



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

Towards applying internet of things and machine learning for the risk prediction of COVID-19 in pandemic situation using Naive Bayes classifier for improving accuracy

N. Deepa^{a,*}, J. Sathya Priya^b, Devi T.^a^a Department of Computer Science & Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India^b Department of Information Technology, Velammal Engineering College, Chennai, India

ARTICLE INFO

Article history:

Available online 24 March 2022

Keywords:

Accuracy
COVID-19
Naive Bayes classifier
Prediction
Random forest

ABSTRACT

Infections such as COVID-19 are affecting the entire world and measures such as social distancing can be done so that the contact among people is reduced. IoT devices usage keeps on increasing every day thereby connecting the environments physically. Among the current technologies, machine learning can be employed along with IoT devices. Predicting the risk related with COVID-19, a novel method employing machine learning is proposed. Random forest and Naive Bayes classifier are used for the prediction from the data collected with the help of sensors. Groups of people are recognized and the disease impact can be reduced for the particular group with more population. The accuracy of RF is 97% and for NB it is 99%.

Copyright © 2022 Elsevier Ltd. All rights reserved.

Selection and peer-review under responsibility of the scientific committee of the International Conference on Innovative Technology for Sustainable Development.

1. Introduction

Device communication has become easier nowadays due to the rise in several new technologies along with the improvements made to the systems available for computing and the protocols in the Internet. A huge number of devices are expected to be connected with the help of the Internet in the year of 2021 [1]. These improvements have led to the rise of a new research domain called Internet of Things abbreviated as IoT. Technologies of wired as well as wireless communication along with sensors and the objects present physically are connected with devices and these all together constitute IoT [2]. There is a necessity for data to be represented for obtaining services at a better level and the framework can be enhanced for accomplishing better performance. As a result of this, raw data gathered from various sources can be accessed by the systems and analysis can be performed for extraction of knowledge [3]. The risk related to prediction of covid are BMI, age and Gender.

New data extraction can be done using IoT and this in combination with data science can generate intelligent applications. Various research domains such as data mining as well as machine learning with other methods can be utilized in identification of

patterns from the available data in data science [4]. Algorithms can be applied in various research domains and the data types involved while using data analytics techniques are the 3Vs called Volume, Velocity and Variety. Neural networks, methods available for performing clustering as well as classification, are the data models that apply algorithms to match the characteristics of the data [5]. As data gets generated from various sources and also the data types get varied, it is necessary to find algorithms that can adapt the characteristics of data. Another reason is that the various resources utilized in generation of data have scaling as well as velocity problems [6].

1.1. Big data analytics Vs machine learning

Collection along with data set analysis is done in the process of big data analytics for discovering the patterns that are hidden [7]. Related information such as choices made by the customer and the trends in the market for helping the organizations to inform them at the right times. The three Vs related to big data are data volume, data type variety and processing velocity of the data [8]. The accuracy of software applications can increase by using the field of artificial intelligence termed as machine learning [9]. Computers are educated by the computer for performing tasks which are complex in nature and these tasks cannot be accomplished by humans.

* Corresponding author.

E-mail address: deepa23narayanan@gmail.com (N. Deepa).

Hadoop is one of the important tool for is.

- i) To understand the cost of big data and its storage functions. On the other hand, machine learning deals with providing the learning ability to the computers while they are not programmed explicitly for those tasks. Discovery of patterns that are being hidden in big data and information extraction are the important functions of big data analytics as analysis of big data is performed. Learning how to respond to the inputs that are unknown and providing outputs that are desirable.

2. Literature review

2.1. Machine learning – an efficient algorithm for prediction

Machine learning is assuming all the possible results using the various problem solving and techniques [10]. The algorithms that are working on the basis of machine learning which can be able to analyze the supervised datasets that are extracted properly and cluster in common group. Such that the algorithm such as linear regression, logistic, random forest are identified as best technique for health care based IoT applications. Table 1 describes the categories of various types ML.

2.2. Machine learning and its control application in Covid'19

Machine learning the most popular machine learning which are used for IOT applications are utilized for social media based projection such as identifying objects, image recognition etc. Big data analytics deal with IoT applications for analyzing and classifying feature extraction [11]. Algorithms such as reinforcement learning which includes both raw and mixed data to understand the pattern. The good accuracy can be obtained from the predicted terms to find the right classifiers based on the probability. Using the input feature and number of samples the stability of covid'19 and its cause due to increase in effects from patients are identified. The descriptive details based on the features that are available in patient's severity are identified using the fields which can be taken to fix the model according to the output [12] (see Tables 2 and 3).

2.3. COVID-19

Several threats do exist in the world periodically as pandemics and have been creating a massive impact in the current world too [13]. COVID-19 has created a huge impact in the economic growth of a country as well as affecting the strengths of a nation. For predicting this disease accurately, it is necessary to analyze the disease progression. The reason for the progress of this disease is mainly infection exposure. As a result of this exposure, hosts for infection are created who are actually a set of people whose infection level is higher. When these people contact others, the spread of disease starts. The spread of diseases using the pathogenic agents such as viruses are called acquired diseases and COVID-19

is a prominent example for this. The factors that are responsible for causing these diseases include the environmental characteristics, agents who act as carriers, hosts who get infected and finally it also includes the pathogens. Carriers are those who act as agents and they form a major part in infection. The host gets affected by this infection when he contacts the host. Vector is another name for pathogens. Organisms that transfer infection through virus or bacteria from one person to another are called vectors. The outbreaks resulting from the pattern of spread are called pandemics. This disease affects both elderly people and children and also affects the individual who has high blood pressure or asthma or cancer and other complications.

2.4. Machine learning algorithms

The disease forecasting can be done using various methods based on the data used. Prediction can be done using machine learning algorithms as accuracy of prediction is high in this case. Issues related to model training as well as parameter selection along with the model to be considered for prediction are to be noted in ML. Naive Bayes classifier in machine learning performs better compared to other traditional models available in the literature. Applications based on time series can use NB models. Most of the nations across the world utilize models based on machine learning algorithms for estimation of the infection of COVID-19 in the future. The researchers compared the several algorithms in machine learning which were used for forecasting COVID-19. Their analysis included heuristic algorithms like Genetic algorithm, PSO termed as Particle Swarm Optimization along with algorithms in ML like MLP termed as Multilayer Perceptron and so on. Accuracy was the parameter used for evaluation of the models used in prediction. Five countries were chosen and their data was analyzed to attain the 0.99 correlation level.

2.5. IoT applications

The internet of things (IoT) is the base of operations in an advanced manner that are related to information extraction. Devices such as wearable which helps in watching virtual views, GPS tracking in tracking the object or patients can be monitored using the effective techniques. Moreover the health systems are also available with sensors to monitor the serious condition of covid affected patients. The isolated patients who are suffering with the idle moment can be avoided using the digitized applications which can continuously monitor the critical conditions. The situation such as low blood pressure, oxygen (O2) level stability can be noted according to the variations of human health condition. To measure the severity and risk prediction based on dataset covid analysis can be done using machine learning algorithms. In terms of understanding the main features such as body mass index, age and heart rate the.

Table 1
Categories of machine learning.

Categories of ML Algorithms	Input	Output	Objective	Applications
Supervised	Data - Labeled	Output - Known	Used for prediction	Image classification, Email Spam Detection
Unsupervised	Data - Unlabeled	Output - Unknown	Used for analysis	Text Mining, Face Recognition
Semi-supervised	Less data - Labeled	Less Output - Unknown	Combines supervised + unsupervised	Speech analytics, Web content classification
Reinforcement	Learning based on Reward	Action	Decision made on experiences	Gaming, Robot Navigation

Table 2
Area under curve.

Area Under the Curve				
Test Result Variable(s): Discharged				
Area	Std. Error	Asymptotic Sig.b	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
0.976	0.024	0.108	0.928	1.000

a. Under the nonparametric assumption.
 b. Null hypothesis: true area = 0.5.
 Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.
 a. The positive actual state is 10.

Table 3
Case Processing Summary.

Case Processing Summary	
Death	Valid N (list wise)
Positive	0
Negative	42

3. Smart devices and its involvement with Machine learning algorithm

In terms of Machine learning (ML) involvement the complex data are processed easily. Whenever consumer-based systems are accessed there is a lot of data produced such as the surfing data, advertisements, social information etc. Internet of things (IoT) and its connection with ML are communicating the data storage and its scattered retrieval procedure. Therefore all applications are almost automated such as IoT, triggering the speed and performance of each action in a fast and efficient way. Therefore the social and environmental processes gather the collective data for implementing various predictions and calculating the probabilistic statistics.

3.1. Proposed approach

The sections of the proposed system is as follows (Fig. 1):

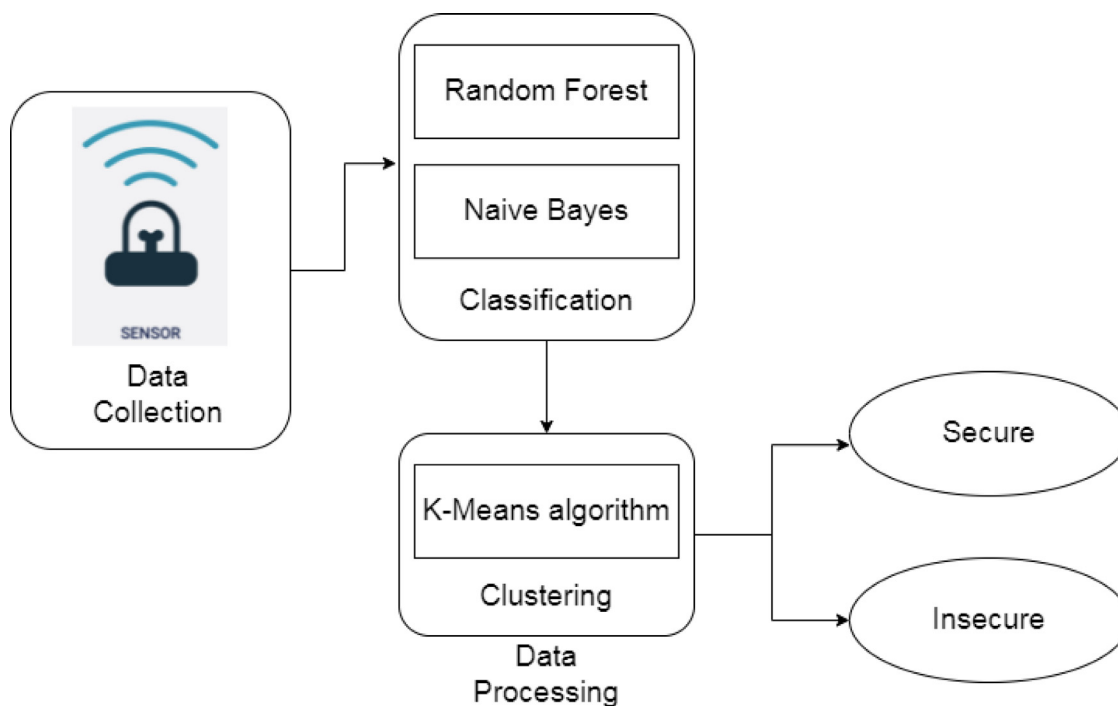


Fig. 1. Proposed system model.

- (i) Collection of Data: Data is collected using IoT sensors and this data contains details regarding the distance among people along with their images. Mask usage by people, their body temperature, age of the specific person and name. Followed by this, this is subjected to processing of data. Data can be collected with the help of social media and also from several domains.
- (ii) Processing of Data: In the module of data processing, classification along with clustering process is done. Random Forest (RF) and Naive Bayes (NB) algorithm for results prediction. Performance analysis is done to evaluate the working of the system. For data interpretation, these machine learning algorithms are necessary. Inputs are analyzed for production of outputs. Classification is performed initially as the data gets classified as “Secure” and “Insecure”. Parameter measurement is done with the help of IoT devices and for detection, the RF and NB are used. For the comparison of input as well as trained data, K-means clustering technique is employed where the instances of data seems to be similar.
- (iii) Machine Learning Algorithms: RF and NB are the two algorithms employed for prediction purpose. Mathematical expressions are used for prediction purpose in RF algorithm whereas NB uses classification rules for automatically classifying data.

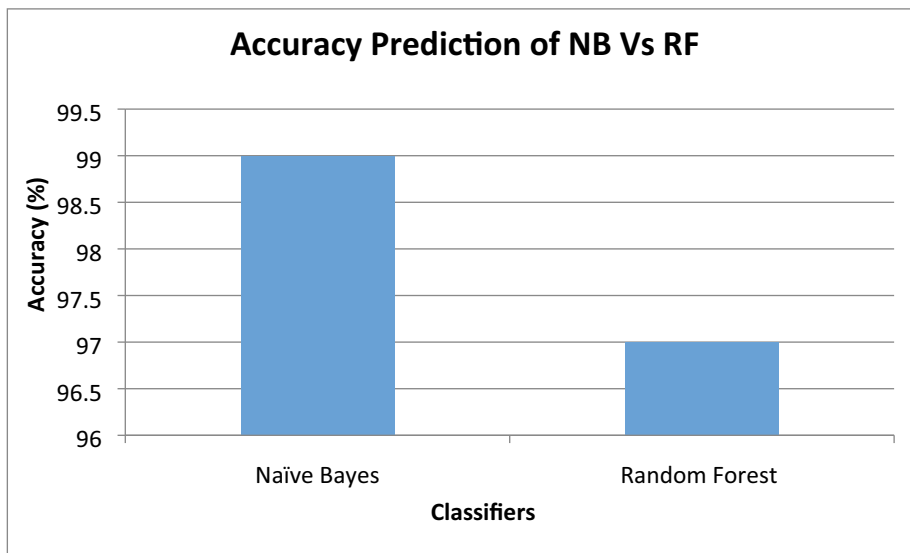


Fig. 2. Proposed system model.

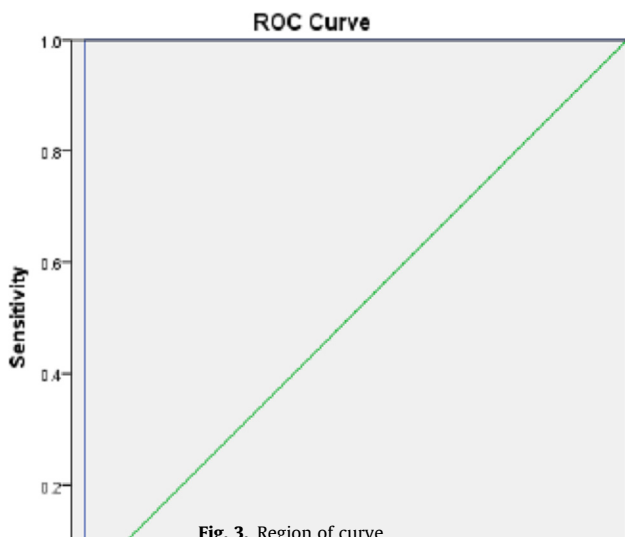


Fig. 3. Region of curve.

Algorithm 1. RF Algorithm.

Input: Feature response variable (R), multiple model ensemble size (M), ADA BOOST sampling dimension (N), FOR m = 1 to M Do
 Build Decision trees for unique sample from R, Generated tree as output(O) where a number of K records are used from immune dimensionality in accordance with |R| feature space is minimized
 Model Rm segregate features for training and testing where Decision tree is not visible during testing without Major voting (stable)
 END
 OUTPUT: Multiple models of decision tree created as Ensemble model

Algorithm 2. RF Algorithm.

Input: Independent dataset features for training (I), classes with its attribute list CA = {CA1,CA2,.. .CAN}
 Determining the features based on conditions from I calculate the probability of occurrence to find the priority based on mean and standard deviation
 Each class display the prediction for each matching attributes Identify the predictors matrix based on probability along with its attributes CA = {CA1,CA2,.. .CAN}
 STOP
 Posterior probability of unique class calculation
 OUTPUT: classifier model displays the class of each testing attributes.

(iv) Results: After the data is collected, it is sent for processing followed by machine learning algorithms. Finally obtained data from the model indicates the safeness associated with the places.

Data vector is created with the data collected from people with the help of parameters. Every cluster can be classified as SECURE or INSECURE thereby enabling people to be aware of the protocols. This helps in reduction of the disease spread.

4. Results and discussion

RF algorithm and the NB algorithm are used for prediction of covid 19 diseases using parameters by the process of classification. The accuracy of classification of the former is 97% whereas the latter takes the accuracy of 99% (Fig. 2). The real time data is gathered for analyzing the performance of RF and NB.

Every cluster is subjected to K-means clustering algorithm for clustering as SECURE or INSECURE. In terms of performance met-

rics various districts with the level of increase and decrease count the analysis have been done to find the region of curve parameters such as discharged in number and death rate which causes risk are taken as result variables.

The specificity and sensitivity based on number of true positive and false positive rate is calculated for case process summary according to the number of samples used in the performance analysis.

Fig. 3 represents the region of curve according to the positive and negative number of cases that are affected by death. Risk factors is highlighted based on the number of patients who got treatment and discharged after regular body mass index and proper functioning of heart rate in their regular checkup.

5. Conclusion

The economy of world is now affected by the spread of COVID-19 disease and social distancing needs to be maintained. The symptoms of COVID-19 resembles those of either bacterial of viral infection. World economy along with lifestyle as well as technology also gets affected. Yet another symptom of this disease is COVID-19. A novel method for monitoring this disease using machine learning algorithms such as Random forest and Naive Bayes algorithm is proposed. The accuracy of NB is 99% whereas for RF it is 97%. Clustering is performed using K-means algorithm as SECURE and INSECURE.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] K.H. Abdulkareem, M.A. Mohammed, A. Salim, M. Arif, O. Geman, D. Gupta, A. Khanna, Realizing an Effective COVID-19 Diagnosis System Based on Machine Learning and IoT in Smart Hospital Environment, *IEEE Internet Things J.* 8 (21) (2021) 15919–15928.
- [2] S. Kushwaha, S. Bahl, A.K. Bagha, K.S. Parmar, M. Javaid, A. Haleem, R.P. Singh, Significant applications of machine learning for COVID-19 pandemic, *J. Indust. Integrat. Managem.* 05 (04) (2020) 453–479.
- [3] Absalom E. Ezugwu, Ibrahim Abaker Targio Hashem, Olaide N. Oyelade, Mubarak Almutari, Mohammed A. Al-Garadi, Idris Nasir Abdullahi, Olumuyiwa Otegbeye, Amit K. Shukla, and Haruna Chiroma, A novel smart city-based framework on perspectives for application of machine learning in combating COVID-19, *Cloud Analytic Based Applications for Bioinformatics*, Vol. 2021.
- [4] M. Golec, R. Ozturac, Z. Poorianian, S. S. Gill, R. Buyya, iFaaSBus: A Security and Privacy based Lightweight Framework for Serverless Computing using IoT and Machine Learning, in: *IEEE Transactions on Industrial Informatics*.
- [5] R. Mukherjee, A. Kundu, I. Mukherjee, et al., IoT-cloud based healthcare model for COVID-19 detection: an enhanced k-Nearest Neighbour classifier based approach. *Computing* (2021).
- [6] D.R. Krithika, K. Rohini, Comparative interpretation of machine learning algorithms in predicting the cardiovascular death rate for Covid-19 data, in: *2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, 2021, pp. 394–400.
- [7] D.N. Le, V.S. Parvathy, D. Gupta, et al., IoT enabled depthwise separable convolution neural network with deep support vector machine for COVID-19 diagnosis and classification, *Int. J. Mach. Learn. Cyber.* 12 (2021) 3235–3248.
- [8] A. Chahal, P. Gulia, N.S. Gill, COVID-19 Vaccination Monitoring Using IoT and Machine Learning, in: D. Gupta, Z. Polkowski, A. Khanna, S. Bhattacharyya, O. Castillo (Eds.), *Proceedings of Data Analytics and Management. Lecture Notes on Data Engineering and Communications Technologies*, Springer, 2022.

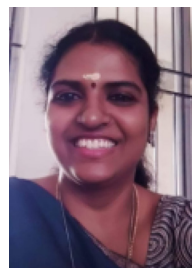
- [9] M. Otoom, N. Otoom, M.A. Alzubaidi, Y. Etoom, R. Banihani, An IoT-based framework for early identification and monitoring of COVID-19 cases, *Biomed. Signal Process. Control* 62 (2020) 102149.
- [10] D. Yacchirema, A. Chura, SafeMobility: An IoT-based System for safer mobility using machine learning in the age of COVID-19, *Procedia Comput. Sci.* 184 (2021) 524–531.
- [11] O. Shahid, M. Nasajpour, S. Pouriyeh, R.M. Parizi, M. Han, M. Valero, F. Li, M. Aledhari, Q.Z. Sheng, Machine learning research towards combating COVID-19: Virus detection, spread prevention, and medical assistance, *J. Biomed. Informat.* 117 (2021) 103751.
- [12] T. Devi, N. Deepa, Predicting the sales of cars using random forest algorithm, *Turkish J. Physiother. Rehabil.* 32 (2) (2021) 1361–1364.
- [13] N. Deepa, T. Devi, Prediction of short text in weblogs for accurate classification using IOHE a machine learning technique, *Turkish J. Physiother. Rehabil.* 32 (2) (2021) 1320–1323.



Mrs.N. Deepa working as a Assistant Professor at the Department of Computer Science and Engineering in Saveetha School of Engineering, SIMATS, Chennai. She obtained her Bachelor degree in B.E. Computer Science and Engineering from Raja College of Engineering, Madurai and Master degree in M.E Computer Science and Engineering from Sathayabama University, Chennai. She is currently pursuing Doctor of Philosophy in Computer Science and Engineering from Saveetha University, SIMATS Chennai majoring in Medical Image Processing and Deep Learning. Her areas of specializations include Deep Learning, Machine Learning, Mobile Computing, Cryptography, Network Security and Block Chain. She published her research work in reputed international Journals and presented papers in both International and National Conference. She has published patent in Computer Science and Engineering Field.



Dr. J. Sathya Priya working as a Associate Professor at the Department of Information Technology in Velammal Engineering College, Chennai. She obtained her Bachelor degree in B.E. Computer Science and Engineering from St. Peters Engineering College, Chennai and Master degree in M.E Computer Science and Engineering from S.A. Engineering College, Chennai. Then she obtained Doctor of Philosophy in Information and Communication Engineering from Anna University majoring Network Security. Her areas of specializations include Wireless Sensor Networks, Internet of Things, Cryptography, Artificial Intelligence, Machine Learning, Block Chain and Data Mining. She published her research work in reputed international Journals and presented papers in both International and National Conference. She authored books namely Cryptography and Network Security, Software Project Management, Computer Networks, Database Management Systems and Mobile Communication.



Ms.Devi.T working as an Assistant Professor at the Department of Computer Science and Engineering in Saveetha School of Engineering, SIMATS, Chennai. She obtained her Bachelor degree in B.Tech. Information Technology from VMKV Engineering College, Salem and Master degree in M.E Computer Science and Engineering from Sri Krishna College of Engineering and Technology, Coimbatore. She is currently pursuing Doctor of Philosophy in Computer Science and Engineering from VIT University, Chennai majoring Cloud Security. Her areas of specializations include Cryptography, Network Security, Cloud Computing, Artificial Intelligence, Machine Learning and Block Chain. She published her research work in reputed international Journals and presented papers in both International and National Conference. She has published patents in Computer Science and Engineering field.