COVID-19 Outbreak Prediction and Analysis of E-Healthcare Data Using Random Forest Algorithms

Debabrata Dansana, KISS University (Deemed), India

iD https://orcid.org/0000-0002-8818-7652

Raghvendra Kumar, GIET University, India*

iD https://orcid.org/0000-0003-1311-7585

Aishik Bhattacharjee, GIET University, India

Chandrakanta Mahanty, GIET University, India

ABSTRACT

The forecasting model used random forest algorithm. From the outcomes, it has been found that the regression models utilize basic linkage works and are exceptionally solid for forecast of COVID-19 cases in different countries as well as India. The current share of worldwide COVID-19 confirmed cases has been predicted by taking the world population, and a comparative study has been done on COVID-19 total case growth for the top 10 worst-affected countries including US and excluding US. The ratio between confirmed cases vs. fatalities of COVID-19 is predicted, and in the end, a special study has been done on India where the authors have forecasted all the age groups affected by COVID-19. Then they have extended the study to forecast the active, death, and recovered cases in India and compared the situation with other countries.

KEYWORDS

COVID-19, Pandemic Forecasting, Random Forest Algorithm, Regression Model

INTRODUCTION

In late December 2019, a novel coronavirus family (COVID-19) was identified by the World Health Organization (WHO) with a high degree of transmission in Wuhan, China. COVID-19 is realized by the contamination known as Severe Acute Respiratory Syndrome coronavirus 2 (SARS-CoV-2) developed by the Coronavirus Study Group of the International Committee on Taxonomy of Viruses (ICTV) (Gorbalenya et al., 2020). The starting point of this infection isnot yet affirmed; however, a sequence-based investigation recommends bats as a potential key repository (Shereen et al., 2020;Cui et al., 2019). The novel COVID-19 has been a genuine danger to public health. Strict measuresare being taken, for example, isolate, social-distancing, quarantine, physical distancing, community observation, to limit the COVID-19 episode by late February. A few sorts of COVID-19 can influence

*Corresponding Author

animals, and now and then, on uncommon events, COVID-19 spread animal species into the human population. The novel COVID-19 may have spread from an animal into a human. An ongoing report has indicated that once the COVID-19 flare-up begins, it will take about a month to overpower the medicinal services framework. When the clinic limit gets overwhelmed, the demise rate bounces (Conghy et al., 2020). In healthcare, an exact forecast is fundamental for the proficient administration of patients while organizing care to the needier. To help in anticipation, a few forecast models have been created utilizing different techniques and tools, including machine learning (Chen et al., 2017; Tseng et al., 2019). In the present work, the regression model is used to forecast the total number of active deaths, recovered cases around the globe in COVID-19. Predicting the total number of COVID-19 infected patients recovered is part of the dataset (https://www.kaggle.com/c/COVID19-global-forecastingweek-5/data). Deaths have occurred in several hotspot countries, including China, the United States, the United Kingdom, Italy, Germany, and others. We have addressed the current proportion of worldwide COVID-19 confirmed cases by taking the world population. We have done comparison research on COVID-19 total cases increase for the top 10 worst impacted nations, including the US and excluding the US.The ratio between confirmed cases vs fatalities of COVID-19 is predicted. The main objectives of this paper are:

- Covid-19 affected Indian patient data analysis.
- Forecast all the age groups affected by COVID-19.
- Analysis to forecast the active, death, and recovered cases, especially in India.
- Compare the situation with other countries around the globe.

BACKGROUND

Machine learning is a field of man-made consciousness where PCs recreate the procedures of human insight and can blend complex data from enormous information sources in a brief time (Benke et al., 2018). Though there have been a couple of expectation instruments on COVID-19, just a bunch have used machine learning (Wynants et al., 2020). As far as we could know, by a wide margin, there is no openly accessible COVID-19 visualization forecast model or device from everybody of affirmed cases utilizing AI. We endeavor to apply AI on the freely accessible COVID-19 information at the network level from South Korea to foresee mortality. Li et al. (2020)used a Gaussian distribution to examine COVID-19 transmission in Hubei Province of China and created forecasts on the epidemic patterns in South Korea, Italy, and Iran; the outcomes show the advancement of the scourge, finding that monumental controls would have a considerable effect. Fanelli et al. (2020)built up a numerical model comprising susceptible-infected-recovered-deaths (SIRDs) to break down the outbreak of the COVID-19 disease in China, Italy, and France, and the outcomes describe that the recovery rate in the nations is the equivalent. In contrast, the infection and mortality rates appear to be different. Petropoulos et al. (2020)introduced an objective approach to predict the continuation of worldwide COVID-19 utilizing models from the exponential smoothing family; the outcomes give a forecast timeline for appropriate planning and decision making. In the previous few months, scientists have created or utilized existing mathematical and statistical strategies to anticipate the quantity of COVID-19 cases and their related results. The fractional time delay dynamic system (FTDD) reflects great forecast understandings of the available information (Chen et al., 2020). The generalized logistic model (GLM) demonstrates that scourge development was exponential in china (Roosa et al., 2020). In light of the forecast, the circumstance will be exacerbated in Europe, and the US will become the epicenter of new cases during the mid of April 2020 (Kumar et al., 2020).Ahmadi et al. (2020)analyzed the prediction of conclusive instances of COVID-19 in Iran through numerical models, and the outcomes show that, via

completing consistency and public behavior interventions, it is possible to control and lessen the COVID-19 epidemic from April 28 to July 2020 in Iran through the Gompertz model. Then again, among computational models to simulate and predict non-linear behaviors, the inverse Artificial Neural Network (ANNi) has as of late exceeded expectations. In writing, the ANNi computational model has simulated and predicted values of interest in engineering fields. As announced by Márquez-Nolasco et al. (2020),created two inverse artificial neural network models to simulate and predict the temperature of a thermal absorber. The outcomes show a decent simulation of the process (R2>0.99). The estimation made it possible to get higher qualities both in the internal heat and outlet temperature as for those obtained experimentally. Researchers across the world have conducted various studies to estimate the possible impact of COVID-19. Kumar et al. (2020)have used the ARIMA model to forecast the outbreak in the top 15 countries. Dehning et al. (2020)used Bayesian inferences with epidemiological parameters to analyze the time dependence of the effective growth rate of new COVID-19 infections in Germany. For the epidemic curve, Yang et al. 2020used the modified SEIR model. They used a 2003 SARS data training artificial intelligence method to forecast the outbreak. Sujatha et al. 2020 proposed a system that could be useful in predicting the extent of COVID-2019 by using vector autoregression, Multilayer Perceptron, and linear regression to forecast the epidemiological trend of the infection and the rate of COVID-2019 cases in India. Bhatnagar et al. 2020 presented a numerical model for projecting COVID-19 distribution in nations using several factors, and their model was evaluated using actual country data.The major studies include stochastic simulations (Boldog et al., 2020), Weibull distribution model (Akhmetzhanov et al., 2020), exponential growth model (Eubank et al., 2004),lognormal distribution (Lauer et al., 2020) and others Liu et al. (2020).The studies were able to predict an average incubation period of 5.1 days and a total of 14 days quarantine necessary for analyzing the virus within a person (Boldog et al. 2020). But none of these studies could estimate the exact reproduction rate, and hence not much has been done to predict how the virus will affect in the coming weeks. Also, all of the studies have been done on China, and not much work has been done concerning the Indian sub-continent.

METHODOLOGY

Regression Model

Regression analysis is a type of predictive modeling method which researches the connection between a dependent (target) and independent variable(s) (indicator). This procedure is utilized for forecasting, time series demonstrating, and finding the causal effect between the factors. Regression analysis is a powerful instrument for displaying and dissecting information.

Here, we fit a curve/line to the data points so that the differences between the distances of data points from the curve or line are minimized.

There are multiple benefits of using regression analysis. They are as follows:

1. It shows the significant relationships between dependent and independent variables.
2. It shows how strong the relationship between multiple independent variables with one dependent variable is.

Regression analysis also permits us to consider the impacts of factors estimated on various scales, such as value changes and the number of limited-time exercises. These advantages help economic researchers dispose of and assess the best arrangement of factors to build predictive models.

RANDOM FOREST ALGORITHM

A Random Forest is an outfit procedure for performing regression and classification using various decision trees and a strategy called Bootstrap and Aggregation, regularly known as Bagging. The essential thought behind this is to consolidate numerous decision trees in deciding the final output instead of depending on singular decision trees. Random Forest has different Decision Trees as base learning models. We arbitrarily perform row sampling and feature sampling from the dataset shaping simple datasets for each model. This part is called Bootstrap.

- **Ensemble Techniques:** An Ensemble technique is a strategy that unites the desires from various Machine Learning calculations to make more accurate estimates than any individual model. A model incorporatingdifferent models is called an Ensemble model.
- **Bootstrap Aggregation (Bagging):** The term "bootstrap" refers to irregular sampling with substitution. Bootstrap allows us to comprehend the bias and variance in the dataset more likely.Bootstrap includes random examining of a small subset of information from thedataset. A general system can be utilized to decrease the fluctuation for that algorithm with significant change, ordinarily decision trees. Bagging makes each model run freely and afterwards totals the output towards the end without bias to any model. Random forest is a Supervised Learning algorithm that utilizes ensemble learning techniques for classification and regression. It works by developing a massive number of decision trees at preparing time and yielding the class withmostcategories (classification) or means prediction (regression) of the individual trees.

A random forest is a meta-estimator (for example, it joins the consequence of numerous forecasts) which totals multiple decision trees, with some accommodating adjustments:

1. The number of features that can be split at each node is constrained to some aggregate level (the hyperparameter). This guarantees the ensemble model doesnot depend too heavily on any individual component and utilizes all conceivably prescient features.
2. Each tree draws a random feature from the original dataset while creating its splits, including a different component of irregularity that prevents overfitting:

Random forest = Decision Tree(base learner) + bagging (Row sampling with arrangements) + feature bagging (column sampling) + aggregation (average/median, maximum votes)

The following steps indicate the working of the Random Forest algorithm.

Random Forest Algorithm

Step 1: Select random samples from a dataset.
Step 2: Here, we construct a decision tree for every sample and then collect the predications from each decision tree.
Step 3: Here, voting is performed for every prediction result.
Step 4: At last, the highest voted result is considered to be the final result.

The above algorithm implies that, at first, test features and rules of each randomly created decision tree are used to forecast the results and store the targeted outcome. Then, the votes of each predicted target are calculated. The highly voted estimated target is then used as the final forecast of the random forest algorithm.

Mathematical Representation of Random Forest Algorithm

Gini Index

$$Gini = 1 - \sum_j p_j^2$$

To compute *Gini* impurity for a set of items with j classes, assume I belong to {1, 2, 3…. J}, let P_i be the fraction of items labeled with class i in the set, and P_j is the probability of occurrence of a feature. Figure 1 depicts a random forest working model:

$$I_G(p) = \sum_{i=1}^{j} p_i \sum_{k \neq i} p_k = \sum_{i=1}^{j} p_i (1 - p_i) = \sum_{i=1}^{j} (p_i - p_i^2) = \sum_{i=1}^{j} p_i - \sum_{i=1}^{j} p_i^2 = 1 - \sum_{i=1}^{j} p_i^2$$

Random Forest Model Description
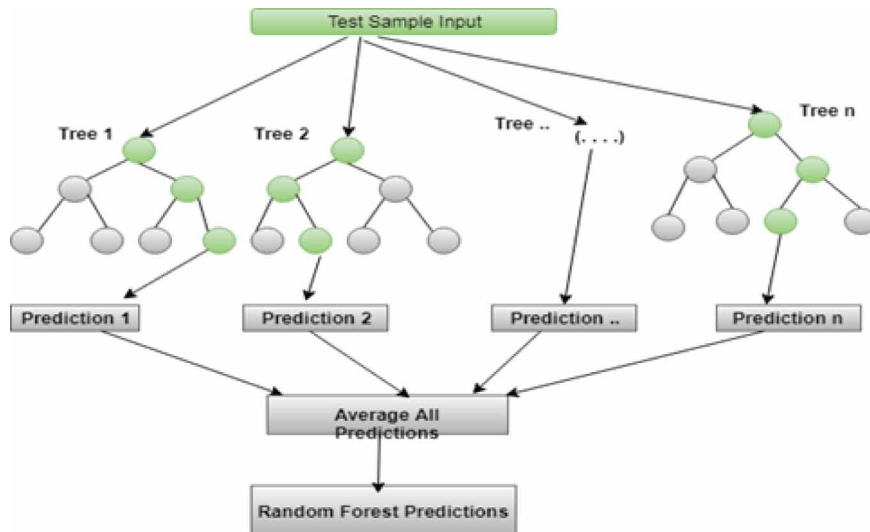
N_jobs = 3 #Number of jobs run in parallel

This parameter tells that how many processors are being used in parallel to perform a job. It is an integer that indicates the number of processors. Here -1 represents all processors:

N_estimators = 200 #Number of trees in the forest

N_estimatorsrepresent the number of trees being used. Since we are using Ensemble Technique, we need to control the number of trees used, which this parameter does:

R2 Score = 95.35%

Figure 1. Working Model of Random Forest

Mean squared error = 3688.70

Features used to predict: Population, Weight, Date, and Target.

DATASET DESCRIPTIONS

The method we have applied for analysis is the COVID-19 dataset collected fromrelevant sources(https://www.kaggle.com/c/COVID19-global-forecasting-week-5/data).Itcontains the day-by-day confirmed cases worldwide from January 23, 2020, to May 18, 2020, as appeared in Table 1. We utilized the whole dataset for training the model. The dataset consists of a total of 969641 columns and eight rows. The COVID-19 dataset includes seven attributes. The attributes are id, prov_state, country_region, confirmed case, fatalities case, weight, and population as it's a time series data having applied several methods for analysis purposes. Time series is a sequence of information that describes the period of each value. Generally, time-series data is used for analysis and forecasting the future based on historical data. Time series data determines the stability of a situation over time and efficiency portfolios. Time series datasets are time-dependent because values for every period are affected by outside factors and by values of the past period. During the dataset loading operation, we took the date as our index column. As a result, the date column was no longer a feature for us because we would do all tasks related to the date in time series data. It is the most often utilized parameter in our technique (Table 1).
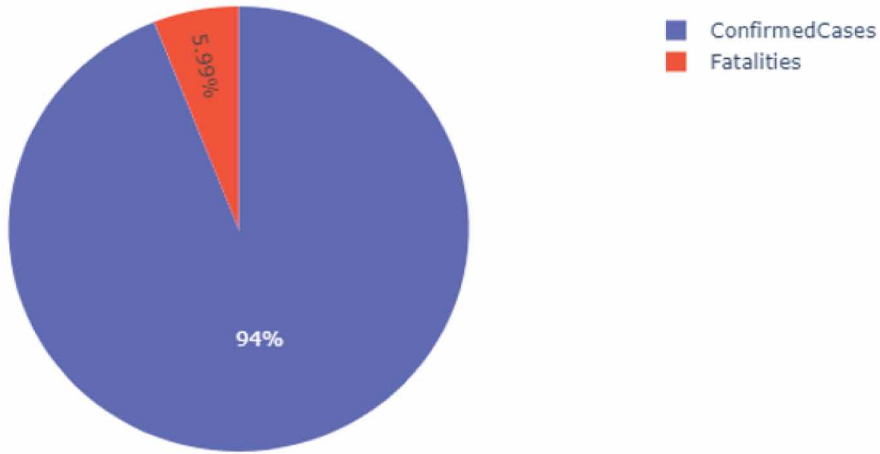
RESULTS AND DISCUSSION

The pie chart in Figure 2 represents the Confirmed case vs Fatalities case ratio of COVID-19 all over the globe; out of the total number of people infected by COVID-19, only 5% to 6%people are dead, which is very low comparatively.

Table 1. Dataset of the COVID-19

County	Province_State	Country_Region	Population	Weight	Date	Target
Autauga	Alabama	U.S.	55869	0.091485	1/23/2020	ConfirmedCases
Autauga	Alabama	US.	55869	0.914848	1/23/2020	Fatalities
Autauga	Alabama	U.S.	55869	0.091485	1/24/2020	ConfirmedCases
Autauga	Alabama	US.	55869	0.914848	1/24/2020	Fatalities
Autauga	Alabama	U.S.	55869	0.091485	1/25/2020	ConfirmedCases
Autauga	Alabama	US.	55869	0.914848	1/25/2020	Fatalities
Autauga	Alabama	U.S.	55869	0.091485	1/26/2020	ConfirmedCases
Autauga	Alabama	US.	55869	0.914848	1/26/2020	Fatalities
Autauga	Alabama	U.S.	55869	0.091485	1/27/2020	ConfirmedCases
Autauga	Alabama	US.	55869	0.914848	1/27/2020	Fatalities
Autauga	Alabama	U.S.	55869	0.091485	1/28/2020	ConfirmedCases
Autauga	Alabama	US.	55869	0.914848	1/28/2020	Fatalities
Autauga	Alabama	U.S.	55869	0.091485	1/29/2020	ConfirmedCases
Autauga	Alabama	US.	55869	0.914848	1/29/2020	Fatalities
Autauga	Alabama	U.S.	55869	0.091485	1/30/2020	ConfirmedCases

Figure 2. Confirmed case vs. Fatalities ratio

A comparison of the total number of cases in different world regions is depicted in Fig. 3. UShas been a highly affected country, i.e., 56.3%, more than half of cases. All the rest of the countries have very few cases as compared to the US. Other than the US, Russia 3.45%, the United Kingdom 3.32%, Brazil 3.2%, Spain 3.17%, Italy 3.03% countries lead the other countries. In Fig.4, a monthly comparison of COVID-19 cases growthfor the top 10 worst affected countries is represented. Here we can observe that up to the middle of March, all the countries had an almost equal number of cases, but after that, the number of cases in the US started increasing rapidly and went far beyond every other country.

In Fig. 5, the population is used as a constraint. When compared to the COVID-19 impacted nations, it can be seen that China and India have populations of 1.2 billion and 1 billion, respectively. However, they are still not that many people affected. Thus the proportion of people affected is lower.
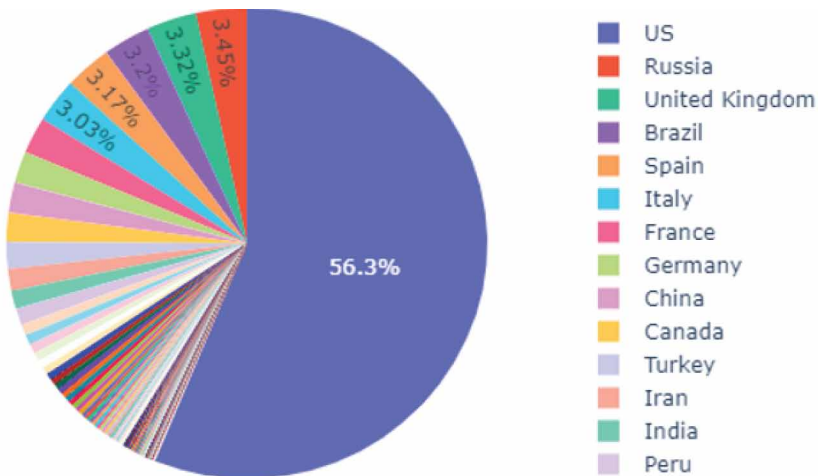
Figure 3. Region wise case comparison

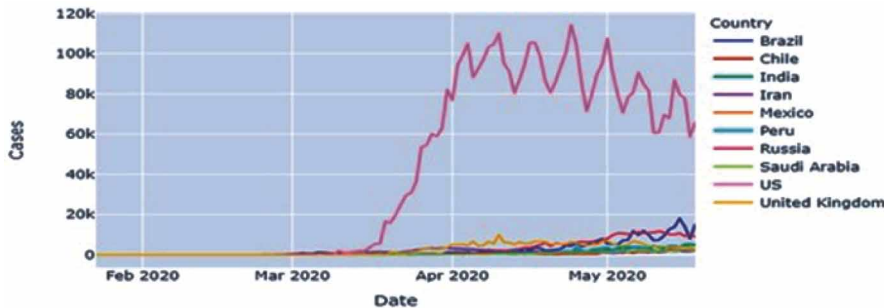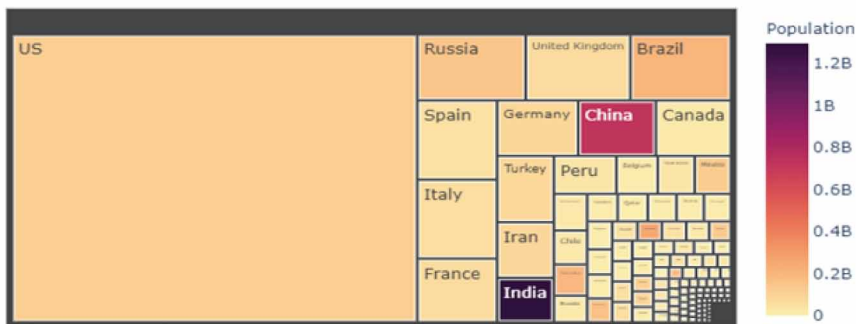**Figure 4. COVID-19 total cases growth for top 10 worst affected countries**

Figure 5. Current Shared of worldwide covid-19 confirmed cases from January 23, 2020, to May 18, 2020

On the other hand, the United States, which occupies about half of the continent and has over 0.2 billion people, is more affected.

Fig 6 shows the condition of the top ten affected countries as of May 15, 2020, and India. The green bar represents Active cases, the red bar represents Death cases, and the gray bar represents Recovered cases. China is the epicenter of COVID-19; the outbreak began on February 1, 2020, increasing active cases. However, by mid-February, the active patients had recovered from their illnesses, and the death of patients had begun. However, within a few days, the recovered cases outnumber the active cases, indicating no more active instances. When the epidemic broke out in China, individuals fled the country and became infected in neighbouring nations. As can be seen in the graphs of other nations, COVID-19 begins to influence other countries around the middle of February.

In the US, UK, Italy, France, and India, the situation seems to be worst and active cases are increasing rapidly from the first week of March. The recovery rate is meager comparatively. Still, after some time, the recovery rate surpasses the active cases. But the death cases increase rapidly. But, the active cases start later in March for Iran, Turkey, South Korea, and Germany, and the people got cured, and the death case is also less compared to other countries.In India, the condition is moderate. Though the cases start pretty late, it increases rapidly. The recovery rate is quite good, and the death rate is meager.

The gender distribution of COVID 19 patients from throughout the world is depicted in Fig.7. 81% of the cases are unknown. The COVID-19 pandemic affects 6.3% of females and 12.7% of males. Males are more severely affected than women when missing values are excluded, and the affected male percentage is 66.8%.

Figure 6. Active, Recovered, Deaths cases in Hotspot Countries and India as of May 15
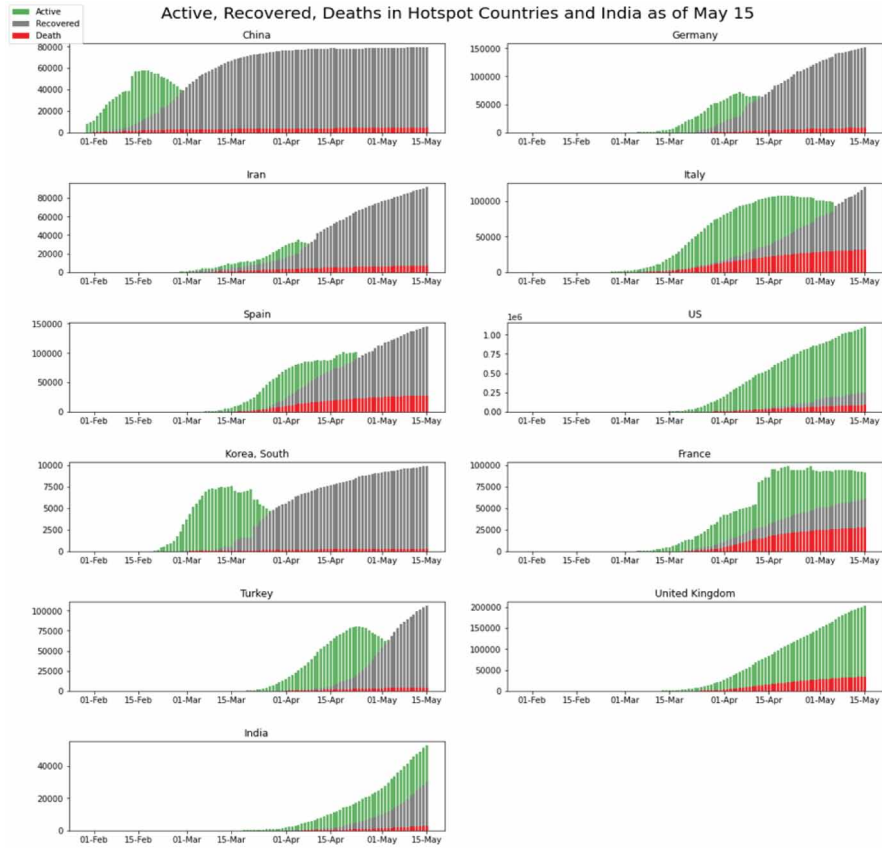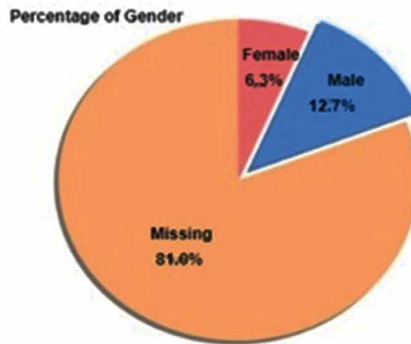

Figure 7. Percentage of Gender distribution

IMPACT OF COVID-19 IN INDIA

Fig 8 shows the age-wise distribution and percentage cases of COVID-19 cases in India. The age groups of 20-29 are highly likely to get infected with COVID-19, i.e., 24.9%, followed by the age group 30-39,40-49 and 60-69 age with 21.1%, 16.2% and 12.9%, respectively, as India's 60% of the population belongs to the age group of 20-50 year. Hence, it affects the youth of India.

Figure 8. Age group-wise distribution in India
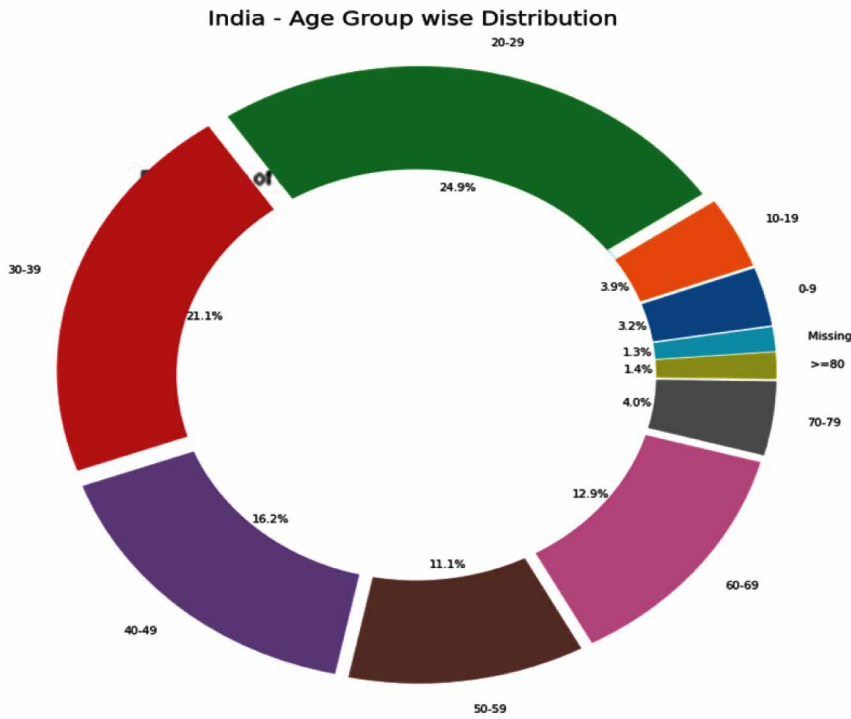

India - Age Group wise Distribution

Fig 9 shows the trend of confirmed cases, deaths, and recoveries in India. The blue spike, i.e., confirmed cases quickly growing, is predicted to exceed the 80,000 confirmed cases by the first week of May. In the case of India, the total number of recovered cases, indicated by the green spike, is relatively high, meaning that the recovery rate is entirely adequate.The trend of deaths due to COVID19 is shallow per day, which is represented in the red spike line. Suppose the total death case is compared with the complete active case in India. In that case, it comes to around almost 3-4%, which is a positive sign for India because it's the 2nd largest populated country globally.

Fig 10 represents the comparison of India with some other countries of the world. Initially, India was doing quite well compared to other countries, but when time increased, the condition deteriorated.

CONCLUSION

In this study, we forecast the total number of COVID-19 infected cases, recovered and death cases in some of the hotspot countries around the world like China, the US, UK, Italy, Germany, etc. Countries such as China, the US, and the UK were the most affected by COVID-19. Then our research exposed the distribution percentage of age groups all around the globe in terms of male and female where affected male rate dominated the female ratio, i.e., males affected percentage is 66.8% and females are 32.2%. By taking the population as a constraint, we have shown the current trend of worldwide COVID-19 confirmed cases where the US shared almost 50% with a smaller number of populations than India and China.A comparative study on COVID-19 total cases growth for the top 10 highly affected countries is discussed. A specific study was conducted in India, where we anticipated all age groups affected by COVID-19. According to our findings, the age group 20–59 years old was more affected than the other age groups.We have extended our study to forecast the active, death, and recovered cases, especially in India, and compared the situation with other countries around

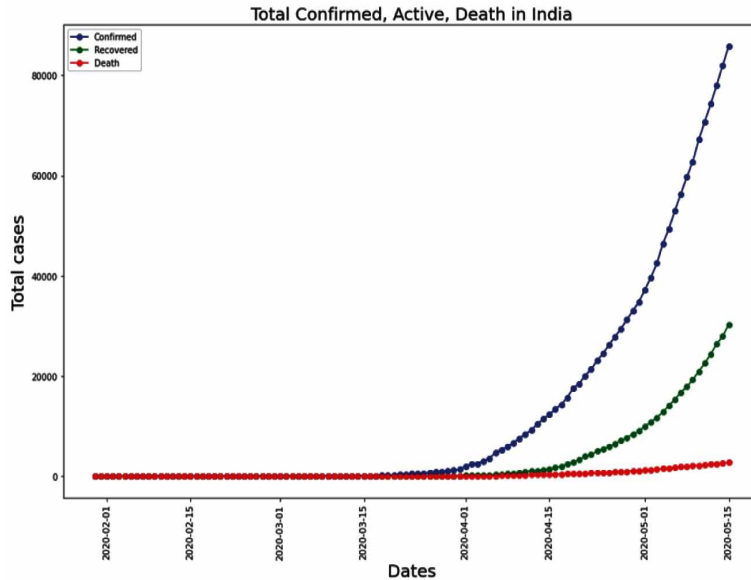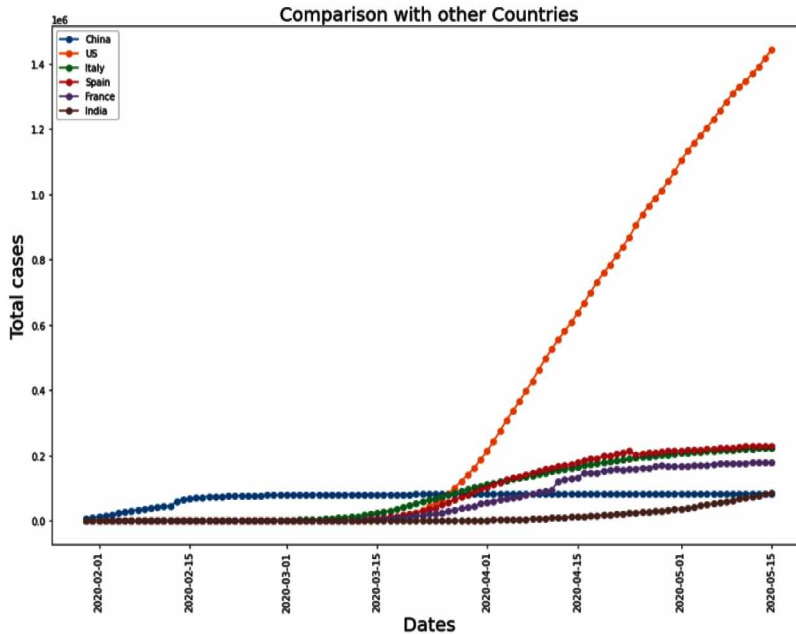**Figure 9. Total confirmed, Active and Death case in India**

Figure 10. Comparison of India with other countries

the globe. Even though the overall accuracy in predicting COVID-19 cases was perfect, we are still trying to implement a forecasting model using different algorithms and models that could give us higher precision. We likewise want to get a solitary standard model that can be utilized for any nation, mixing various algorithms.

REFERENCES

Ahmadi, A., Fadaei, Y., Shirani, M., &Rahmani, F. (2020). Modeling and forecasting trend of COVID-19 epidemic in Iran until May 13, 2020. *Medical Journal of the Islamic Republic of Iran, 34*, 27.

Bhatnagar, M. R. (2020). *COVID-19: Mathematical modeling and predictions*. ResearchGate., 10.

Boldog, P., Tekeli, T., Vizi, Z., Dénes, A., Bartha, F. A., & Röst, G. (2020). Risk assessment of novel coronavirus COVID-19 outbreaks outside China. *Journal of Clinical Medicine*, *9*(2), 571. doi:10.3390/jcm9020571 PMID:32093043

Chen, J. H., & Asch, S. M. (2017). Machine learning and prediction in medicine—Beyond the peak of inflated expectations. *The New England Journal of Medicine*, *376*(26), 2507–2509. doi:10.1056/NEJMp1702071 PMID:28657867

Chen, Y., Cheng, J., Jiang, X., & Xu, X. (2020). *The reconstruction and prediction algorithm of the fractional TDD for the local outbreak of COVID-19*. arXiv preprint arXiv:2002.10302.

Conaghy, T., Pon, B., & Anderson, E. (2020). *When does hospital capacity get overwhelmed in USA*. https://medium.com/@trentmc0/when-does-hospital-capacity-get-overwhelmed-in-usa-germany-a06cf2835f89

Cui, J., Li, F., & Shi, Z. L. (2019). Origin and evolution of pathogenic coronaviruses. *Nature Reviews. Microbiology*, *17*(3), 181–192. doi:10.1038/s41579-018-0118-9 PMID:30531947

Dehning, J., Zierenberg, J., Spitzner, F. P., Wibral, M., Neto, J. P., Wilczek, M., & Priesemann, V. (2020).Inferring COVID-19 spreading rates and potential change points for case number forecasts.medRxiv.

Eubank, S., Guclu, H., Kumar, V. A., Marathe, M. V., Srinivasan, A., Toroczkai, Z., & Wang, N. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, *429*(6988), 180–184. doi:10.1038/nature02541 PMID:15141212

Fanelli, D., & Piazza, F. (2020). Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos, Solitons, and Fractals*, *134*, 109761. doi:10.1016/j.chaos.2020.109761 PMID:32308258

Gorbalenya, A. E., Baker, S. C., Baric, R. S., De Groot, R. J., Drosten, C., Gulyaeva, A. A., & Ziebuhr, J. et al. (2020). The species Severe acute respiratory syndrome-related coronavirus: Classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology*, *5*(4), 536–544. doi:10.1038/s41564-020-0695-z PMID:32123347

Kumar, P., Kalita, H., Patairiya, S., Sharma, Y. D., Nanda, C., Rani, M., . . . Bhagavathula, A. S. (2020). Forecasting the dynamics of COVID-19 pandemic in top 15 countries in April 2020: ARIMA model with machine learning approach.MedRxiv. 10.1101/2020.03.30.20046227

Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., Azman, A. S., Reich, N. G., & Lessler, J. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Annals of Internal Medicine*, *172*(9), 577–582. doi:10.7326/M20-0504 PMID:32150748

Li, L., Yang, Z., Dang, Z., Meng, C., Huang, J., Meng, H., Wang, D., Chen, G., Zhang, J., Peng, H., & Shao, Y. (2020). Propagation analysis and prediction of the COVID-19. *Infectious Disease Modelling*, *5*, 282–292. doi:10.1016/j.idm.2020.03.002 PMID:32292868

LiuT.HuJ.KangM.LinL.ZhongH.XiaoJ.HeG.SongT.HuangQ.RongZ.DengA.ZengW.TanX.ZengS.ZhuZ.LiJ.WanD.LuJ.DengH.MaW. (2020). Transmission Dynamics of 2019 Novel Coronavirus (2019-nCoV). 10.2139/ssrn.3526307

Márquez-Nolasco, A., Conde-Gutiérrez, R. A., Hernández, J. A., Huicochea, A., Siqueiros, J., & Pérez, O. R. (2018). Optimization and estimation of the thermal energy of an absorber with graphite disks by using direct and inverse neural network. *Journal of Energy Resources Technology*, *140*(2), 020906. doi:10.1115/1.4036544

Petropoulos, F., & Makridakis, S. (2020). Forecasting the novel coronavirus COVID-19. *PLoS One*, *15*(3), e0231236. doi:10.1371/journal.pone.0231236 PMID:32231392

Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J. M., Yan, P., & Chowell, G. B. (2020). Real-time forecasts of the COVID-19 epidemic in China from February 5 to February 24, 2020. *Infectious Disease Modelling*, *5*, 256–263. doi:10.1016/j.idm.2020.02.002 PMID:32110742

Shereen, M. A., Khan, S., Kazmi, A., Bashir, N., & Siddique, R. (2020). COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. *Journal of Advanced Research*, *24*, 91–98. doi:10.1016/j.jare.2020.03.005 PMID:32257431

Sujath, R., Chatterjee, J. M., & Hassanien, A. E. (2020). A machine learning forecasting model for COVID-19 pandemic in India. *Stochastic Environmental Research and Risk Assessment*, *34*(7), 959–972. doi:10.1007/s00477-020-01827-8 PMID:32837309

Tseng, Y. J., Huang, C. E., Wen, C. N., Lai, P. Y., Wu, M. H., Sun, Y. C., Wang, H.-Y., & Lu, J. J. (2019). Predicting breast cancer metastasis by using serum biomarkers and clinicopathological data with machine learning technologies. *International Journal of Medical Informatics*, *128*, 79–86. doi:10.1016/j.ijmedinf.2019.05.003 PMID:31103449

Wynants, L., Van Calster, B., Collins, G. S., Riley, R. D., Heinze, G., Schuit, E., ... van Smeden, M. (2020). Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal. *BMJ*, 369.

Yang, Z., Zeng, Z., Wang, K., Wong, S. S., Liang, W., Zanin, M., Liu, P., Cao, X., Gao, Z., Mai, Z., Liang, J., Liu, X., Li, S., Li, Y., Ye, F., Guan, W., Yang, Y., Li, F., Luo, S., & He, J. et al. (2020). Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease*, *12*(3), 165–174. doi:10.21037/jtd.2020.02.64 PMID:32274081

Debabrata Dansana is working as an Assistant Professor, in the Department of Computer Science, Kalinga Institute of Social Sciences (KISS) Deemed to be University, Bhubaneswar, Odisha. His researches areas are Cognitive Radio Ad-Hoc Networks, WSN, IoT, Machine Learning, and Deep Learning.

Raghvendra Kumar is working as Associate Professor in Computer Science and Engineering Department at GIET University, India. He received B. Tech, M.Tech and Ph.D. in Computer Science and Engineering, India, and Postdoc Fellow from Institute of Information Technology, Virtual Reality and Multimedia, Vietnam. He serves as Series Editor Internet of Everything (IOE): Security and Privacy Paradigm, Green Engineering and Technology: Concepts and Applications, publishes by CRC press, Taylor & Francis Group, USA, and Bio-Medical Engineering: Techniques and Applications, Publishes by Apple Academic Press, CRC Press, Taylor & Francis Group, USA. He also serves as acquisition editor for Computer Science by Apple Academic Press, CRC Press, Taylor & Francis Group, USA. He has published number of research papers in international journal (SCI/SCIE/ESCI/Scopus) and conferences including IEEE and Springer as well as serve as organizing chair (RICE-2019, 2020), volume Editor (RICE-2018), Keynote speaker, session chair, Co-chair, publicity chair, publication chair, advisory board, Technical program Committee members in many international and national conferences and serve as guest editors in many special issues from reputed journals (Indexed By: Scopus, ESCI, SCI). He also published 13 chapters in edited book published by IGI Global, Springer and Elsevier. His researches areas are Computer Networks, Data Mining, cloud computing and Secure Multiparty Computations, Theory of Computer Science and Design of Algorithms. He authored and Edited 23 computer science books in field of Internet of Things, Data Mining, Biomedical Engineering, Big Data, Robotics, and IGI Global Publication, USA, IOS Press Netherland, Springer, Elsevier, CRC Press, USA.

Chandrakanta Mahanty is a research scholar of the Department of Computer Science and Engineering, GIET University, Gunupur. He received his MTech in Computer Science and Engineering, India.