priyadarshini engineering college

COVID-19 CASE ANALYSIS

October 17,2023

Phase 3

Presented To

Wasim Akram.R

Register.no: 511921104086

Email : wasimmsd030@gmail.com

ABSTRACTION

Today world thinks about coronavirus disease that which means all even this pandemic disease is not unique. The purpose of this study is to detect the role of machine-learning applications and algorithms in investigating and various purposes that deals with COVID-19. Review of the studies that had been published during 2020 and were related to this topic by seeking in Science Direct, Springer, Hindawi, and MDPI using COVID-19, machine learning, supervised learning, and unsupervised learning as keywords. The total articles obtained were 16,306 overall but after limitation; only 14 researches of these articles were included in this study. Our fndings show that machine learning can produce an important role in COVID-19 investigations, prediction, and discrimination. In conclusion, machine learning can be involved in the health provider programs and plans to assess and triage the COVID-19 cases. Supervised learning showed better results than other Unsupervised learning algorithms by having 92.9% testing accuracy. In the future recurrent supervised learning can be utilized for superior accuracy.

keywords: COVID-19 · Artifcial intelligence AI · Machine learning · Machine learning tasks · Supervised and un-supervised learning

INTRODUCTION

Recently, the world gained rapid progression in technology and it shows an important role in the developed countries. Nowadays all daily life sectors such as education, business, marketing, militaries, and communications, engineering, and health sectors are dependent on the new technology applications. The health care center is a crucial field that strongly needs to apply the new technologies from defining the symptoms to the accurate diagnosis and digital patient's triage. Coronavirus-2 (SARSCoV-2) causes severe respiratory infections, and respiratory disorders, which results in the novel coronavirus disease 2019 (COVID-19) in humans who had been reported as the first case in Wuhan city of China in December 2019. Later, SARS-CoV-2 was spread worldwide and transmitted to millions of people and the world health organization (WHO) have announced the outbreak as a global pandemic since the number of infected people is still increasing day by day. As of 16th December 2020, the total (global) coronavirus cases were approximately 73,806,583 with reported deaths of 1,641,635 (Pasupuleti et al. 2021).

The novel coronavirus appeared in December 2019, in the Wuhan city of China and the World Health Organization (W.H.O) reported it on 31st December 2019. The virus produced a global risk and W.H.O named it COVID-19 on 11th February 2020 (Wu 2020). Up to the present time, there was no specific medication that deals directly with this new generation of COVID-19 virus, but some of the companies produced several combination drugs that basically made up from ethanol, isopropyl alcohols, and hydrogen peroxides in different combinations show a significant reaction to the novel virus and had been confirmed and accepted by WHO to be used in the world (Mahmood et al. 2020). The artificial intelligence and deep learning algorithm show the ability to diagnose COVID-19 in precise which can be regarded as a supportive factor to improve the common diagnostic methods including Immunoglobulin M (IgM), Immunoglobulin (IgG), chest x-ray, and computed tomography(CT) scan, also reverse transcription-polymerase chain reaction (RT-PCR) and immunochroma to graphic fluorescence assay. The developments of a potential technology are one of the currently used methods to identify the infection, such as a drone with thermal screening without human intervention, which needs to be encouraged (Manigandan 2020).

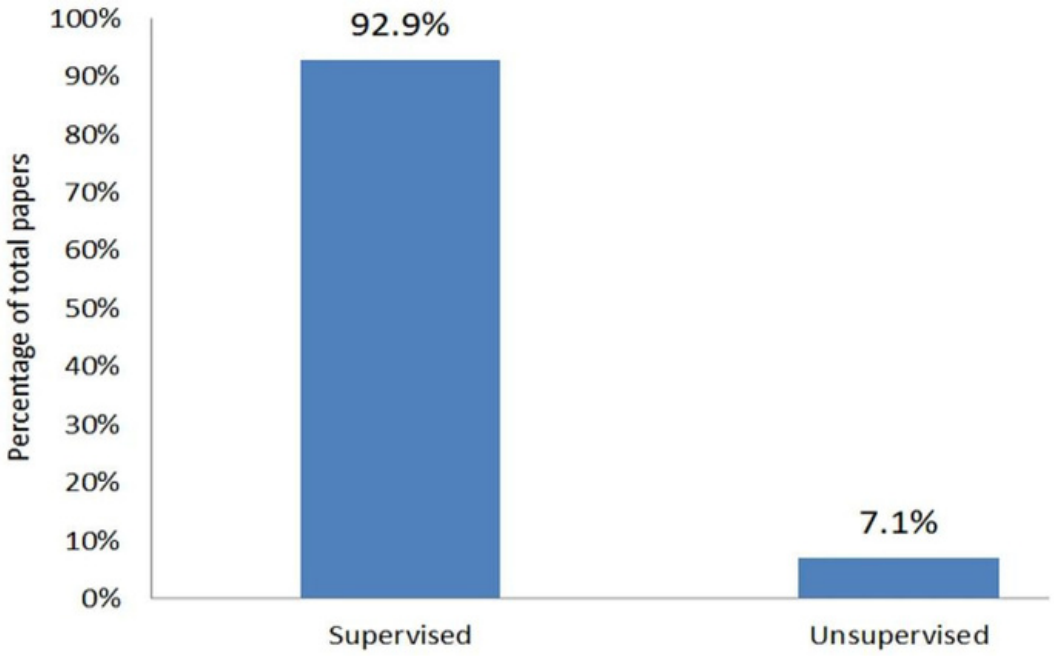
COVID-19 WITH MACHINE LEARNING

Recently there are three different perspectives of work that had been done on edge computing and the detection of (COVID-19) Cases. The viewpoints are including the recognizing of (COVID-19) cases by machine-learning systems (Table 1). The algorithms for the recognition of activity from machine learning and the approaches which used in edge computing are considered the Imaging workflows that can inspire machine-learning methods that are able of supporting radiologists who search for an analysis of complex imaging and text data. For the novel COVID-19 there are models capable of analyzing medical imaging and recognizing COVID-19 (Shirzadi 2018). Artificial intelligence AI has various types, machine learning (ML), is one of these applications, it had been applied successfully to different fields of medicine for detection of new genotype–phenotype associations, diagnosis, which showed effects on assessment, prediction, diseases classification, transcriptomic, and minimizing the death ratio(Gao 2020).

The technique of automatic classification of COVID-19 can be applied by comparing general deep learning-based feature extraction frameworks to achieve the higher accurate feature, which is an important module of learning, MobileNet, DenseNet, Xception, ResNet, InceptionV3, InceptionResNetV2, VGGNet, NASNet were selected among a group of deep convolutional neural networks CNN. The classification then achieved by running the extracted features into some of machine-learning classifiers to recognize them as a case of COVID-19 or other diseases (Bishop 2006). Progressive machine-learning algorithms can integrate and evaluate the extensive data that is related to COVID-19 patients to provide best understanding of the viral spread pattern, increase the diagnostic accuracy, improve fresh, and effective methods of therapy, and even can recognize the individuals who, at risk of the disease depending on the genetic and physiological features (Khanday 2020).

MACHINE-LEARNING TYPE APPLIED

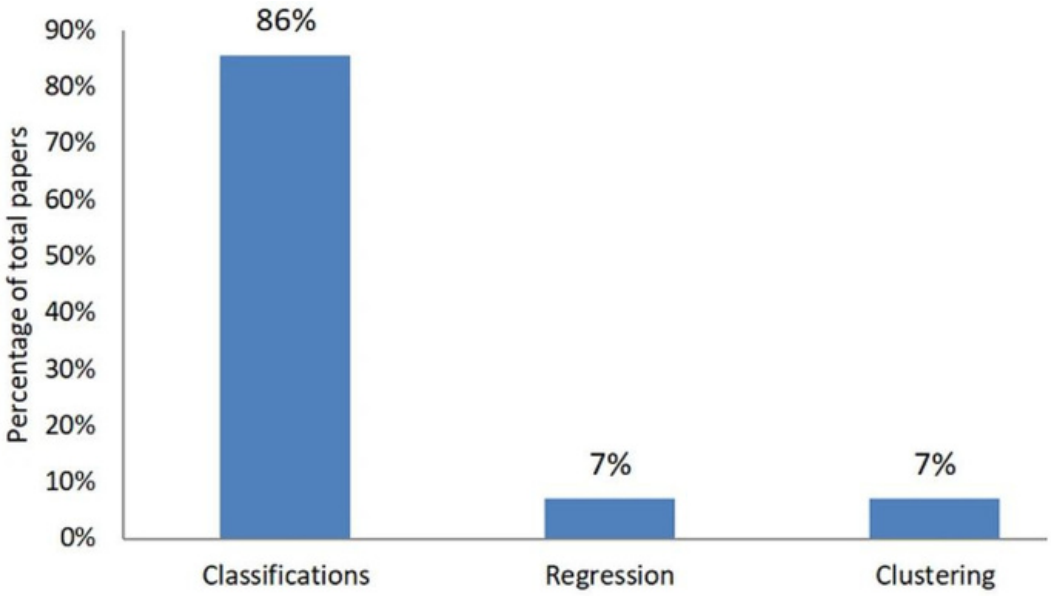
supervised learning is the dominant machine-learning type applied for production lines. The majority of studies used both supervised learning methods which were (92.9%), whereas unsupervised learning was (7.1%).



Distribution of machine-learning types

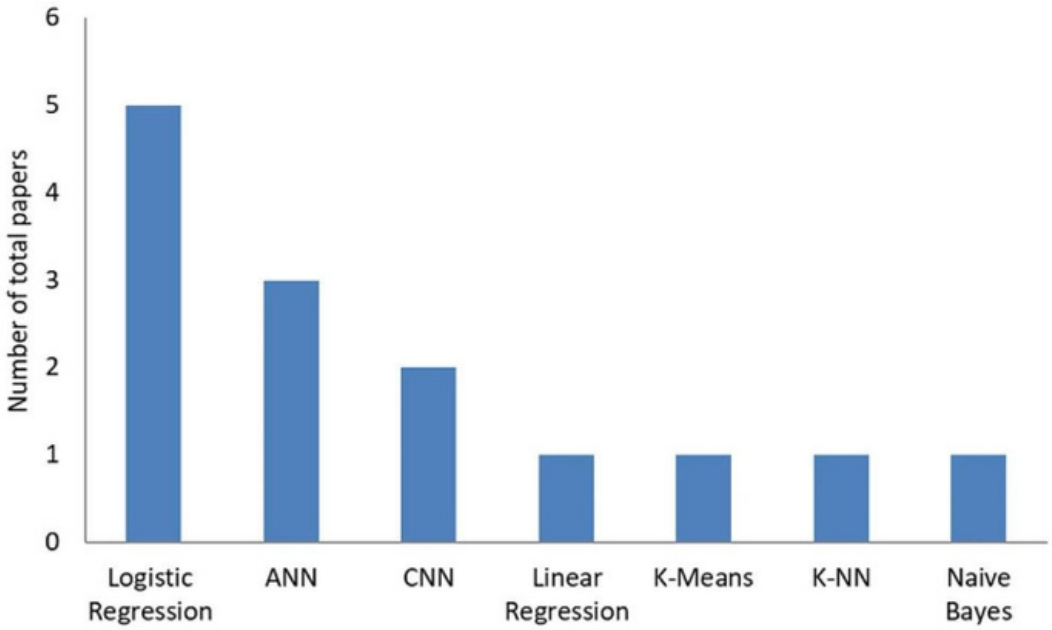
MACHIMACHINE-LEARNING TASKS ADDRESS

which accounts for about (86%) of all selected papers. There are about (7%) of papers that applied for each of the regression and clustering.



MACHINE-LEARNING ALGORITHMS USED

shows that the logistic regression is largely applied in production lines. Logistic regression is the most frequently applied machine-learning algorithm, including five papers in 14 papers. Artificial neural network algorithm (ANN) and CNN (convolutional neural network) are in the second and third ranks which were three and two papers in 14 papers, respectively. Linear regression, K-Means, KNN (K-nearest neighbors), and Naive Bayes are the other algorithms applied for production lines.



DISCUSSIONS AND IMPLICATIONS

The new transmitted virus was discovered and spread out from Wuhan city of China in December 2019 and affected more than (100) countries around the world in a very short time (Wu 2020). It was represented and introduced to the World Health Organization (W.H.O) on 31st December 2019. The virus was then termed COVID-19 by W.H.O on 11th February 2020, because it formed a global risk (Wu 2020). This family of viruses also includes SARS, ARDS. W.H.O confirmed this eruption as a public health emergency (Manigandan et al. 2020). Technology progressions have a fast effect on each field of life; the medical field is one of the important direct daily related to people's lives. Recently Artificial intelligence AI had been introduced to the medical field and it has shown promising outcomes in health care due to the high accuracy of data analysis which makes an exact decision making. Researchers all over the world tried to find a method to improve the clinical diagnosis and minimize the rapid spread of this virus so that they involved AI algorithms in the diagnosis of this disease. This review paper explains various AI algorithms that people used in their researches and will compare their results to demonstrate the best accurate method that shows the most improving in COVID-19 diagnosis. The total studies that used in this research are (14) original articles, all of them used supervised learning as the main method, but the algorithms were differed among them according to the research purpose.

A study recently published 2020 in India they extracted their dataset from GitHub which was 212 reports of 1000 cases, they used supervised learning as their main method in machine-learning application, and the algorithm that they applied was classification logistic regression and multinominal Nia″ve Bayes. The findings showed that Logistic regression and multinominal Nia″ve Bayes are better than the commonly used algorithms according to 96% accuracy obtained from the findings (Khanday 2020). Scientists in the USA published an article 2020 they relied on United States health systems to custom 197 patients as their data, the main method that they used was supervised learning, while the algorithm was classification logistic regression, their results showed that this algorithm displays higher diagnostic odds ratio (12.58) for foreseeing ventilation and effectively triage patients than a comparator early warning system, such as Modified Early Warning Score (MEWS) which showed (0.78) sensitivity, while this algorithm showed (0.90) sensitivity which leads to higher specificity ($p < 0.05$), also it shows the capability of accurate identification 16% of patients more than a commonly used scoring system which results in minimizing false-positive results (Burdick 2020a). Varun et al.used 184,319 reported cases as a ataset in his article in which he applied the same method supervised learning but with a different algorithm which was convolutional neural network CNN and their outcomes were in response to this crisis, the medical and academic centers in New York City issued a call to action to artificial intelligence researchers to leverage their electronic medical record (EMR) data to better understand SARS-COV-2 patients. Due to the scarcity of ventilators and a reported need for a quick and accurate method of triaging patients at risk for respiratory failure, our purpose was to develop a machine-learning algorithm for frontline physicians in the emergency department and the inpatient floors to better risk-assess patients and predict who would require intubation and mechanical ventilation (Arvind 2020). Meanwhile, another study had been published in Italy by (Luca et al. 2020)

The research findings stated that countries with the highest death ratio were those who had a high consumption of fats, while countries with a lower death rate have a higher level of cereal consumption followed by a lower total average intake of kilocalories (García-Ordás, et al. (2020)). A study conducted to (Shinwoo et al. 2020) their research data were extracted from the immigrant Korean COVID-19 patients who were 290 cases from 12 states all of them older than 18 years, the study observed the ability to the prediction of discrimination-related variables, such as racism effects, and sociodemographic factors that influence the psychological distress level during the COVID-19 pandemic, they nominated the supervised learning as the method and then using the Artificial Neural Network ANN as the main algorithm, their result showed The Artificial Neural Network (ANN) analysis, which is a statistical model and able to examine complex non-linear interactions of variables, was applied. The algorithm perfectly predicted the person's flexibility, familiarities of everyday discernments, and the racist actions toward Asians in the U.S. since the beginning of the COVID-19 pandemic which finally provides important suggestions for public health practitioners (Choi 2020). During the same time, a study presented by (Yigrem et al. 2020) conducted a cross-study based on 244 of the healthcare providers in Dilla, Southern Ethiopia. Supervised learning was used in the methodology and then they analyzed the data by logistic regression algorithm to find the association between the perceived stress of COVID-19 and the health care providers. Results showed that more than half of the research participants were presented with perceived stress of coronavirus disease, which means that there is a strong correlation between the health care staff and perceived stress of COVID-19 (Chekole, et al. (2020) Finally, the last article conducted by (Abolfazl et al. 2020) their study used 57 samples of COVID-19 cases from the USA to find out the relationship between the sociodemographic and environmental variables, other diseases, such as chronic heart disease, leukemia, and pancreatic cancer, also socioeconomic factors and the death ratio due to COVID-19 disease. Results showed that the presented model (logistic regression) shown that these factors and variables describe the presence/absence of the hotspot of the COVID-19 incidence which was clarified by Getis-Ord Gi ($p < 0.05$) in a geographic information system. As a result, the findings provided valuable insights for public health decision makers in categorizing the effect of the potential risk factors associated with COVID-19 incidence level (Mollalo et al. 2020)

EMAIL

wasimmsd030@gmail.com

REG.NO

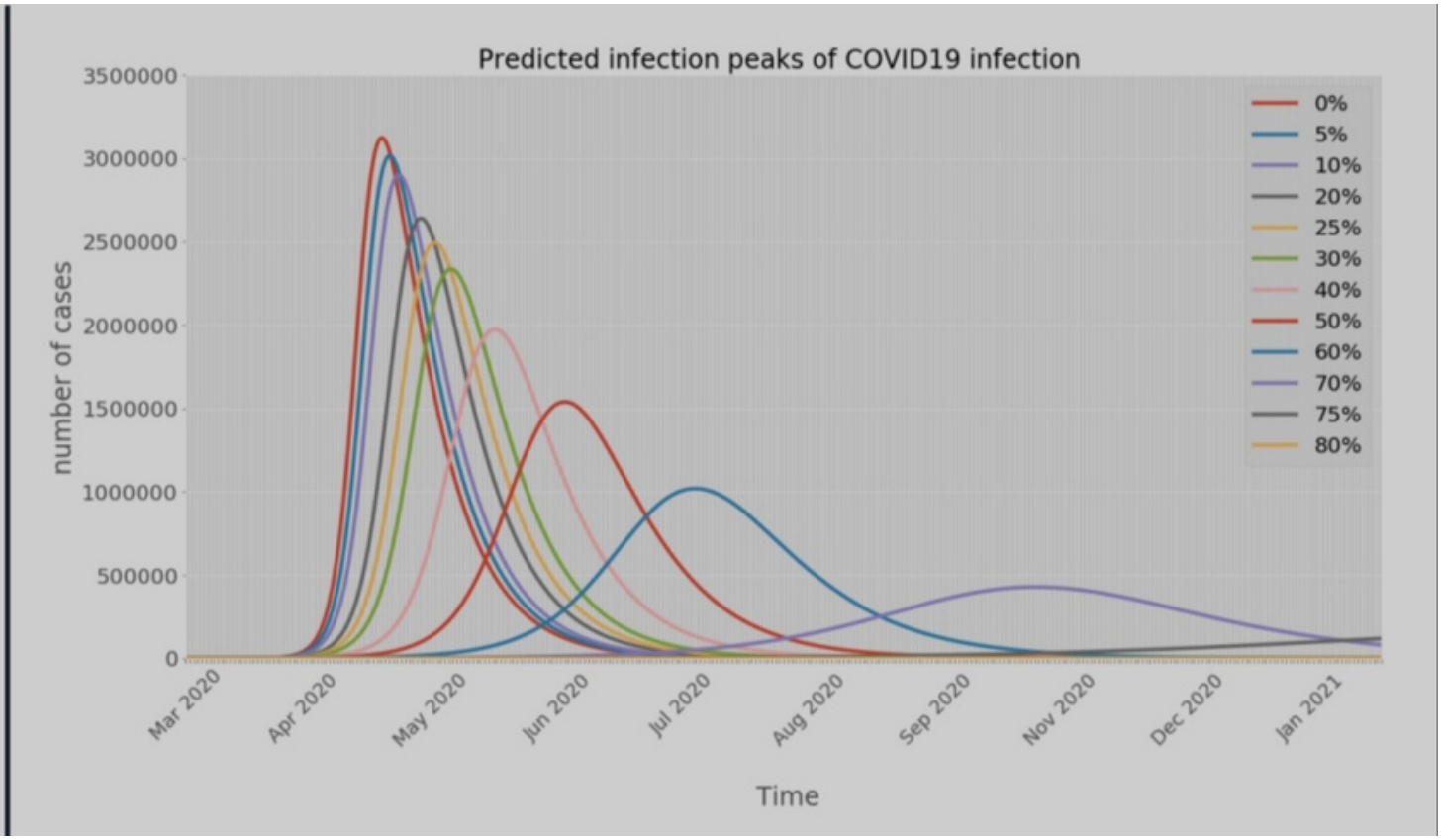
511921104086

SAMPLE PROGRAM USING PYTHON PROGRAM

```
# Simulations visualization over several months and computed model predictions
DATES = ["Mar 2020"]
M = [ "Apr 2020 " , "May 2020 ", "Jun 2020 ", "Jul 2020 ", "Aug 2020 ", "Sep 2020 ", "Nov 2020 ", "Dec
2
020 ", "Jan 2021 ", "Feb 2021 " ]
K = [ 29, 60 , 91, 121 , 152, 183 , 213, 244, 274, 305, 336 ] # the correspondent number of days
z= 0
for j, L in enumerate(K):
 for i in range( z, L ):
 if i <= 300:
 DATES.append("")
 if i <= 300:
 DATES.append(M[j])
 z = K[j]

plt.clf()
plt.xlabel('\n Time', size = 24 )
plt.ylabel('number of cases', size =24 )
plt.title("Predicted infection peaks of COVID19 infection ", size =24)
plt.tick_params(axis='both', labelsize = 20) # set the font of axis values
plt.ticklabel_format(axis = "x", style = "sci", useMathText = "True")
for l in range(len(Infected)):
 plt.plot(Time, Infected[l], label = perturbations[l], linewidth = 3.5 )
plt.xticks(np.arange(600), DATES , fontsize=18, rotation = 45)
plt.legend( fontsize = 20)
plt.axis([0, 300, 0, 3500000] ) # max([max(Y12), max(Y3), max(Y4) ]) *1.1 ] )
plt.show()
# Get the peak times and expected hospital peaks
alpha = 354.0/4268 # estimated the current number of hostilazations per confirmed cases
beta = 76.0/4268 # estimated current number of deaths per confirmed cases
print ("Hospitalization fraction = ", round(alpha, 3), " (" , round(alpha*100,1), " %)" )
print ("Death fraction = ", round(beta, 3) , " (" , round(beta*100,1), " %)" , "\n" )
for i, sim in enumerate(Infected):
 Quarentine = perturbations[i]
 Peak = max(sim)
 Total = TotalInf[i][sim.index(Peak)]
 T = Time[sim.index(Peak)]
 Hospital = Peak* alpha
 Death = Peak*beta
 print ( Quarentine , " --> pk time = ", int(T) , " days Hospital = ",
 int(Hospital) , " Deaths = ", int(Death), " Total infected = ",
 int(Total ) , " infected (pk)= ", int(Peak) )
```

PREDICTED INFECTION PEAK OF COVID-19 INJECTION



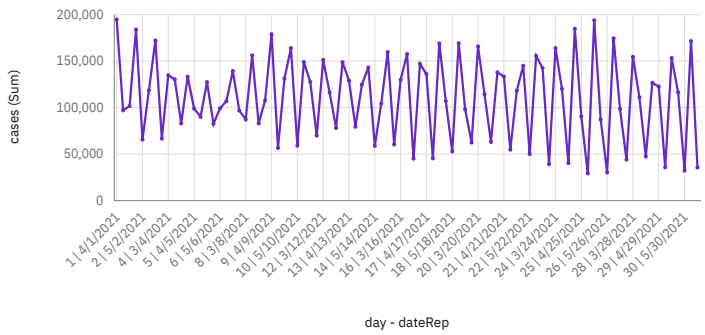
Hospitalization fraction = 0.083 (8.3 %)
Death fraction = 0.018 (1.8 %)
0% --> pk time = 49 days Hospital = 259055 Deaths = 55616 Total infected = 5156932 infected (pk)= 3123301
5% --> pk time = 51 days Hospital = 250125 Deaths = 53699 Total infected = 5094458 infected (pk)= 3015636
10% --> pk time = 53 days Hospital = 240573 Deaths = 51648 Total infected = 5081874 infected (pk)= 2900468
20% --> pk time = 59 days Hospital = 219034 Deaths = 47024 Total infected = 4892567 infected (pk)= 2640788
25% --> pk time = 62 days Hospital = 206911 Deaths = 44421 Total infected = 4791783 infected (pk)= 2494627
30% --> pk time = 66 days Hospital = 193740 Deaths = 41594 Total infected = 4648087 infected (pk)= 2335835
40% --> pk time = 77 days Hospital = 163696 Deaths = 35143 Total infected = 4342700 infected (pk)= 1973605
50% --> pk time = 95 days Hospital = 127656 Deaths = 27406 Total infected = 3910457 infected (pk)= 1539093
60% --> pk time = 128 days Hospital = 84499 Deaths = 18141 Total infected = 3295201 infected (pk)= 1018771
70% --> pk time = 213 days Hospital = 35509 Deaths = 7623 Total infected = 2202358 infected (pk)= 428124
75% --> pk time = 350 days Hospital = 12685 Deaths = 2723 Total infected = 1354146 infected (pk)= 152938
80% --> pk time = 500 days Hospital = 54 Deaths = 11 Total infected = 22886 infected (pk)= 662

CONCLUSION

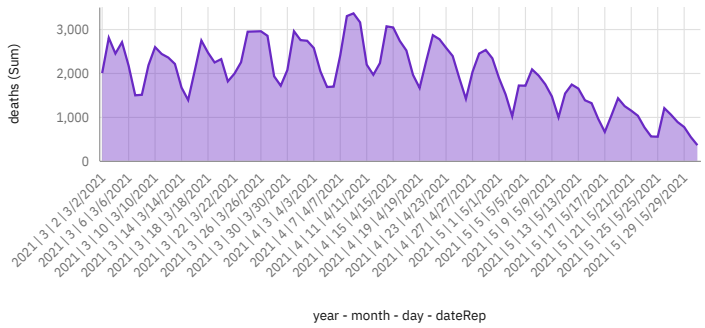
This study focused on the articles that applied machine-learning applications in COVID-19 disease for various purposes with different algorithms, 14 from 16 articles used supervised learning, and only one among them used unsupervised learning another one used both methods supervised and unsupervised learning and both of them shows accurate results. The studies used different machine-learning algorithms in different countries and by different authors but all of them related to the COVID-19 pandemic, (5) of these articles used Logistic regression algorithm, and all of them showed promising results in the COVID-19 health care applications and involvement. While (3) of the articles used artificial neural network (ANN) which also shows successful results, the rest of the 14 articles used different supervised and unsupervised learning algorithms and all of the models showed accurate results. Our conclusion is ML applications in medicine showed promising results with high accuracy, sensitivity, and specificity using different models and algorithms. In general, the paper results explored the supervised learning is more accurate to detect the COVID-19 cases which were above (92%) compare to the unsupervised learning which was mere (7.1%).

Tab 1

cases by day and dateRep



deaths by year, month, day and dateRep



deaths and cases for countriesAndTerritories regions