Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2659 – 2681
Received 05 March 2021; Accepted 01 April 2021.

A study on the challenges of predictive analysis on the outbreak of pandemic

Anitha Christy Angelin[1*], Sheikh Faritha Begum[2]
[1]PSNA College of Engineering and Technology, Dindigul
[2] PSNA College of Engineering and Technology, Dindigul
*anitha.angelin@yahoo.com

ABSTRACT

With a rapid outbreak of corona virus pandemic disease 2019 (COVID-19), caused by a novel severe acute respiratory syndrome corona virus 2 (SARS-CoV-2), numerous measures are taken by the countries in order to contain the primary spread of the disease. The use of IoT and cloud for collecting and analysing data to prevent the spread of the disease through mobile applications was a major challenge. The conventional centralized cloud computing used for storing the victim's data is encountering severe challenges, such as high latency and high energy consumption. Also, the possibility of occurrence of coverage hole can lead to the loss of actual data causing the system to provide erroneous information. This paper addresses a detailed review of various predictive models to identify the challenges in determining the outset of a pandemic like COVID and thereby health measures can be taken.

Keywords
Covid-19; Predictive analysis; Loss of data;Energy efficiency; Edge computing.

Introduction

A coronavirus is a kind of common virus that causes an infection in your nose, sinuses, or upper throat. At the end of 2019, a novel coronavirus was recognized as the cause of a cluster of pneumonia cases in Wuhan, a city in the Hubei Province of China. It quickly spread, resulting in an epidemic throughout China, charted an accumulative number of cases in other states all through the world [7]. World countries have used a wide range of technologies in their fight against the pandemic. However the statistical analysis in Figure 2 shows that the number of recovered cases compared with the total cases is significantly low which augments stress analysis towards finding an efficient solution towards the outbreak of the disease. Though cloud based expertise are devised, it raises questions about excessive surveillance and the violation of citizens' privacy [1, 7]. Indian government usage of Aarogya Setu app for tracking people's whereabouts through the location information provided by their phones has been crucial to identifying where an infected person went before being quarantined and how many people were in close proximity to the patient. The app in
Figure 3 collects the following personal information during registration and stores it in the cloud: (i) name; (ii) phone number; (iii) age; (iv) sex; (v) profession; (vi) countries visited in the last 30 days; and (vii) whether or not you are a smoker and a person's current medical condition collected through a series of questions when the app is run for the first time to assess the condition of theuser. Moreover, the app continuously collects the location data of the registered user and maintains a record of the places where the user had come in contact with other registered users. With the increase of cloud storage, many consumers have integrated it into their IoT networks. Some companies have provided other services to their customers other than cloud storage [2]. Apps are now available online without users needing to install them to build cloud computing.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2659 – 2681
Received 05 March 2021; Accepted 01 April 2021.

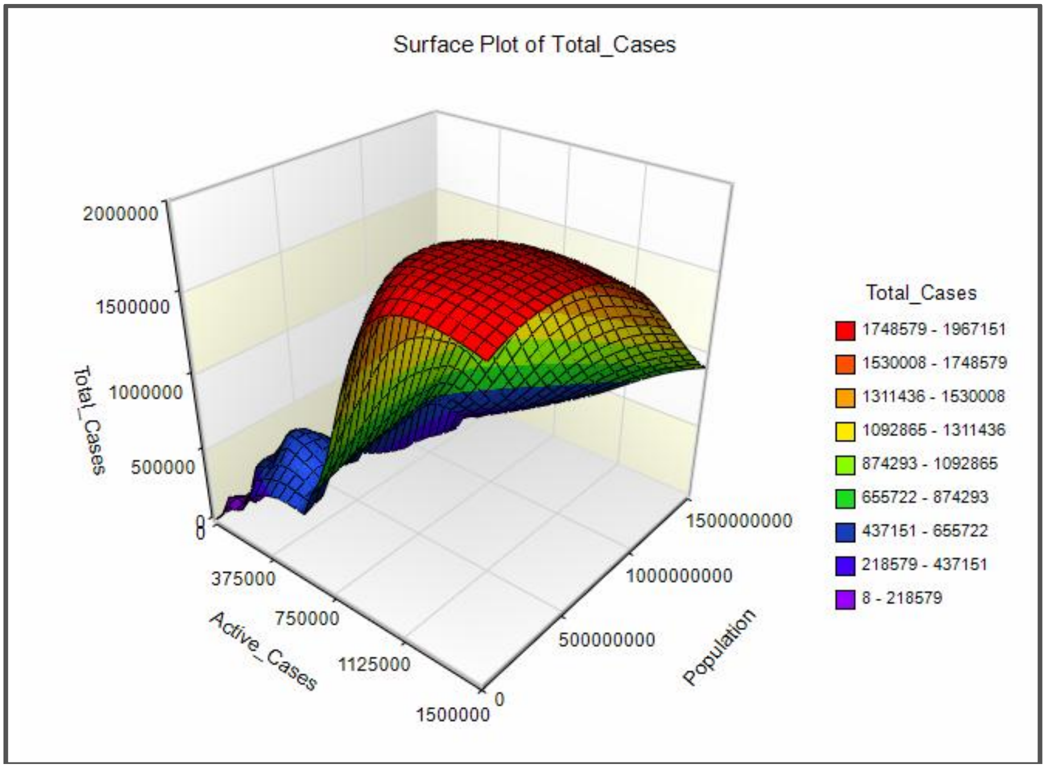


Figure 1 Surface plot of test cases in total world population

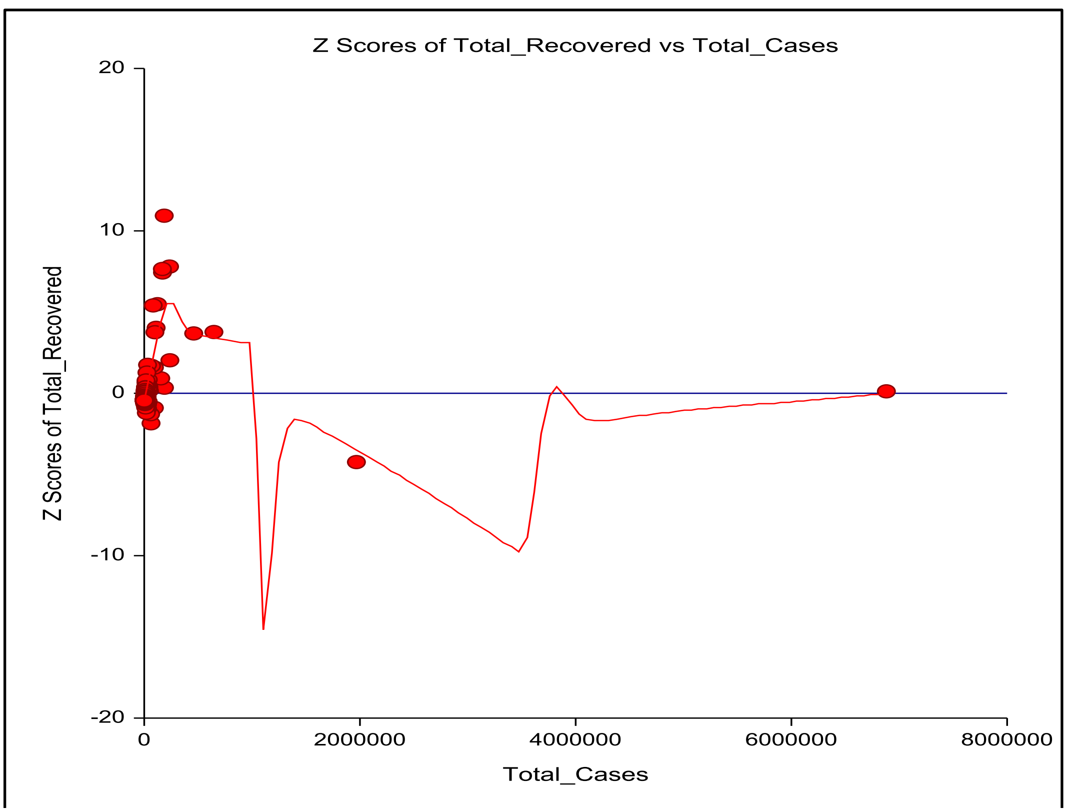


Figure 2 Z Scores of Total Recovered vs Total cases

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2659 – 2681
Received 05 March 2021; Accepted 01 April 2021.

Cloud computing is a way of remotely storing all types of data on the cloud server. Many applications use these servers to implement large-scale algorithms such as natural language translation, image processing and sound classification. As a result, cloud usage removes most of the processing strains on user devices. This makes the process too centralized, however. Then, this burden is dumped on the data centre. There is indeed a barrier that restricts the quality of the service. The server's software and hardware architecture now governs the boundaries of the study

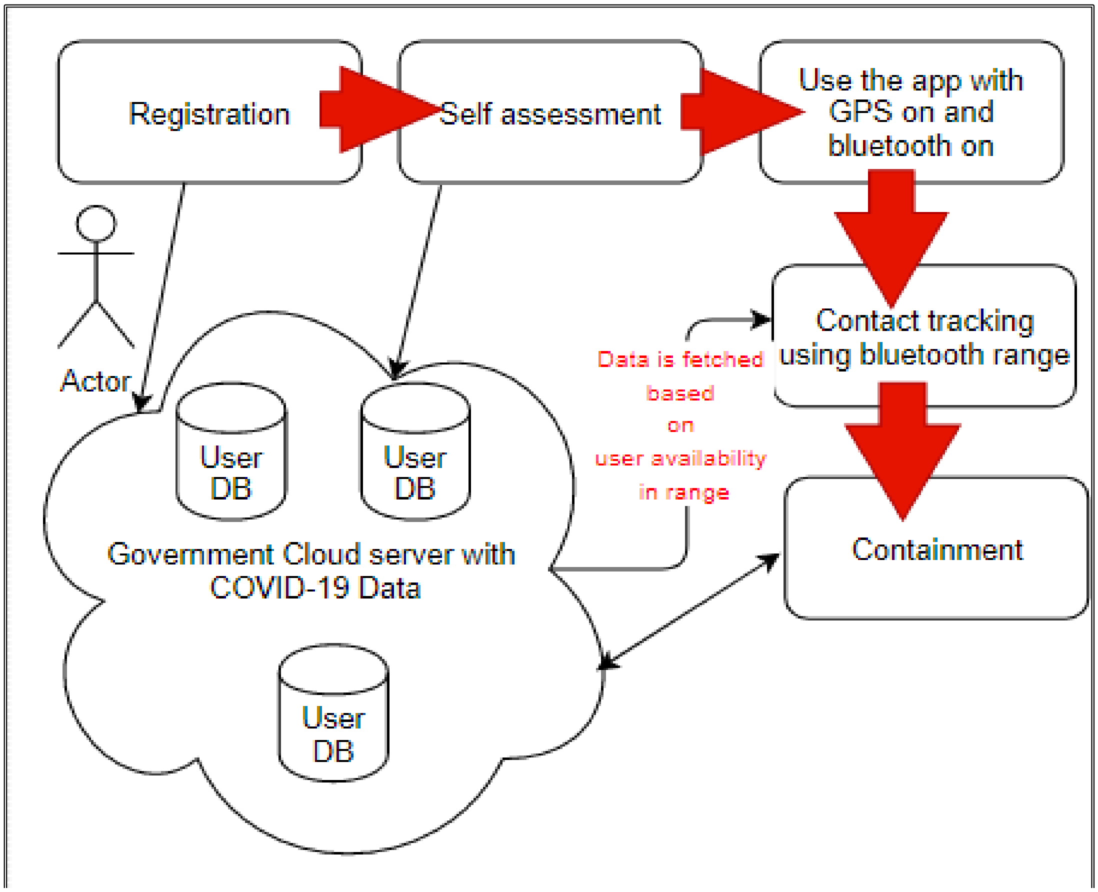


Figure 3 Cloud computing system for COVID 19 tracking and detection

of the service. The influence of users on the cloud has been shown to be dangerous. Data protection, efficiency and availability are some of the aspects that can be risky for cloud users [4]. It may be dangerous to depend too much on the cloud to manage all data processes. Again with dramatic rise in the advent of the internet of Things and 5G networks in the smart city environment, a significantamount of data is expected to grow substantially, leading to an increased latency for the conventional cloud computing technology. Despite the exponential increase in the number of mobile devices, traditional centralized cloud computing is failing to make the QoS for a variety of applications. With future 5G technologies on the horizon, EDGE computing will become the crucial solution to this situation. The Radio-Access Network (RAN) is one of the major hurdles associated with 5 G technology. Mobile edge computing offers real-time RAN data [4]. While using real-time RAN details, network providers will enhance the Quality-of-Experience (QoE) to end-users, as RAN can provide context-aware services in real-time.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2659 – 2681
Received 05 March 2021; Accepted 01 April 2021.

The new MEC architecture for storing, classifying and analysing IoT data streams was also implemented. The MEC server is responsible for managing different protocols, distributing messages and processing analytics. The MEC framework [5] creates a new supply chain and an energized system, providing new possibilities for mobile carriers and providers of software and content. EDGE computing offers a tolerable computational capacity, enough storage space, and fast response time to satisfy IoT application requirements. The rest of the paper is organized as follows. Different categories of applications of EDGE computing in healthcare is presented in Section 2. Also the predictive data analytic models are compared to analyse the recovery rate in COVID detection using cloud based approaches.

Literature Review

New trends in mobile edge computing, such as remote surveillance of patients with central cloud computing and their growing volumes of multimedia, have attracted all people in industry and government. Clear visualization, a high degree of sensing, and better service quality (QoS) are therefore the top priority [8].Numerous applications use this recent technology in order to have a id data flow with a better QoS. Static monitoring healthcare applications are used to monitor thephysiological parameters of a patient whereas dynamic monitoring healthcare applications are used to monitor the epidemic spread and location tracking of the patients.

Static Monitoring applications

Ali et al. [8] suggested that in mobile edge computing based Tele-surgery applications, the Window-based Control rate algorithm (w-RCA) will optimize medical service quality (m-QoS) by considering the Network parameters such as peak-to-mean ratio standard deviation delay and jitter in 8 min Medical Video streams called "Navigation towards the Uterine Horn, horn transection and re-anastomosis". Though w-RCA produces better and effective results at small buffer and window, it cannot provide optimized results for a long stream of data. Min et al. [9] proposed the Edge-Cognitive-Computing (ECC-based) healthcare system that can monitor and analyse users' physical health with cognitive computing. The system optimizes computing resources and improves patient survival in a sudden emergency significantly. However this system can be used to monitor the health of the patient depending upon the network availability and high latency.

Golam et al. [10] proposed portfolio-optimization and ADMM-based approaches to ensure that ECaaS and CCaaS providers are identified in a cost-effective way for analysis of the data obtained from wearable sensors in e-health or m-health applications. Though this approach cuts costs of the diagnosis of health data through the integration of the EoT framework, security and privacy cost of the data is not optimized. Pham et al. [11] proposed a cloud based smart home environment for health monitoring which is a cooperative fog to cloud architecture, in which data pre-processing, indoor location and algorithms for activity recognition are done via a home gateway while a private cloud is being used for remote data storing. This system suffers from high latency and storage issues.

Dynamic monitoring applications

Dynamic monitoring applications are becoming a viable solution in diagnosis and treatment of epidemic diseases. In these perspective, real time set-up, mobility, location awareness and smart data aggregation from different sensors are truly a difficult task. Such systems prove to be effective when diagnosing virus infections in an initial state so that proper treatment can be given

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2659 – 2681
Received 05 March 2021; Accepted 01 April 2021.

on time to enable fast recovery. Sanjay et al. [12] proposed a system to prevent and control the spread of Zika virus disease using integration of Fog computing, cloud computing, mobile phones and the IoT-based sensor devices. It is used to represent each Zika virus (ZikaV)-infected user, mosquito-dense sites and breeding sites on the Google map that help the government healthcare authorities to control such risk-prone areas effectively and efficiently. However this system cannot identify the asymptotic spread of the diseases. In order to predict COVID-19 by using multimodal AI approaches. Maghdid et al[13] proposed design studies for the reading of integrated sensors on smartphones. The ability to share location information makes intelligent IoT devices possible to maximize their owners' security.

Al-Hamadi and Chen et al. [14] discussed a trust-based information sharing approach in the IoMT framework in which data from sensors is used to derive a user health loss risk according to their vulnerabilities from entering a given location at a certain time. In both the works, sensing reports submitted by individual IoT systems are collected and analysed in a centralized cloud which actually suffers from latency and security issues.Most of the static and dynamic monitoring applications suffers from the challenges of latency [13] and data security [14]. Numerous research works are identified and implemented to tackle these issues. However, the problem of energy efficiency of the sensor devices containing the patient’s critical data has been still under research. Especially in dynamic monitoring and tracking applications, when the sensor carrying data is exhausted, it causes coverage hole [19, 20, 21] causing serious issues in healthcare applications.

Predictive analysis for COVID data

Predictive analytics models [29] are designed to assess the chronological COVID-19 case data, determine patterns, detect trends and use that information to draw up predictions about the recovery rate and the spread of the disease across the globe. One of the most popular models for predictive analytics is a forecast model. It manages the estimation of metric values by estimating the values of new data based on historical data learnings. It is also used to produce numerical values when there is none to be found in historical data. One of predictive analytics' key strengths is the ability to input several parameters. There are different types of predictive models. Classification models [26, 27] are one of the most popular models of predictive analysis. These models are focused on historical data to categorize knowledge.

Classification models are used by various industries because new data can be readjusted quickly and the questions can be addressed in a thorough analysis. Although classification and prediction models work with historical data, anomalous data entries within a dataset are used for the outliers’ model.Outlier data, as its name suggests, refers to data veering away from the norm. It works by recognizing extraordinary data, either separately or with respect to various categories and numbers. The time series model [29] concentrates on the data when the input parameter is time. The time series model is used to generate a number metric that forecasts trends in a particular timeframe e using different data points taken from previous year data. The clustering model divides the data into various classes, using similar characteristics. In some applications it is particularly useful to separate data to various datasets based on unique attributes. The analysis of data is performed based on the following models: i) Generalized Linear Model [29],ii) Deep Learning Model [28], iii) Decision Tree Model [27], iv) Random Forest Model [26], v) Support Vector Machine [25]. These models forecast the recovery rate from the epidemic in the future.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2659 – 2681
Received 05 March 2021; Accepted 01 April 2021.

Generalized Linear Model [29]

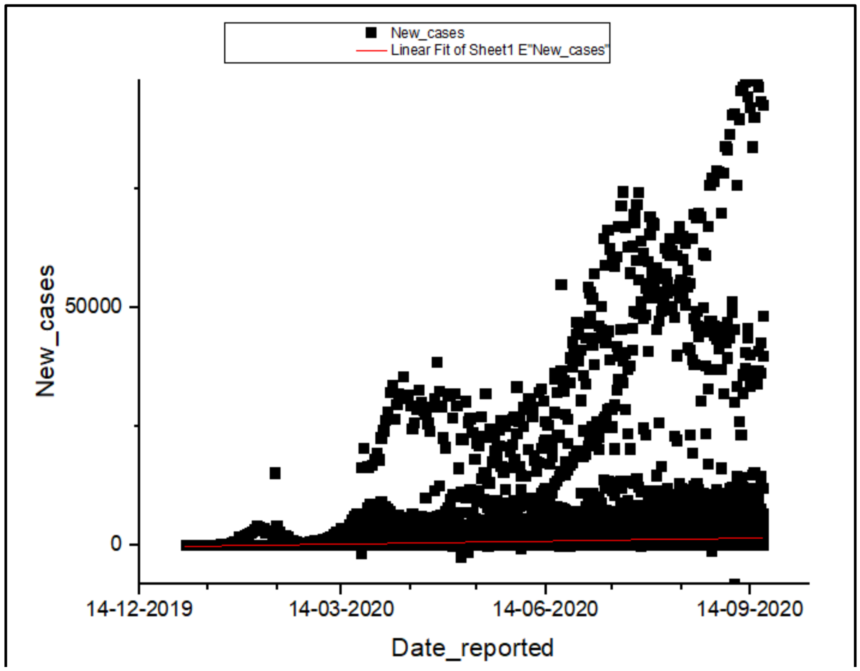


Figure 4 Linear Fit of Newly affected covid cases

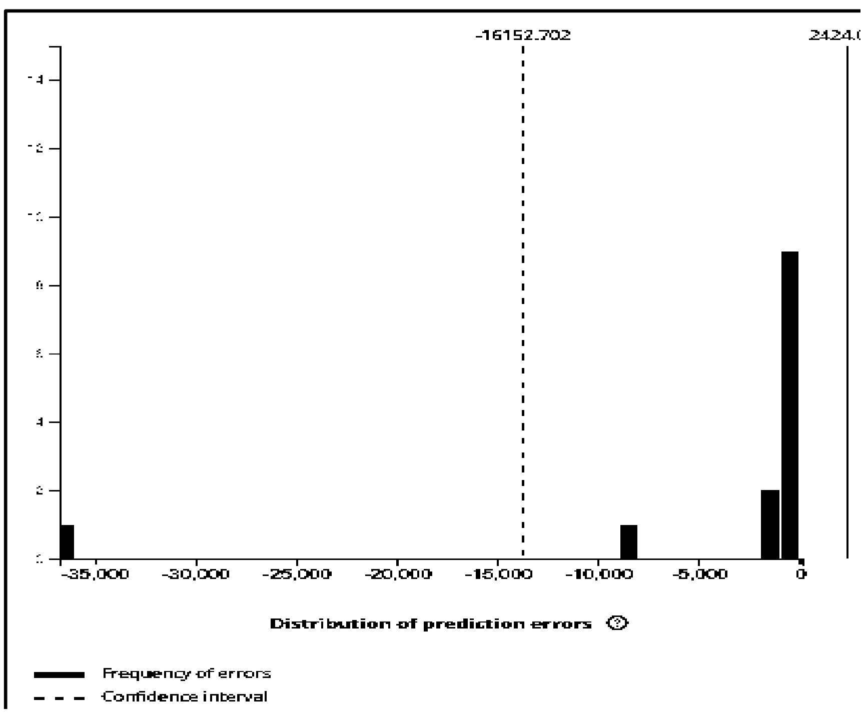


Figure 5 Prediction values of Active cases using Generalized Linear Model.

Figure 4 shows the linear fit of newly affected cases.Figure 5 shows the predictions vs the actual values. Ideally, all predictions equal the actual values and lie on a diagonal line. The closer they are to the diagonal line the better the predictions are. With 95% probability the points are

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2659 – 2681
Received 05 March 2021; Accepted 01 April 2021.

betweenthe dashed lines. With 95% probability the predictions do not deviate more than +/- 16152.702 from the actual value.

Table 1 Statistical parameters for Linear fit

Parameters	**New cases**
Number of Points	42779
Degrees of Freedom	42777
Residual Sum of Squares	8.21E+11
Pearson's r	0.09097
Adj. R-Square	0.00825

Deep Learning Model[28]

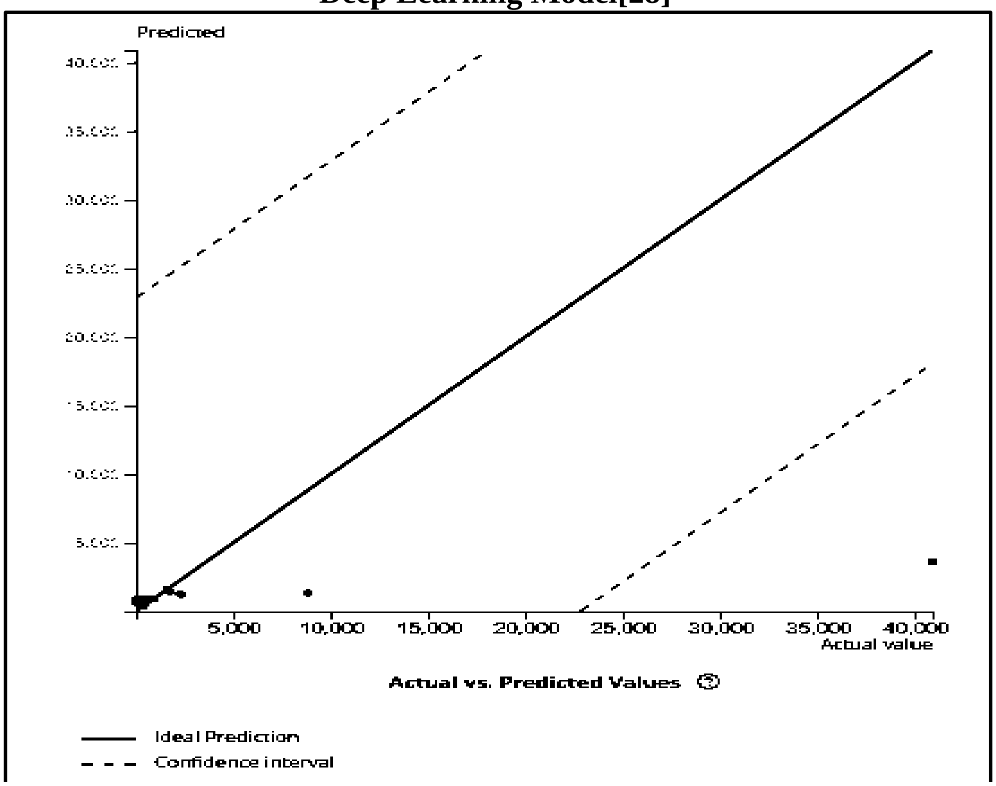


Figure 6 Prediction values of Active cases using Deep Learning Model

A multitude of promising performance is obtained through deep learning-based method specified in Figure 6 and Figure 7. For processing problems, CNNs play a major role, in which case-based reasoning, analogy mining, and cluster analysis, which in the other fields of NLP are subsumed under this classifier. With 95% probability the points are between the dashed lines. With 95% probability the predictions do not deviate more than +/- 16127.184 from the actual value.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2659 – 2681
Received 05 March 2021; Accepted 01 April 2021.

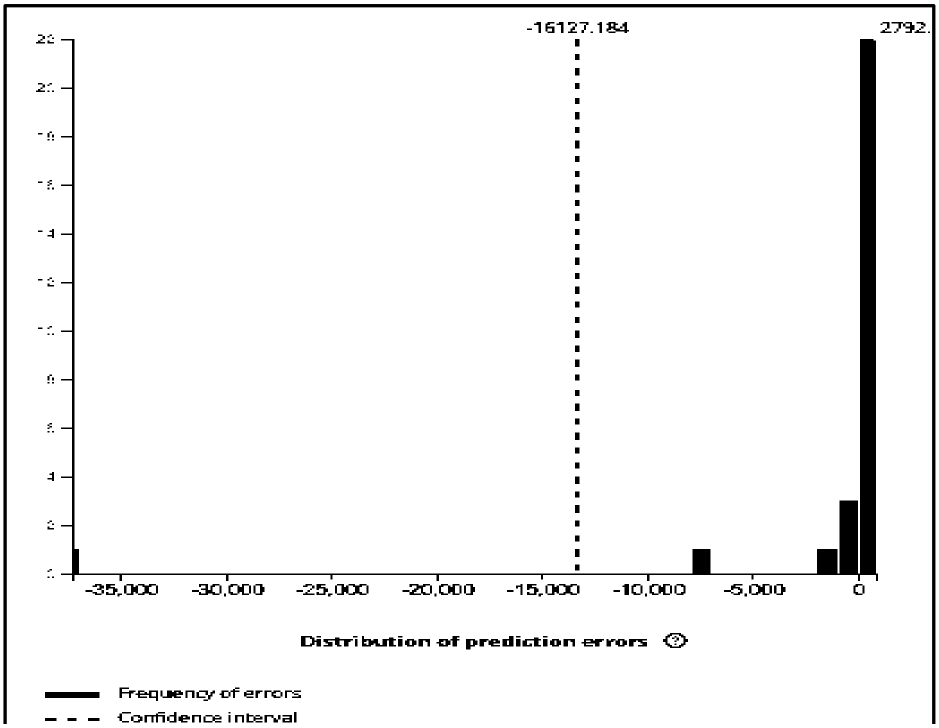


Figure 7 Prediction values of Active cases using Deep Learning Model.

Decision tree model [27]

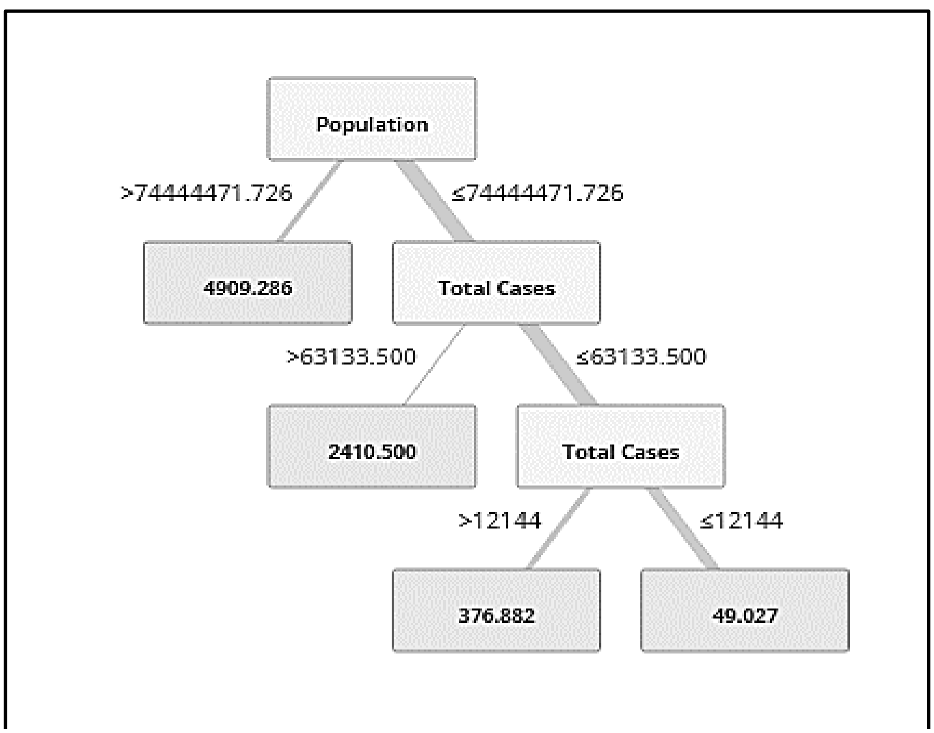


Figure 8 Decision Tree visualization of total cases

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2659 – 2681
Received 05 March 2021; Accepted 01 April 2021.

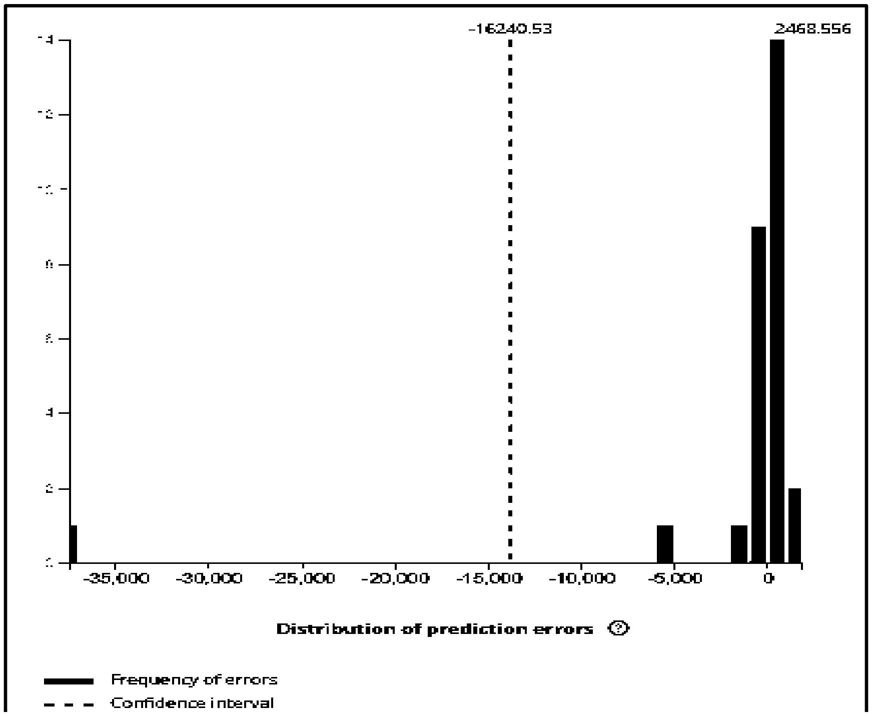
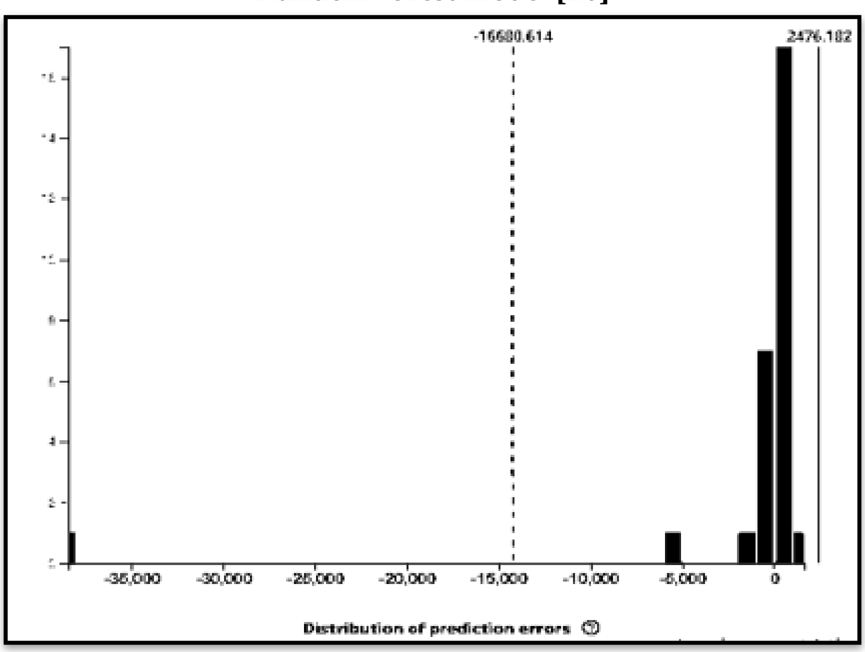


Figure 8 Prediction values of Active cases using Decision Tree Model

Figure 8shows the Decision tree and Figure 9 shows the predictions vs. the actual values. Ideally, all predictions equal the actual values and lie on a diagonal line. The closer they are to the diagonal
line the better the predictions are. With 95% probability the points are between the dashed lines. With 95% probability the predictions do not deviate more than +/- 16240.53 from the actual value.

Random forest model [26]



Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2659 – 2681
Received 05 March 2021; Accepted 01 April 2021.

Figure 9 Prediction values of Active cases using Random forest Model

Figure 9 shows the predictions vs the actual values. Ideally, all predictions equal the actual values and lie on a diagonal line. The closer they are to the diagonal line the better the predictions are. With 95% probability the points are between the dashed lines. With 95% probability the predictions do not deviate more than +/- 16681.614 from the actual value.

Support Vector Machine [25]

An SVM (Supported Vector Machine) is an impossible binary linear classification method. Its main attribute is the non-probabilistic component. This concept is opposed to probabilistic classificatory. In other words, an SVM removes knowledge from a plane. Only a small dataset (feature vectors) can assess. The data subset supporting the decision boundary is referred to as vectors of support. The data collection does not change the location of the boundary in the feature function.

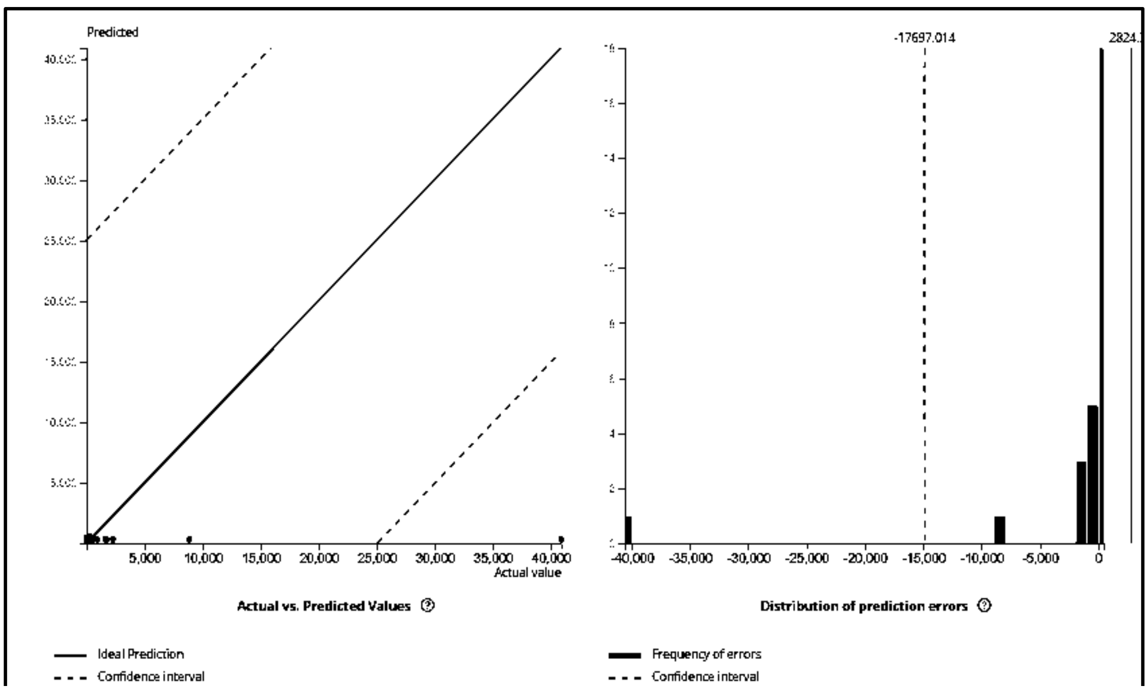


Figure 10 Prediction values of Active cases using SVM

Figure 10 shows the prediction of active cases using SVM.With 95% probability the points are between the dashed lines. With 95% probability the predictions do not deviate more than +/- 17697.014 from the actual value. The biggest challenge to predictions has been the lack of accurate surveillance data and the ability to test for the presence of virus in individuals exhibiting symptoms of COVID-19. As a result, the flow of information has been uneven, and this, in turn, makes the predictions fluctuate. Those fluctuations impact predictions on the timing of the peak as well as the magnitude. The proposed EDGE computing based COVID detection framework for coverage hole detection aims at detecting and predicting the COVID infected person’s location with data accuracy, minimum energy consumption and efficient delivery of critical data with a low latency.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2659 – 2681
Received 05 March 2021; Accepted 01 April 2021.

Conclusion

All medical devices are linked to the internet, and during any critical situation, it automatically conveys a message to the healthcare staff. Asymptotic cases can be handled properly in a remote location with integrated IoT devices. Compared to cloud services EDGE based COVID detection and prevention system seems to be an excellent way to monitor the infected patient. In recent critical times, this technology is supportive to sustain quality observation with accurate real-time data. By using a computational geometry based method, EDGE based COVID detection and prevention system is efficient to predict an upcoming situation of this disease with low latency and high energy efficiency. With appropriate execution of this capability, researchers and healthcare workers can create a healthier atmosphere to combat this disease.

References

[1] Ai, Y., Peng, M., & Zhang, K. (2018). Edge computing technologies for Internet of Things: a primer. Digital Communications and Networks. https://doi.org/10.1016/j.dcan.2017.07.001.

[2] Chen, M., Li, W., Hao, Y., Qian, Y., & Humar, I. (2018). Edge cognitive computing based smart healthcare system. Future Generation Computer Systems. https://doi.org/10.1016/j.future.2018.03.054.

[3] Hassanalieragh, M., Page, A., Soyata, T., Sharma, G., Aktas, M., Mateos, G., Kantarci, B., & Andreescu, S. (2015). Health Monitoring and Management Using Internet-of-Things (IoT) Sensing with Cloud-Based Processing: Opportunities and Challenges. Proceedings - 2015 IEEE International Conference on Services Computing, SCC 2015. https://doi.org/10.1109/SCC.2015.47.

[4] Rothan, H. A., & Byrareddy, S. N. (2020). The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. In Journal of Autoimmunity. https://doi.org/10.1016/j.jaut.2020.102433.

[5] Safavat, S., Sapavath, N. N., & Rawat, D. B. (2019). Recent advances in mobile edge computing and content caching. Digital Communications and Networks. https://doi.org/10.1016/j.dcan.2019.08.004.

[6] Sodhro, A. H., Luo, Z., Sangaiah, A. K., & Baik, S. W. (2019). Mobile edge computing based QoS optimization in medical healthcare applications. International Journal of Information Management. https://doi.org/10.1016/j.ijinfomgt.2018.08.004.

[7] Sung, W. T., & Chang, K. Y. (2013). Evidence-based multi-sensor information fusion for remote health care systems. Sensors and Actuators, A: Physical. https://doi.org/10.1016/j.sna.2013.09.034

[8] Verma, P., & Sood, S. K. (2018). Fog assisted-IoT enabled patient health monitoring in smart homes. IEEE Internet of Things Journal. https://doi.org/10.1109/JIOT.2018.2803201

[9] Yu, W., Liang, F., He, X., Hatcher, W. G., Lu, C., Lin, J., & Yang, X. (2017). A Survey on the Edge Computing for the Internet of Things. In IEEE Access. https://doi.org/10.1109/ACCESS.2017.2778504.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2659 – 2681
Received 05 March 2021; Accepted 01 April 2021.

[10] Sareen, S., Gupta, S. K., & Sood, S. K. (2017). An intelligent and secure system for predicting and preventing Zika virus outbreak using Fog computing. Enterprise Information Systems. https://doi.org/10.1080/17517575.2016.1277558.

[11] Pham, M., Mengistu, Y., Do, H., & Sheng, W. (2018). Delivering home healthcare through a Cloud-based Smart Home Environment (CoSHE). Future Generation Computer Systems. https://doi.org/10.1016/j.future.2017.10.040.

[12] Alam, M. G. R., Munir, M. S., Uddin, M. Z., Alam, M. S., Dang, T. N., & Hong, C. S. (2019). Edge-of-things computing framework for cost-effective provisioning of healthcare data. Journal of Parallel and Distributed Computing. https://doi.org/10.1016/j.jpdc.2018.08.011.

[13] Wang, X. X., Han, S., Wu, Y., & Wang, X. X. (2013). Coverage and energy consumption control in mobile heterogeneous wireless sensor networks. IEEE Transactions on Automatic Control, 58(4). https://doi.org/10.1109/TAC.2012.2225511.

[14] Fekete, S. P., Kröller, A., Pfisterer, D., Fischer, S., & Buschmann, C. (2004). Neighborhood-based topology recognition in sensor networks. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 3121. https://doi.org/10.1007/978-3-540-27820-7_12.

[15] Chao, C. M., Lu, M. W., & Lin, Y. C. (2015). Energy-Efficient Multichannel MAC Protocol Design for Bursty Data Traffic in Underwater Sensor Networks. IEEE Journal of Oceanic Engineering, 40(2). https://doi.org/10.1109/JOE.2014.2304311.

[16] Zhang, Y., Zhang, X., Wang, Z., & Liu, H. (2013). Virtual edge based coverage hole detection algorithm in wireless sensor networks. IEEE Wireless Communications and Networking Conference, WCNC. https://doi.org/10.1109/WCNC.2013.6554783.

[17] Funke, S., & Klein, C. (2006). Hole detection or: "How much geometry hides in connectivity?" Proceedings of the Annual Symposium on Computational Geometry, 2006.

[18] Kumar, P., & Lee, H. J. (2012). Security issues in healthcare applications using wireless medical sensor networks: A survey. In Sensors (Vol. 12, Issue 1). https://doi.org/10.3390/s120100055.

[19] Dhingra, R. K. S. (2015). Coverage Hole Detection in Wireless Sensor Networks. International Journal of Science and Research (IJSR), 4(6).

[20] Gupta, H. P., Rao, S. V., Yadav, A. K., & Dutta, T. (2015). Geographic routing in clustered wireless sensor networks among obstacles. IEEE Sensors Journal, 15(5). https://doi.org/10.1109/JSEN.2014.2385734.

[21] Nazrul Alam, S. M., & Haas, Z. J. (2015). Coverage and connectivity in three-dimensional networks with random node deployment. Ad Hoc Networks, 34. https://doi.org/10.1016/j.adhoc.2014.09.008.

[22] Si, P., Wu, C., Zhang, Y., Xia, Z., & Cheng, L. (2011). A hole detection and redeployment strategy in wireless sensor network. Journal of Information and Computational Science, 8(13).

[23] Khedr, A. M., Osamy, W., & Salim, A. (2018). Distributed coverage hole detection and recovery scheme for heterogeneous wireless sensor networks. Computer Communications, 124. https://doi.org/10.1016/j.comcom.2018.04.002.

[24] Li, W., & Wu, Y. (2016). Tree-based coverage hole detection and healing method in wireless sensor networks. Computer Networks, 103. https://doi.org/10.1016/j.comnet.2016.04.005.

[25] Wang, X., He, X., Feng, F., Nie, L., & Chua, T.-S. (2018). TEM: Tree-enhanced Embedding Model for Explainable Recommendation. Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW ’18, 10. https://doi.org/10.1145/3178876.3186066.

[26] Zafari, A., Zurita-Milla, R., & Izquierdo-Verdiguier, E. (2019). Evaluating the performance of a Random Forest Kernel for land cover classification. Remote Sensing, 11(5). https://doi.org/10.3390/rs11050575.

[27] Matasci, G., Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., Hobart, G. W., & Zald, H. S. J. (2018). Large-area mapping of Canadian boreal forest cover, height, biomass and other structural attributes using Landsat composites and lidar plots. Remote Sensing of Environment, 209. https://doi.org/10.1016/j.rse.2017.12.020.

[28] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15.

[29] Schaid, D. J., Tong, X., Batzler, A., Sinnwell, J. P., Qing, J., & Biernacka, J. M. (2019). Multivariate generalized linear model for genetic pleiotropy. Biostatistics, 20(1). https://doi.org/10.1093/biostatistics/kxx06.