

Data Analysis and Prediction of COVID-19 Using Machine Learning Models

¹Dr. E. Laxmilydia,²Dr.Pavan Kumar,³Vydarsi Sitarama Prasad,⁴Chinmaya
Ranjanpattanaik

¹Professor, Department of Computer Science and Engineering, Vignan's Institute of Information Technology (A), Visakhapatnam (Andhra Pradesh), India.

²department of Computer Science Engineering, IFHE, HYDERABAD.

³Assistant professor in department of Computer Science Engineering, Vignan's Institute of Information Technology for Women, Andhra Pradesh, India.

⁴Department of computer science and engineering, AjayBinay Institute of Technology, Cuttack, Odisha, India

ABSTRACT

Breakout of novel corona virus from China got spread to various countries all over the world. The symptoms of the affected patients were hard to be sorted through self defection. The rapid increase of the virus cases spread through air, physical touch, and direct interaction from one person to another and to things. COVID-19 begins with respiratory illness which also includes fever, sore throat, cough and water from eyes. Increase of patients at various places, this paper identifies prediction model to predict and analyze COVID-19 virus using machine learning Time series forecasting prediction models such as Holt's Linear Model, Holt's winter Model, Prophets Predictions. By comparing the worst affected countries those reside beside India using machine learning prediction models from kaggle Coronavirus COVID-19 repository dataset, we can predict the rapid changes in India.

Keywords: Corona virus, COVID-19, Time series forecast models, Machine Learning

I INTRODUCTION

Overlap of symptoms through respiratory virus were collected from ARIs (Acute Respiratory virus illnesses) form those who suffer with severe throat infection, feverish body temperature. In severe conditions, this may lead to death. After several testing of virus, scientists, doctors have identified that this virus has faster transmission compared to other viruses. It is identified that above 60 age group people with diabetes, renal failure, and chronic lung disease were at high risk. COVID-19 patient has to be undergone diagnosis by collecting blood samples. Blood samples were tested for pneumonia, sepsis which has bacteria (RT-PCR testing of nasopharyngeal and oropharyngeal) and also samples from lower respiratory tract. Suspected patients were tested in the laboratory. Identified positive cases were isolated and diagnosed.

Precautions before reaching COVID-19 virus affected people were carried out at different stages such as droplet precautions, contact precautions, airborne precautions. At the initial stage of the patient, triple layer facial masks were provided to avoid direct contact with others. Minimum one meter distance is maintained from contact and instructs to take care of cough and sneezing which comes from respiratory secretions. Patient will be undergone through self quarantine or under medical treatment at hospital for clinical diagnosis depending upon the epidemiological risk factor to avoid droplet transmission of respiratory viruses [3]. Respiratory droplets such as coarse aerosols, fine-particle aerosols transfer quickly through direct contact. For the direct interaction of the patient, Personal Protective Equipment such as gloves, mask, eye protection, gown are very essential. Avoid touching of surfaces at patients surroundings and maintaining continuous hand hygiene is mandatory.

II LITERATURE SURVEY

COVID-19 was also known as Wuhan virus, where its roots got spread all over world. It's global panic situation of virus spreading from one infected person to multiple people have many complicated

symptoms that comes from pneumonia, acute respiratory distress syndrome, and in some cases symptoms were not observed within the infected patients were spotted and kept in observations. COVID-19 cases vary in different mild forms and symptoms were identified in between 2 to 14 days. [9] Uncertain situations implement common models for respiratory modes of transmission to identify COVID-19. [8] The transmission routes of respiratory droplets and aerosol were primary focus to transfer Coronavirus and other respiratory viruses, infections such as influenza, rhinoviruses exhaled through our breath. [4][5] Virus particles from influenza are captured through Gesundheit-II. They result breath coarse particles with Teflon impactor under $>5\mu\text{m}$, other particles gathered into buffer i.e., aerosols. The transmission can be avoided and reduced through facial masks. [6]

Patient with high fever with more than or equal to 38 degrees of temperature lasting from 10 days are tested for severe acute respiratory infection and requested to be hospitalized before getting any further delay. Diagnosis and treatment of COVID-19 clinical syndromes, case classification result with proper result to implement prevention, control measures for infections. Before starting antimicrobial therapy, they were tested by collecting blood samples and identify bacteria related to pneumonia, sepsis and also nasopharyngeal, oropharyngeal to start antimicrobial therapy. Patients were treated till complete cure on testing and discharge process is performed. [7]

Importance of facial surgical masks and its usage was prescribed by World Health Organization. Symptoms that exist more than a week was identified as worsen situation and sent to have clinical consultations having chest X-ray, Lung USG. In case of cardiac failure were excluded with objective assessment i.e., ECHO. India is the least worsen country which got affected and increased rapidly with 1.7% at the beginning of COVID-19 [11].

Similar to Covid-19, spread of an epidemic disease in 1897 more number of countries and most of the union territories has shaken people to shut down all educational institutions, commercial business, cancelled visas to all tourists. [12]. The first step to take serious action by Narendra Modi on COVID-19 was India Janta Curfew on 22nd March 2020 from 7am to 9pm. Depending on the severity of COVID-19, Government of India declared lockdown for seventy five districts for COVID-19 and it's been extending till date (April 2020) [13]. He suggested every Indian citizen to work from home and stay safe at home. In rare cases, Government has granted permission to use Lopinavir or Ritonavir. Statewise economic response was to the government to address the nation on COVID-19 information and increase of positive in every state was listed in Wikipedia and other Indian COVID-19 portals. [10]

Antiviral drug for nucleotide analogs i.e., Remdesivir was invented to fight against virus such as Ebola and Marburg [14]. Remdesivir drug was also applied for the cure of other diseases. This was first given to the United states patient lately in January 2020 who was identified as pneumonia patient [15] with SARS-COV-2. As patient got cured slowly, the same Remdesivir drug was treated for other patients declared by WHO for the launch of four-arm pragmatic clinical trial. [16]

The rapid increase of positive cases from China provided patients with randomized clinical trials and observed that Remdesivir drug is most effective for the quick recovery of COVID-19. Multiple Harvard-affiliated hospitals in United States have undergone survey on Remdesivir drug. They carried their work to predict the patients who were treated with this drug and concluded that it is safe to use and generate it in large amounts. [18]

In case of Prophylaxis, Indian Council of Medical Research, the national task force for COVID-19 has suggested the importance of hydroxychloroquine. This can be given to the patients who were at high risk i.e., Prophylaxis of SARS-CoV-2. [19]

Indian Medical council identified that Antiviral activity of Prophylaxis was faster than nasopharyngeal and provide 400mg of hydroxychloroquine to patient twice a day, later once a week. [1] Shortage of Chloroquine for malaria-endemic countries such as India was pointed to worst situation of facing spread of covid-19 ten million cases. Health intervention approaches to control Coronavirus through mathematical model based techniques. [2]

To promote the pH as it is weak base for membrane fusion possible medical approach is provided to intracellular organelles. SARS-CoV-2 virus infection from different stages, a crucial stage of acidification for function and endosome maturation, intermediate phases transformations were blocked to summarize endocytosis. [20]

Indian Government and doctors have kept their maximum effort to fight against COVID-19 by spreading awareness and precautions to all citizens.[17]

III METHODOLOGY

Machine learning models in search of correlations within large volumes of data among variables independently. After large number of emerging reviews, the subtle correlations were also be detected based on the data points. Focusing on the trained dataset, similar type of unlabeled data files were merged and made inferences using various algorithms [21-30]. Following are the three implemented time series forecasting prediction models:

Time series forecasting model Prediction using Holt's Linear Model: Data with high variations mostly use Holt's linear model among several existing models. Time series forecast decompose dataset data according to the seasonality, residual and trend. Forecasting is allowed to extend exponential smoothing in simpler manner using Holt approach to level series and trend. To obtain expected forecast, the entire process is carried out using three equations at each level and trend and for the combination of level and trend i.e, forecast equation. Equations were generated to forecast equation and also to achieve multiplicative instead of merging. With respect to results linear increase or decrease, additive equation is applied. Exponential decrease in trend, multiplicative equation is applied.

Time series forecasting model Prediction using Holt's Winter Model: Similar patterns of data with constant intervals of time described in the datasets varying seasonality. This model is the best approach when compared to other forecasting time series models because of its seasonality factor. It comprises of three smoothing equations similar to Holt's linear model but differs with its seasonality factor. Smoothing parameters ranges in between $0 \leq \alpha \leq 1, 0 \leq \beta \leq 1, 0 \leq \gamma \leq 1$. The seasonal index and current index draws the weighted average for the seasonal equation of the same season.

Time series forecasting model Prediction using Prophets Prediction: The regressive components, average components from the primary input parameters are automated. Intuitive parameters are simple to tune through Prophet Package. Dealing with variety of problems in forecasting models draws meaningful predictions use Prophet Predictions. It is a decomposable time series model with balances the changes of periodic, non-periodic, seasonality, holidays information. It is the resultant of logistic growth $g(t)$, weekly and yearly seasonality $s(t)$, user data at irregular schedules i.e., holidays $h(t)$, unusual changes that generate error terms accommodated through the models (ϵt). Represented by $y(t) = g(t) + s(t) + h(t) + \epsilon t$
Linear and non-linear functions at different time components use prophet prediction model. Prophet Prediction is similar to Holt-Winters technique that uses exponential smoothing for Additive component Modeling seasonality.

Testing Strategy for Indian People:

People who traveled from other countries to India,

- Initially, requested to take 14 days home quarantine and kept under observation
- They were observed for primary symptoms of fever, difficulty in breathing.
- If they come across the primary symptoms of COVID-19, blood samples will be collected and tested.
- Positive report in test will be treated as COVID-19 patient and treated accordingly by isolating.

People with severe acute respiratory distress were tested in same procedure. Following figure 1 is the flow diagram to discharge COVID-19 patient:

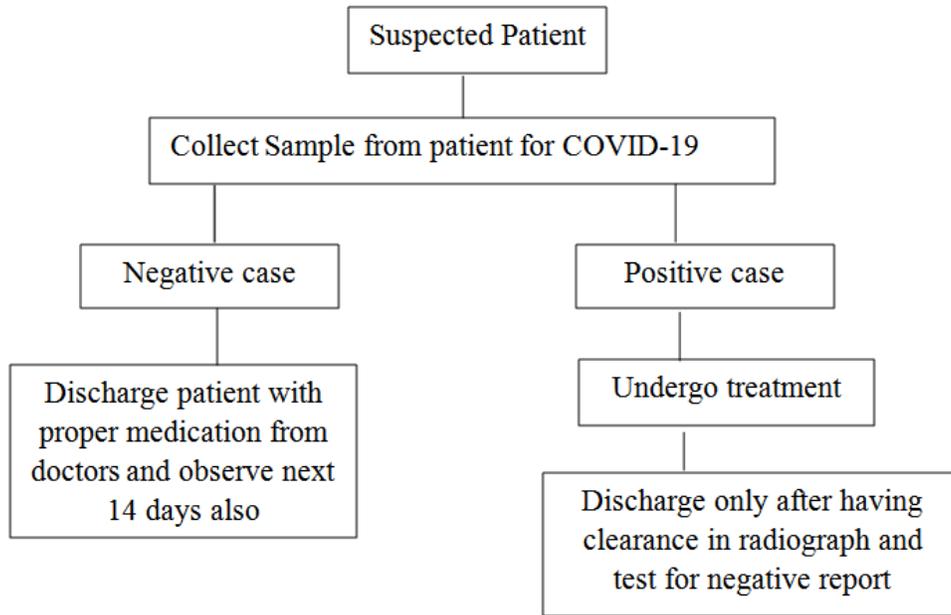


Fig1: Flowchart for COVID-19 patient

IV RESULT ANALYSIS

This paper work is carried out using Kaggle novel-corona-virus-2019 (nCoV) dataset, which consists of different countries COVID-19 statistics such as confirmed cases, deaths, recovered cases basing on observation date, province. Country wise dataset were available. Indian dataset is compared with neighboring dataset for the analysis and prediction of COVID-19.

Figure 2 describes the COVID-19 rapid increases of confirmed positive cases, recovered cases and death cases in India from February to May 2020. Figure 3 describes the COVID-19 mean Recovery, mean morality rate of India February to May 2020. Figure 4 describes 14 weeks of COVID-19 confirmed positive cases from India and other neighboring countries. Figure 5 describes 14 weeks of COVID-19 confirmed death cases from India and other neighboring countries.

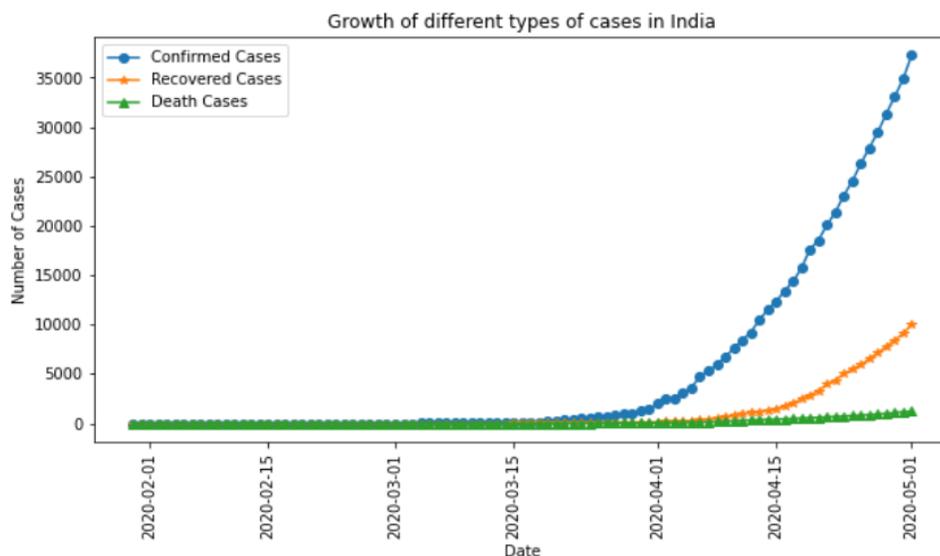


Fig2: Graph representation of COVID-19 cases in India

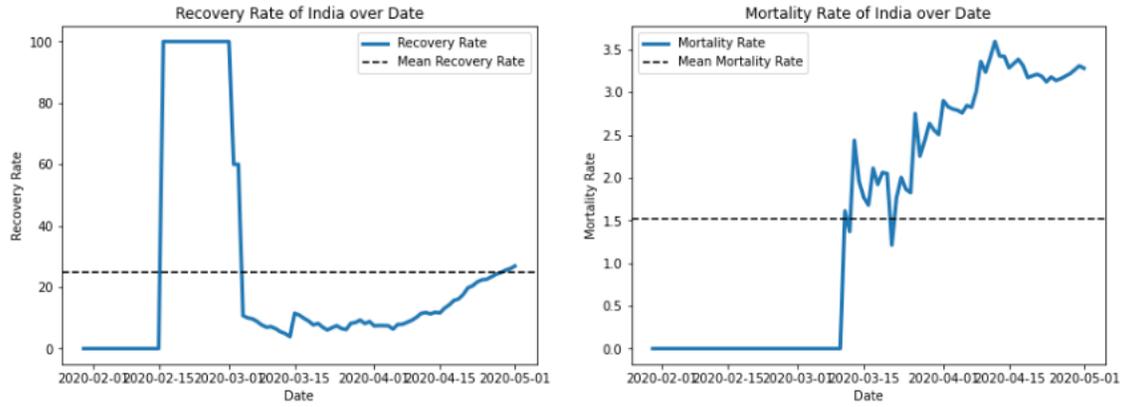


Fig3: Graph representation of COVID-19 Recovery, morality rate of India

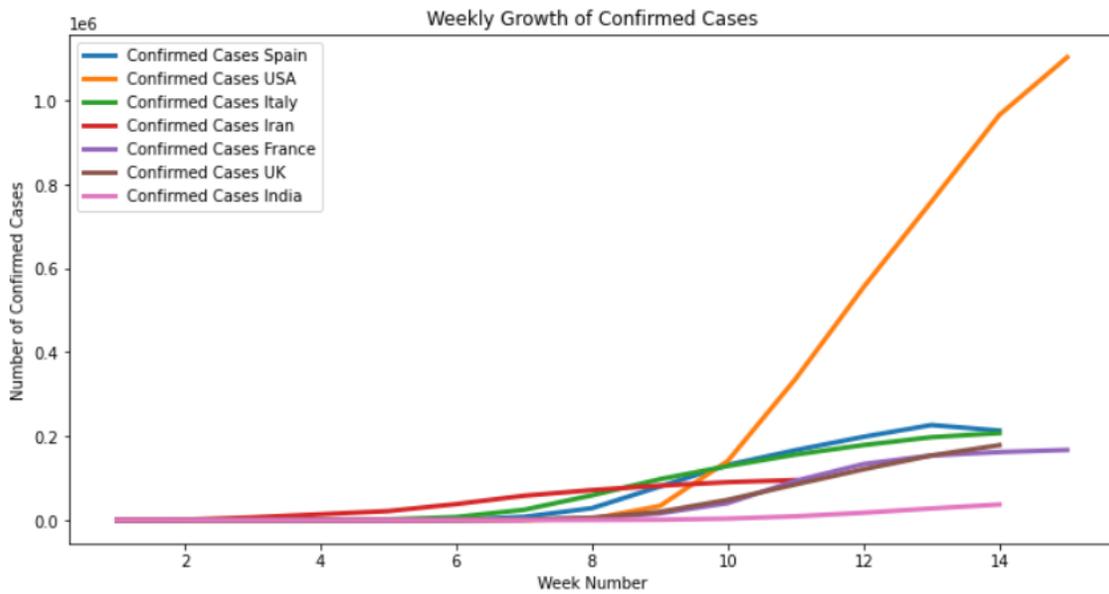


Fig5:Graph representation of COVID-19 confirmed positive cases from India and other neighboring countries weekly

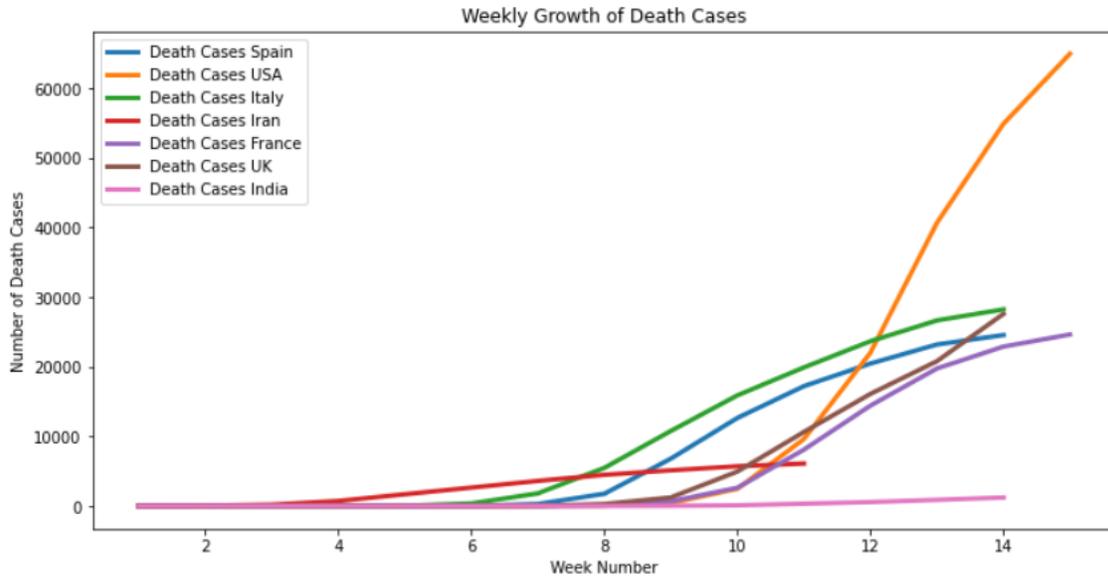


Fig6: Graph representation of COVID-19 death cases from India and other neighboring countries weekly

V CONCLUSION

India being a Republic country, populous democratic country, seventh-largest country in area with 1,387,297,452 estimated population. Virus originated from China COVID-19, India implementing lockdown showing its “unity in diversity policy” to fight against COVID-19. Government has divided every state and district into red, orange, green Zones. This paper concentrates on data analysis and rapid increase and drop of the COVID-19 predications. Machine learning prediction models such as Holt’s linear, winter model and Prophets Prediction were applied to forecast the confirmed positive, recovered and death cases. Based on the weekly increase of the cases, neighboring countries of India were compared and forecasted. Results were tested considering kaggle novel-corona-virus-2019 (nCoV) dataset.

REFERENCES

1. Gautret P, Lagier J-C, Parola P, et al. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *Int J Antimicrob Agents* 2020; DOI: 10.1016/j.ijantimicag.
2. Mandal S, Bhatnagar T, Arinaminpathy N, et al. Prudent public health intervention strategies to control the Coronavirus disease 2019 transmission in India: a mathematical model- based approach. *Indian J Med Res* 2020. DOI:10.4103/ijmr.IJMR_504_20.
3. Shiu, E.Y.C., Leung, N. H. L. & Cowling, B.J. Controversy around airborne versus droplet transmission of respiratory viruses: implication for infection prevention. *Curr. Opin. Infect. Dis.* 32, 372-379, 2019.
4. Tellier, R., Li, Y., Cowling, B. J. & Tang, J. W. Recognition of aerosol transmission of infections agents: a commentary. *BMC Infect.Dis.* 19, 101(2019).
5. Xiao, J. et al. Nonpharmaceutical measures for pandemic influenza in nonhealthcare settings-personal protective and environmental measures. *Emerg. Infect. Dis.* <https://doi.org/10.3201/eid2605.190994> (2020)
6. Yan, J. et al. Infectious virus in exhaled breath of symptomatic seasonal influenza cases from a college community. *Proc. Natl Acad.Sci. USA* 115, 1081-1086 (2018).
7. Cowling, B. J. & Leung, G. M. Epidemiological research priorities for public health control of the ongoing global novel Coronavirus (2019-nCoV) outbreak. *Euro Surveill*, <https://doi.org/10.2807/1560-7917.ES.2020.25.6.2000110> (2020).
8. Han, Q., Lin, Q., Ni, Z. & You, L. Uncertainties about the transmission routes of 2019 novel Coronavirus, *Influenza other Respir. Viruses* <https://doi.org/10.1111/irv.12735> (2020)

9. Bradley M. the Essential guide to the Wuhan Virus(Symptoms, transmission and Prevention). Corona Virus; 2020.
10. Guidelines on clinical management of COVID-19. Available from: <https://www.mohfw.gov.in/pdf/GuidelinesonClinicalManagementof COVID1912020.pd>
11. Sinha A. One COVID-19 positive infects 1.7 in India, lower than in hot zones. The Indian Express; 23 March 2020. Available from: <https://indianexpress.com/article/coronavirus/coronavirus-india-infection-rate-china-6321154/>
12. Sanyal A. "India Suspends all tourist Visas till april 15 over Coronavirus: 10 Facts". <https://www.ndtv.com/india-news/coronavirus-impact-visas-to-india-suspended-till-april-15-2193382>.
13. Hebbar, Nistula. Coronavirus: Union government announces lockdown in 75 districts till march 31. The Hindu; 2020. <https://www.thehindu.com/news/national/coronavirus-india-locks-down-80-districts-to-contain-virus-spread/article31134567.ece>
14. Warren TK, Jordan R, Lo MK, Ray AS, Mackman RL, Soloveva V, et al. "Therapeutic efficacy of the small molecule GS-5734 against Ebola virus in rhesus monkeys". *Nature* 2016;531(7594): 381-5
15. Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. (Washington state 2019-nCoV Case Investigation Team). "First Case of 2019 Novel Coronavirus in the United States". *The N Eng J of Med* 2020; 382(10):929-36.
16. Harmon A. "Inside the Race to contain America's First Coronavirus case ". The New York Times Feb 2020.
17. Kachroo V. Novel Coronavirus (COVID-19) in India: current scenario. *International Journal of Research and review.* 2020; 7(3):435-447.
18. Haung, C. et al. Clinical features of patients infected with 2019 novel Coronavirus in Wuhan, China. *Lancet* 2020; 395:497-506
19. Liu J, Cao R, Xu M, Wang X, Zhang H, Hu H. Hydroxychloroquine, a less toxic derivative of chloroquine is effective in SARS-CoV2 infection in vitro. *Cell Discov* 2020; 6(16)
20. Mauthe, M. et al. Chloroquine inhibits autophagic flux by decreasing autophagosome lysosome fusion. *Autophagy* 2018; 14:1435-55
21. Shankar, K., Lakshmanprabu, S. K., Khanna, A., Tanwar, S., Rodrigues, J. J., & Roy, N. R. (2019). Alzheimer detection using Group Grey Wolf Optimization based features with convolutional classifier. *Computers & Electrical Engineering*, 77, 230-243.
22. Elhoseny, M., & Shankar, K. (2019). Optimal bilateral filter and convolutional neural network based denoising method of medical image measurements. *Measurement*, 143, 125-135.
23. Krishnaraj, N., Elhoseny, M., Thenmozhi, M., Selim, M. M., & Shankar, K. (2019). Deep learning model for real-time image compression in Internet of Underwater Things (IoUT). *Journal of Real-Time Image Processing*, 1-15.
24. Lakshmanprabu, S. K., Mohanty, S. N., Krishnamoorthy, S., Uthayakumar, J., & Shankar, K. (2019). Online clinical decision support system using optimal deep neural networks. *Applied Soft Computing*, 81, 105487.
25. Lakshmanprabu, S. K., Mohanty, S. N., Shankar, K., Arunkumar, N., & Ramirez, G. (2019). Optimal deep learning model for classification of lung cancer on CT images. *Future Generation Computer Systems*, 92, 374-382.
26. Elhoseny, M., Bian, G. B., Lakshmanprabu, S. K., Shankar, K., Singh, A. K., & Wu, W. (2019). Effective features to classify ovarian cancer data in internet of medical things. *Computer Networks*, 159, 147-156.
27. Kathiresan, S., Sait, A. R. W., Gupta, D., Lakshmanprabu, S. K., Khanna, A., & Pandey, H. M. (2020). Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model. *Pattern Recognition Letters*.
28. Iswanto, I., Lydia, E. L., Shankar, K., Nguyen, P. T., Hashim, W., & Maselena, A. (2019). Identifying diseases and diagnosis using machine learning. *International Journal of Engineering and Advanced Technology*, 8(6 Special Issue 2), 978-981.
29. Khasanah, S., Nguyen, P. T., Lydia, E. L., & Shankar, K. (2019). Exploration of Retinopathy Disease using Machine Learning Methodology. *International Journal of Engineering and Advanced Technology*, 8(6S), 914-921.
30. Sibirian, R. H. S., Karolina, R., Nguyen, P. T., Lydia, E. L., & Shankar, K. (2019). Leaf Disease Classification using Advanced SVM Algorithm. *International Journal of Engineering and Advanced Technology*, 8(6S), 712-718.