in 70 epochs, and (b) *Adagrad* optimizer in 80 epochs while deep learning simulation



(a)



(b)

### 6.3 Deep learning prediction model

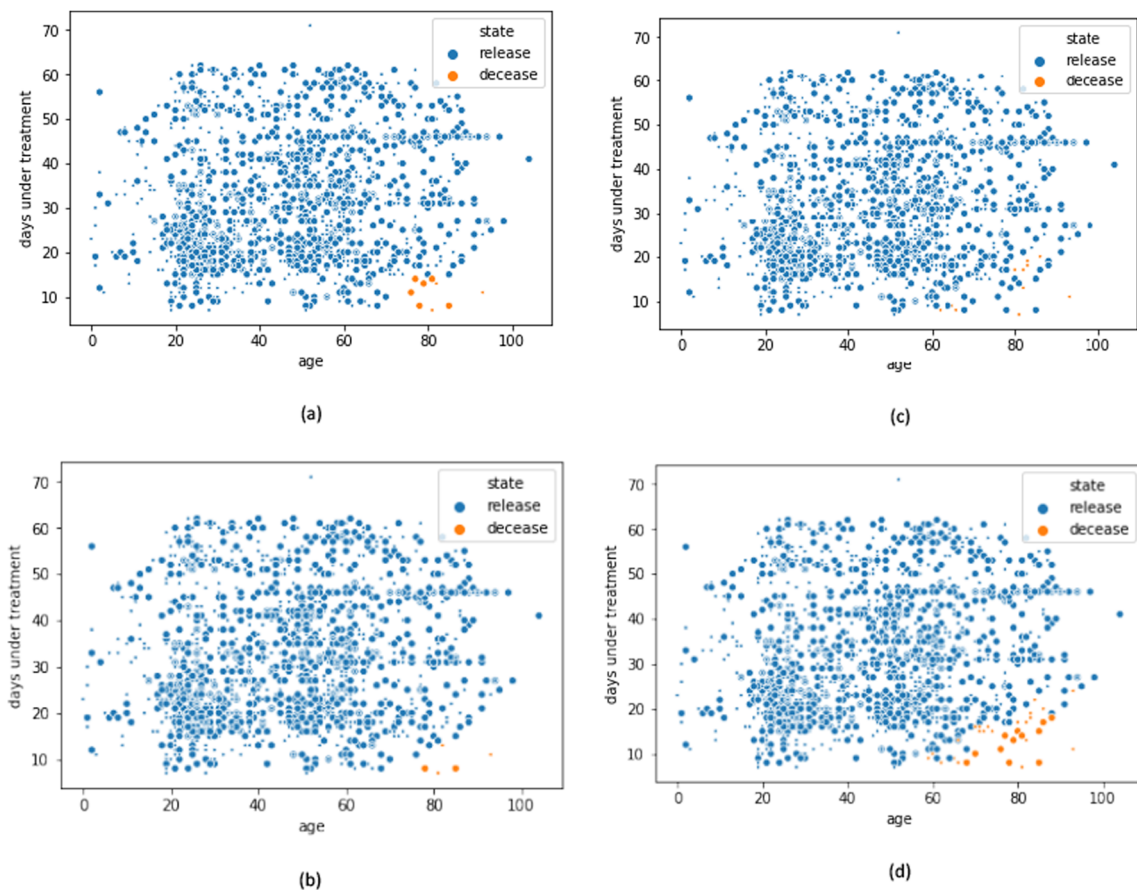
In the second phase, a deep learning approach is used to develop an Artificial Neural Network (ANN) for identifying the fraction of the population who bear higher chances of survival based on the average length of their stay in the hospital, age group, gender, region of belongingness, number of contacts, and comorbidities. The result in Fig. 17a-b monitors the training and validation loss during the deep learning simulation with *Adam* and *Adagrad* optimizers, respectively.

Both the results show a desired alignment of validation with training loss, depicting the validity of the ANN models on the dataset. Binary cross-entropy with *Adam* optimizer is employed to evaluate the validation performance (Fig. 17a). Moreover, early stopping criteria are applied to the simulation to restrict the over-fitting of the model due to a large number of epochs while training. The number of epochs in the case of the *Adam* optimizer was recorded to be 70. The accuracy recorded from our ANN model is 99 %. In Fig. 17b, though the *Adagrad* optimizer gave better alignment on the losses, but required 80 epochs before early stopping criteria could be

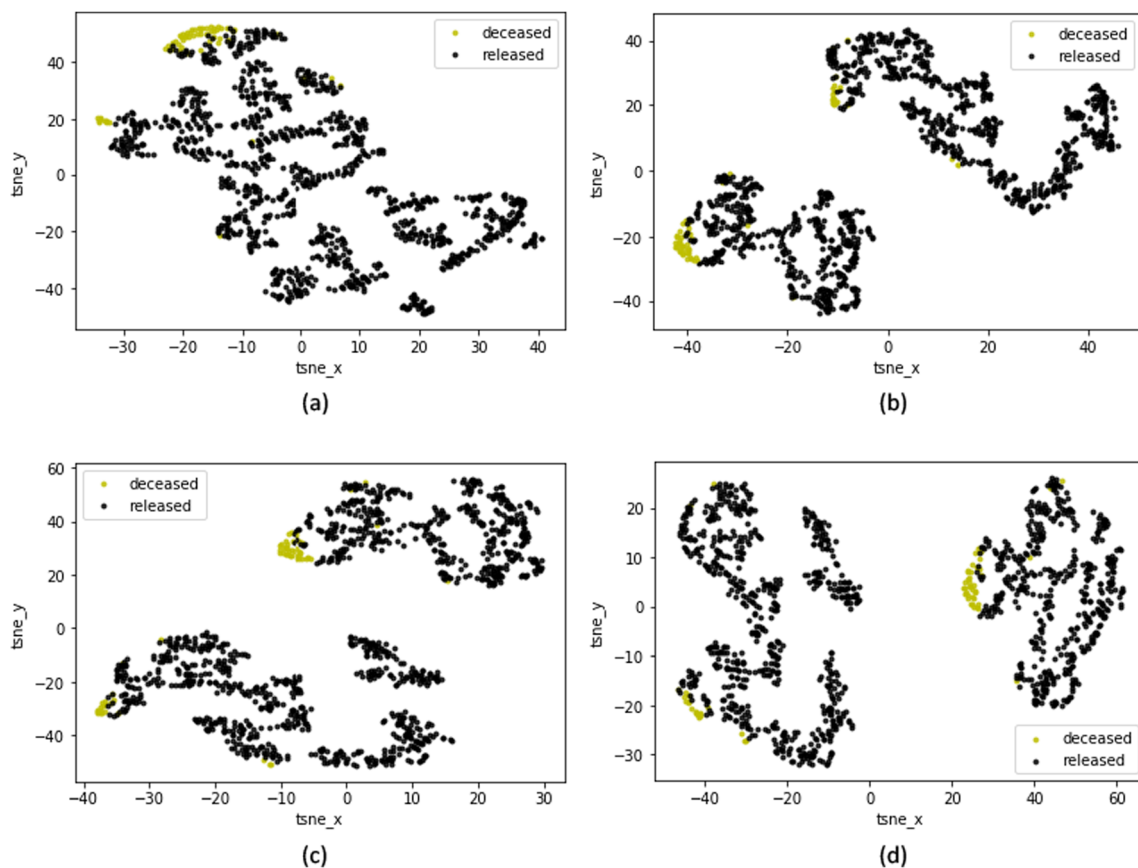
satisfied. Also, there was no improvement in the model accuracy and was fairly found to be the same, i.e. 99 %, as in the case of ANN with *Adam* optimizer. Furthermore, the Rectified Linear Unit (ReLU) activation function was used for the 6 denser intermediate layers having 30 neurons. The sigmoid activation function was used for the last layer, as our model had a binary outcome in the form of released or deceased state. Our trained and tested model was further applied for the survival prediction of the isolated patients in South Korea, and the result is highlighted in Fig. 18a.

This is followed by the outcome of applying SVM to isolated cases in Fig. 18b. The model simulation performed with SVM gave an accuracy of 97 %. The male and female corona patients are represented with dots and filled-in circles. Further, the autoencoder based approach is used to refine the outcome of LR and SVM, as highlighted in Fig. 18c-d. The accuracy obtained during model testing was recorded to be 97.07 % for auto-encoder LR, and 99.16 % for auto-encoder SVM, respectively.

However, according to the results obtained from the t-SNE plots in Fig. 19, it is apparent that even after 100 epochs, the auto-encoder based deep learning approach fails to



**Fig. 18** Survival prediction for isolated cases in South Korea with (a) ANN with binary cross-entropy & Adam optimizer, and (b) SVM with parameter optimization (c) Auto-encoder LR (d) Auto-encoder SVM



**Fig. 19** t-SNE plots for evaluating deceased & released cases while model training with auto-encoder. The graphs show (a) original divergence, and outcomes for (b) 10 epochs (c) 50 epochs (d) 100 epochs

appropriately classify the deceased and released cases. Hence, the outcome obtained from applying auto-encoder cannot be considered reliable in predicting the health criticality of isolated patients. Therefore, among all AI-based model predictions, the survival chances predicted by the ANN model with *Adam* optimizer is found to be quite high and is also achieved with the best accuracy. This equally justified with the present scenario of South Korea, where the death curve has considerably flattened, thereby validating our model predictions.

#### 6.4 Theoretical investigation of existing approaches

An extensive theoretical investigation of various existing approaches is illustrated in Table 2. The exploratory study is performed in terms of the suitability of the proposed approach alongside each of its existing counterparts in the context of COVID-19 prediction analysis.

## 7 Conclusions

The research provided an analysis of the disease spread to discover the death estimators for the COVID-19 pandemic

in South Korea. Several factors including the time-series data (age, gender, living province), location data, and epidemiological data (number of days under treatment, distribution of confirmed and deceased cases, temporal pattern in the disease evolution) have impressively influenced our study and hence, were utilized for assessing the pandemic outbreak. The analysis of the COVID-19 positive cases with the help of the proposed analytical model indicated that deceased cases are at their peak at the starting and mid of the disease outbreak, but the death rate degraded drastically, indicating individual resistance to the virus. Also, the results concluded that the mortality rate was high for old aged people, and males were found to be at higher risk than females in South Korea.

Our research work presented a predictive analysis of quarantined COVID-19 cases using several artificial intelligence-based models with the tuning of hyperparameters and exhaustive exploration of model learning capacities. The model was trained and tested on 5165 coronavirus instances and was validated over 1533 quarantined patients. Logistic regression analysis resulted in an average accuracy of 91 % over total corona causalities, and SVM model predictions were recorded to be 97 % with hyperparameters tuning.



**Table 2** Comparative Analysis of our proposed approach with other recent studies

S.No.	Description	Approach	Results (Accuracy)
1.	We have implemented machine learning techniques along with ensemble deep learning techniques for achieving improved predictive performance.	Proposed Approach (1) LR (2) SVM (3) ANN with Adam Optimizer (4) ANN with Adagrad Optimizer (5) ANN with Autoencoder based SVM (6) ANN with Autoencoder based LR	(1) 91% (2) 97% (3) 99% (4) 99% (5) 99.16% (6) 97.07%
2.	An ensemble approach using Convolutional Neural Network (CNN) and Support Vector Machine is developed [41, 42]	Support Vector Machine+Convolutional Neural Network (CNN)	97%
3.	Research predicting the cumulative number of daily cases, deaths, and recovered cases using SVM [43]	Support Vector Regression	97% in death prediction, and 87% in forecasting daily new cases
4.	A comparative analysis of various traditional prediction techniques [44]	(1) SVM (2) Neural Network (3) Naïve Bayes (4) K-Nearest Neighbor (5) Decision Table (6) Decision Stump (7) OneR (8) ZeroR	(1) 92.95% (2) 92.89% (3) 90.58% (4) 92.89% (5) 92.95% (6) 70.73% (7) 68.36% (8) 57.86%
5.	An Embedded approach was developed for prediction purposes. It is the amalgamation of Boosting and Random Forest [45]	Boosted Random Forest	94%
6.	Several machine learning techniques analyzed on COVID-19 data for analyzing the prediction performance [46]	(1) IPCRidge (2) CoxPH (3) Coxnet (4) Stagewise GB (5) Componentwise GB (6) Fast SVM (7) Fast Kernel SVM	(1) 49.05% (2) 70.63% (3) 70.72% (4) 71.47% (5) 70.60% (6) 70.65% (7) 61.05%
7.	A stacked autoencoder detector model is proposed to improve overall statistical parameter of prediction models like accuracy, precision rate, and recall rate [47]	Stacked autoencoder	94.70%
8.	A convolutional neural network (CNN) architectures along with transfer learning are proposed for medical classification [48]	Transfer learning with convolutional neural networks	96.78%
9.	A forecasting model for COVID-19 spread is created using a neural network [49]	Neural Network	Average accuracy approx. 97%
10.	An artificial neural network approach is developed to detect COVID-19 disease using capsule networks [50]	Artificial Neural Network	97.24% for binary class, and 84.22% for multi-class

Finally, an accuracy of 99 % was found using deep learning sequential model with 6 hidden layers that were built over 30 neurons in each layer, along with *Adam* optimizer and binary entropy cross-validation to prevent the model from overfitting. Moreover, we have also extended our study with ANN autoencoders, and an accuracy of 97.07 %, and 99.16 % were recorded for autoencoder-based LR, and autoencoder-based SVM respectively. Based on our results, we presented trustworthy estimates for the health practitioners for their prudent groundwork.

**References**

1. Liu Y, Lee JM, Lee C (2020) The challenges and opportunities of a global health crisis: the management and business implications of COVID-19 from an Asian perspective. *Asian Bus Manag* 19:277–297. <https://doi.org/10.1057/s41291-020-00119-x>
2. Bedford J, Enria D, Giesecke J, Heymann DL, Ihekweazu C, Kobinger G, Lane HC, Memish Z, Oh MD, Schuchat A, Ungchusak K (2020) COVID-19: towards controlling of a pandemic. *Lancet* 395(10229):1015–1018. [https://doi.org/10.1016/S0140-6736\(20\)30673-5](https://doi.org/10.1016/S0140-6736(20)30673-5)

3. Shim E, Tariq A, Choi W, Lee Y, Chowell G (2020) Transmission potential and severity of COVID-19 in South Korea. *Int J Infect Dis* 93:339–344. <https://doi.org/10.1016/j.ijid.2020.03.031>
4. Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, Ji R, Wang H, Wang Y, Zhou Y (2020) Prevalence of comorbidities in the novel Wuhan coronavirus (COVID-19) infection: a systematic review and meta-analysis. *Int J Infect Dis* 94:91–95. <https://doi.org/10.1016/j.ijid.2020.03.017>
5. Ruan Q, Yang K, Wang W, Jiang L, Song J (2020) Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive Care Med* 46:846–848. <https://doi.org/10.1007/s00134-020-05991-x>
6. Li B, Yang J, Zhao F, Zhi L, Wang X, Liu L, Bi Z, Zhao Y (2020) Prevalence and impact of cardiovascular metabolic diseases on COVID-19 in China. *Clin Res Cardiol* 109:531–538. <https://doi.org/10.1007/s00392-020-01626-9>
7. Lippi G, Lavie CJ, Sanchis-Gomar F (2020) Cardiac troponin I in patients with coronavirus disease 2019 (COVID-19): Evidence from a meta-analysis. *Prog Cardiovasc Dis* 63(3):390–391. <https://doi.org/10.1016/j.pcad.2020.03.001>
8. Shi S, Qin M, Shen B, Cai Y, Liu T, Yang F, Gong W, Liu X, Liang J, Zhao Q, Huang H (2020) Association of cardiac injury with mortality in hospitalized patients with COVID-19 in Wuhan, China. *JAMA Cardiol* 5(7):802–810. <https://doi.org/10.1001/jamacardio.2020.0950>
9. Guo T, Fan Y, Chen M, Wu X, Zhang L, He T, Wang H, Wan J, Wang X, Lu Z (2020) Cardiovascular implications of fatal outcomes of patients with coronavirus disease 2019 (COVID-19). *JAMA Cardiol* 5(7):811–818
10. Driggin E, Madhavan MV, Bikdeli B, Chuich T, Laracy J, Biondi-Zoccai G, Brown TS, Der Nigoghossian C, Zidar DA, Haythe J, Brodie D (2020) Cardiovascular considerations for patients, health care workers, and health systems during the coronavirus disease 2019 (COVID-19) pandemic. *J Am Coll Cardiol* 75:2352–2371. <https://doi.org/10.1016/j.jacc.2020.03.031>
11. Lippi G, Plebani M, Henry BM (2020) Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: A meta-analysis. *Clin Chim Acta* 506:145–148. <https://doi.org/10.1016/j.cca.2020.03.022>
12. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, Tao Q, Sun Z, Xia L (2020) Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology*. <https://doi.org/10.1148/radiol.2020200642>
13. Bernheim A, Mei X, Huang M, Yang Y, Fayad ZA, Zhang N, Diao K, Lin B, Zhu X, Li K, Li S (2020) Chest CT findings in coronavirus disease-19 (COVID-19): relationship to duration of infection. *Radiology*, 200463. <https://doi.org/10.1148/radiol.2020200463>
14. Shi H, Han X, Jiang N, Cao Y, Alwalid O, Gu J, Fan Y, Zheng C (2020) Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *Lancet Infect Dis* 20(4):425–434. [https://doi.org/10.1016/S1473-3099\(20\)30086-4](https://doi.org/10.1016/S1473-3099(20)30086-4)
15. Xu X, Jiang X, Ma C, Du P, Li X, Lv S, Yu L, Ni Q, Chen Y, Su J, Lang G (2020) A deep learning system to screen novel coronavirus disease 2019 pneumonia. *Engineering* 6(10):1122–1129. <https://doi.org/10.1016/j.eng.2020.04.010>
16. Wang L, Lin ZQ, Wong A (2020) COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images. *Sci Rep* 10:19549. <https://doi.org/10.1038/s41598-020-76550-z>
17. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, Bai J, Lu Y, Fang Z, Song Q, Cao K (2020) Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct. *Radiology* 292(6):65–71
18. Zhang J, Xie Y, Li Y, Shen C, Xia Y (2020) Viral Pneumonia Screening on Chest X-rays Using Confidence-Aware Anomaly Detection. *IEEE transactions on medical imaging* 1:1–12. arXiv preprint 2003.12338
19. Elfiky AA (2020) Anti-HCV, nucleotide inhibitors, repurposing against COVID-19. *Life Sci* 117477. <https://doi.org/10.1016/j.lfs.2020.117477>
20. Gao J, Tian Z, Yang X (2020) Breakthrough: Chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies. *Biosci Trends* 14:72–73. <https://doi.org/10.5582/bst.2020.01047>
21. Touret F, de Lamballerie X (2020) Of chloroquine and COVID-19. *Antiviral Res* 177:104762. <https://doi.org/10.1016/j.antiviral.2020.104762>
22. Cortegiani A, Ingoglia G, Ippolito M, Giarratano A, Einav SA (2020) A systematic review on the efficacy and safety of chloroquine for the treatment of COVID-19. *J Crit Care* 57:279–283. <https://doi.org/10.1016/j.jcrc.2020.03.005>
23. Kalil AC (2020) Treating COVID-19—off-label drug use, compassionate use, and randomized clinical trials during pandemics. *JAMA* 323(19):1897–1898
24. Gautret P, Lagier JC, Parola P, Meddeb L, Mailhe M, Doudier B, Courjon J, Giordanengo V, Vieira VE, Dupont HT, Honoré S (2020) Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *Int J Antimicrob Agents* 105949. <https://doi.org/10.1016/j.ijantimicag.2020.105949>
25. Baron SA, Devaux C, Colson P, Raoult D, Rolain JM (2020) Teicoplanin: an alternative drug for the treatment of coronavirus COVID-19. *Int J Antimicrob Agents* 105944. <https://doi.org/10.1016/j.ijantimicag.2020.105944>
26. Colson P, Rolain JM, Lagier JC, Brouqui P, Raoult D (2020) Chloroquine and hydroxychloroquine as available weapons to fight COVID-19. *Int J Antimicrob Agents* 105932. <https://doi.org/10.1016/j.ijantimicag.2020.105932>
27. Cai Q, Yang M, Liu D, Chen J, Shu D, Xia J, Liao X, Gu Y, Cai Q, Yang Y, Shen C (2020) Experimental treatment with favipiravir for COVID-19: an open-label control study. *Engineering* 6:1192–1198. <https://doi.org/10.1016/j.eng.2020.03.007>
28. Wu D, Yang XO (2020) TH17 responses in cytokine storm of COVID-19: An emerging target of JAK2 inhibitor Fedratinib. *J Microbiol Immunol Infect* 53(3):368–370. <https://doi.org/10.1016/j.jmii.2020.03.005>
29. Shan F, Gao Y, Wang J, Shi W, Shi N, Han M, Xue Z, Shi Y (2020) Lung infection quantification of COVID-19 in CT images with deep learning. arXiv preprint arXiv:2003.04655
30. Dong L, Hu S, Gao J (2020) Discovering drugs to treat coronavirus disease 2019 (COVID-19). *Drug Discov Ther* 14(1):58–60. <https://doi.org/10.5582/ddt.2020.01012>
31. Mienye ID, Sun Y, Wang Z (2020) Improved sparse autoencoder based artificial neural network approach for prediction of heart disease. *Inform Med Unlocked* 100307. <https://doi.org/10.1016/j.imu.2020.100307>
32. Pereira IG, Guerin JM, Junior AG, Distanto C, Garcia GS, Goncalves LM (2020) Forecasting Covid-19 dynamics in Brazil: a data driven approach. *Int J Environ Res Public Health* 17(14): 5115
33. Chen J, Yuan W, Cao J, Lv H (2019) Traffic-flow prediction via granular computing and stacked autoencoder. *Granul Comput* 5: 449–459. <https://doi.org/10.1007/s41066-019-00167-5>
34. Bas E, Egrioglu E, Yolcu U, Grosan C (2019) Type 1 fuzzy function approach based on ridge regression for forecasting. *Granul Comput* 4(4):629–637. <https://doi.org/10.1007/s41066-018-0115-4>
35. Amezcua J, Melin P (2019) A new fuzzy learning vector quantization method for classification problems based on a granular approach. *Granul Comput* 4(2):197–209. <https://doi.org/10.1007/s41066-018-0120-7>

36. Liu H, Cocea M (2019) Granular computing-based approach of rule learning for binary classification. *Granul Comput* 4(2):275–283. <https://doi.org/10.1007/s41066-018-0097-2>
37. Liu H, Cocea M (2019) Nature-inspired framework of ensemble learning for collaborative classification in granular computing context. *Granul Comput* 4(4):715–724. <https://doi.org/10.1007/s41066-018-0122-5>
38. Soh CW, Njilla LL, Kwiat KK, Kamhoua CA (2020) Learning quasi-identifiers for privacy-preserving exchanges: A rough set theory approach. *Granul Comput* 5(1):71–84. <https://doi.org/10.1007/s41066-018-0127-0>
39. Melin P, Sánchez D (2019) Optimization of type-1, interval type-2 and general type-2 fuzzy inference systems using a hierarchical genetic algorithm for modular granular neural networks. *Granul Comput* 4(2):211–236. <https://doi.org/10.1007/s41066-018-0133-2>
40. Egrioglu E, Yolcu U, Bas E (2019) Intuitionistic high-order fuzzy time series forecasting method based on pi-sigma artificial neural networks trained by artificial bee colony. *Granul Comput* 4(4):639–654. <https://doi.org/10.1007/s41066-018-00143-5>
41. Kaggle Dataset (2020) DS4C: Data Science for COVID-19 in South Korea. License CC BYNC-SA 4.0. <https://www.kaggle.com/kimjihoo/coronavirusdataset>
42. Novitasari DC, Hendradi R, Caraka RE, Rachmawati Y, Fanani NZ, Syarifudin A, Toharudin T, Chen RC (2020) Detection of COVID-19 chest X-ray using support vector machine and convolutional neural network. *Commun Math Biol Neurosci* 42. <https://doi.org/10.28919/cmbn/4765>
43. Parbat D, Chakraborty M.A python based support vector regression model for prediction of COVID19 cases in India. *Chaos Solitons Fractals* 138:109942. <https://doi.org/10.1016/j.chaos.2020.109942>
44. Otoom M, Otoum N, Alzubaidi MA, Etoom Y, Banihani R (2020) An IoT-based framework for early identification and monitoring of COVID-19 cases. *Biomed Signal Process Control* 62:102149. <https://doi.org/10.1016/j.bspc.2020.102149>
45. Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, Mishra R, Pillai S, Jo O (2020) COVID-19 Patient health prediction using boosted random forest algorithm. *Front Public Health* 8:357. <https://doi.org/10.3389/fpubh.2020.00357>
46. Nemati M, Ansary J, Nemati N (2020) Machine-learning approaches in COVID-19 survival analysis and discharge-time likelihood prediction using clinical data. *Patterns* 1(5):100074. <https://doi.org/10.1016/j.patter.2020.100074>
47. Li D, Fu Z, Xu J (2020) Stacked-autoencoder-based model for COVID-19 diagnosis on CT images. *Appl Intell* 9:1–3. <https://doi.org/10.1007/s10489-020-02002-w>
48. Apostolopoulos ID, Mpesiana (2020) TA Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Phys Eng Sci Med* 3:1. <https://doi.org/10.1007/s13246-020-00865-4>
49. Wiczorek M, Siłka J, Woźniak M (2020) Neural network powered COVID-19 spread forecasting model. *Chaos Solitons Fractals* 140:110203. <https://doi.org/10.1016/j.chaos.2020.110203>
50. Toraman S, Alakus TB, Turkoglu I (2020) Convolutional capsnet: A novel artificial neural network approach to detect COVID-19 disease from X-ray images using capsule networks. *Chaos Solitons Fractals* 140:110122. <https://doi.org/10.1016/j.chaos.2020.110122>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.