Nile Journal of Communication & Computer Science

Volume 1, Number 1, August 2021

Journal Webpage: https://njccs.journals.ekb.eg

Print ISSN: 2805-2366
Online ISSN: 2805-2374

A Review on COVID-19 Patients Detection Using Data Mining and IoT Technology

Asmaa H. Rabie[1], Nehal A.Mansour[1,2], Amal Al-Husseiny[1], and Ahmed I. Saleh[1]

1: Computers and Systems Department, Faculty of Engineering, Mansoura University, Egypt

2: Nile Higher Institute for Engineering and Technology, Artificial intelligence Lab., Mansoura, Egypt

Abstract

Early detection of COVID-19 patients is an important issue for disease cure and control. COVID-19 has not been previously identified in humans because it is a new species appeared in 2019. Unfortunately, COVID-19 spreads so quickly between people in few months. The most common symptoms of COVID-19 that can progress to a severe form of pneumonia with critical complications are dry cough, sore throat, and fever. In fact, clinical characteristics alone cannot determine the diagnosis of COVID-19 patients at the early-onset of symptoms. Thus, it is an important to find fast and accurate COVID-19 patients detection model that can quickly and accurately diagnose the patients. In this paper, many classification methods which can early detect COVID-19 patients are discussed. In fact, data mining is an effective tool that can be used in predicting medical conditions. Data mining can enable caregivers to accurately make medical decisions. Nowadays, Internet of Things (IoT) is implemented in the infrastructure of the medical system that leads to make the system more automated and enable the medical staff to monitor COVID-19 patients remotely.

Keywords: *COVID-19, Classification, Data mining, IoT.*

1. Introduction

Coronavirus is highly threatening for human and animal life in which some kinds of coronavirus can affect animals and then it jumps from animal species into the human population [1-4]. Among nucleic acid-based tests, Reverse Transcription Polymerase Chain Reaction (RT-PCR) test has been used as the ‘gold standard’ for confirming COVID-19 positive patients [4-6]. In fact, a negative result of RT-PCR test does not negate the possibility of COVID-19 infection because it has low sensitivity although it has high sp ecificity. Hence, laboratory test is more

accurate than an imaging test to diagnose COVID-19 patients because an imaging test suffers from false positive and false negative cases [4-6].

The direct diagnosis of COVID-19 patients by the doctors and nurses is extremely dangerous as it may expose their lives to death [4-6]. To overcome these challenges, data mining play an effective role in the medical system as it made the system be automated. Thanks to data mining techniques, COVID-19 patients can be remotely diagnosed. Classification methods are important to early detect COVID-19 patients. There are many classification methods that can quickly and accurately diagnose COVID-1 cases. These methods such as; K-Nearest Neighbor (KNN), Naïve Bayes (NB), Fuzzy Logic (FL), Artificial Neural Network (ANN), and Deep Learning (DL) [4-6].

Although these classification techniques can early detect COVID-19 patients, it may suffer from noise data. Thus, preprocessing data before learning the classifier is an important process to filter the data from irrelevant features. Thus, feature selection process is an important process that should be implemented on COVID-19 dataset before starting to learn the diagnosis system [4-6]. Feature selection methods are categorized into two main groups, called; filter and wrapper [7-12]. While filter methods can provide a fast subset of features, wrapper methods can provide accurate subset of features that effect on COVID-19 diagnosis. Thus, using a hybrid feature selection methods that combine both filter and wrapper is important to provide fast and accurate subset of the most significant features in the COVId-19 dataset. Thus, many COVID-19 diagnosis strategies based on two layers; preprocessing layer to perform feature selection process and classification layer to diagnose the patients based on the selected features [4-6].

IoT is a new technology that help the medical system to remotely collect the data from patients without coming to medical system (i.e., hospitals) [13-16]. IoT sensors enable the hospitals to perform many procedures at a remote location in a real-time manner such as blood testing, diabetic monitoring, and pressure monitoring. Consequently, the collected real-time data from IoT sensors enables the medical system to give accurate real-time reactions. In fact, IoT technology enable a medical system to remotely monitor the patients by using wireless networking connections such as Bluetooth, Wi-Fi, etc. [13-16]. While IoT supports the medical systems with real-time data, it cannot formulate this data in appropriate form for analysis. Hence, IoT transmit data that is sent to be stored in the cache server of the medical system for further processing.

In this paper, a review of different used classification methods applied as COVID-19 patients diagnosis techniques will be discussed. The rest of the paper is organized as follows: *Section2* describes the COVID-19 problem. *Section3* shows an overview of the widely used classification methods. *Section 4* provides the role of IoT on COVID-19 patients. Finally, conclusions are discussed in *Section 5*.

2. Problem Definition

Once the coronavirus epidemic starts, it will take around four weeks to break the basic healthcare system. Cases detection and isolation is the golden solution for protecting the healthcare system from becoming overwhelmed, and accordingly will flat the epidemic curve as shown in figure 1.

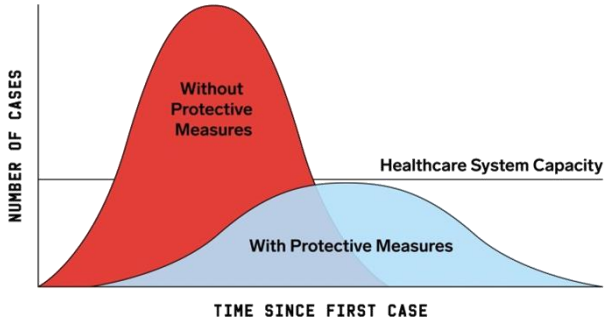

Figure 1, COVID-19 epidemic curve with and without protective measures.

In fact, a negative result of RT-PCR test does not negate the possibility of COVID-19 infection because it has low sensitivity although it has high specificity. Hence, laboratory test is more accurate than an imaging test to diagnose COVID-19 patients because an imaging test suffers from false positive and false negative cases [4-6].

3. Classification Techniques

Over the centuries, several researchers directed their biggest attention to improve classification techniques [7-12]. These techniques are extended from traditional to soft computing or data mining techniques. Some of classification techniques which are widely applied are called NB classifier, KNN classifier, ANN, FL, and DL [4-6].

3.1. Naïve Bayes Classifier

One of most widely used as a statistical classifier is called Bayesian classifier. NB algorithm can predict class membership probabilities depending on a simplified version of Bayes formula. These probabilities, such as the probability that a given item belongs to a specific target class [4,5]. According to a given target class, all NB classifiers suppose that specific feature's value does not depend on the values of other features. Abstractly, NB is a conditional probability model that can be formulated by (1):

$$P(Ci|f1, f2, \dots\dots fn) = \frac{P(Ci)\prod_{j=1}^{n} P(fj|Ci)}{P(f1,f2,\dots\dots fn)} \qquad (1)$$

In equation (1), the left side represents the posterior probability of class C_i according to observed feature values that are given in the item to be categorized, $< f1, f2, \ldots \ldots fn >$. Additionally, the denominator of the right side is a constant, hence, it is often omitted. Indeed, this constant is easily calculated if one requires that the posterior probabilities of the classes sum to one. Thus, learning with NB classifier is straightforward. It also includes simply estimating the probabilities in the right side of equation (1) from the training items. So that, the result represents a probabilistic summary for every one of the possible classes.

3.2. K-Nearest Neighbor Classifier

A nearest neighbor classifier supposes all items related to objects in the n-dimensional space. KNN classifier is a popular technique in both data mining and statistics [6,11]. It considered one of the simplest and straightforward techniques that can handle multi-class problems [6,11]. KNN is based on a distance measure which is required to determine the "closeness" of items [6,11]. An item can be categorizes by arriving to its nearest neighbors and then electing the most popular class through its neighbors, as shown in figure 1. To determine the class of a new item, KNN can be expressed as following pseudo-code:

For each item ***S*** in the set
  Calculate distance ***D(S,G)*** between ***S*** and every other item ***G***
  Neighbors = the ***K*** neighbors in the training set that are most similar to ***S***
  Get the majority vote from neighbors.
End For

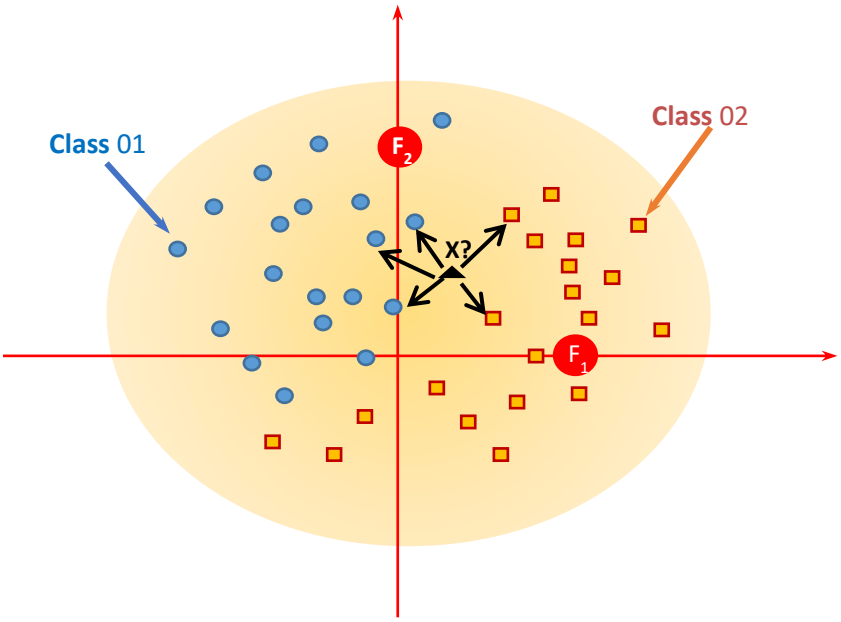

Figure 1: The K-Nearest Neighbor Method.

The execution of KNN depended on two major operators which are; no. of neighbors to be used, and similarity measure [6,11]. Widely, no. of neighbors to be used is dependent on experimental results. Generally, the similarity can be measured by using Euclidean distance that is the most common way [11]. Euclidean distance is the most universal, between two vectors A_{ir} and B_{ir}, the Euclidean distance is defined by (2):

$$D(A_i, B_i) = \sqrt{\sum_{r=1}^{n} (A_{ir} - B_{ir})^2} \qquad (2)$$

3.3. Artificial Neural Network Techniques

ANN is a data-driven machine learning technique that is used to solve non-linear problems. Actually, ANN is inspired from neural networks in the human brain [1-6]. According to figure 2, the main components of ANN are represented in three layers which are; (i) input layer, (ii) hidden layer, and (iii) output layer [4-6]. The use of input layer is to train the model in the training phase or to get forecasting in the testing phase from input data set. The hidden layer is used between the input and output layers, while the output layer is used to give the final result(s). There are three different criteria to design ANN which are; the interconnections between the different neurons of different layers, the learning methods, and each neuron's activation function based on the input.

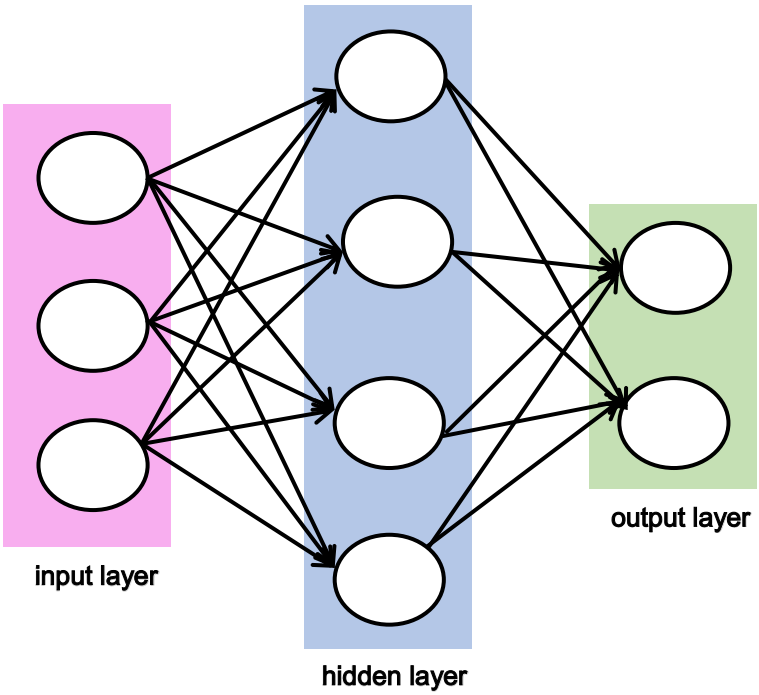

Figure 2: Main components of ANN.

ANN is a flexible and adaptable technique that can solve non-linear forecasting problems. ANN models can be used for COIVD-19 patients diagnosis using some of indicators as inputs. One of the ANN's drawbacks is that its models are subject to overfitting as the application of ANN is the black-box nature of the model. Additionally, ANN suffers from no transparency related to physical interpretation.

3.4. Fuzzy Logic

FL describes systems in terms of a combination of numeric and linguistic [9,17]. FL is a data mining technique that has many benefits such as it is simple and robust. Additionally, FL is flexible and easy to be implemented. Thus, it can be used to diagnose the COVID-19 patients. Although FL is a very convenient method for uncertain or approximate reasoning, it should be built with the full guidance of experts. The membership functions for the small, medium, and large fuzzy sets can be illustrated according to one input in figure 3. Additionally, these functions are formulated in (3)-(5) [9]:

$$\mu(x)_{small} = \begin{cases} 1 & x \preccurlyeq \propto \\ \dfrac{\beta - x}{\beta - \alpha} & \alpha \prec x \preccurlyeq \beta \\ 0 & x \succ \beta \end{cases} \qquad (3)$$

$$\mu(x)_{\text{medium}} = \begin{cases} 0 & x \preccurlyeq \propto \\ \frac{x-\alpha}{\beta-\alpha} & \alpha \prec x \preccurlyeq \beta \\ \frac{\gamma-x}{\gamma-\beta} & \beta \prec x \preccurlyeq \gamma \\ 0 & x \succ \gamma \end{cases} \quad (4)$$

$$\mu(x)_{\text{large}} = \begin{cases} 0 & x \preccurlyeq \beta \\ \frac{x-\beta}{\gamma-\beta} & \beta \prec x \preccurlyeq \gamma \\ 1 & x \succ \gamma \end{cases} \quad (5)$$

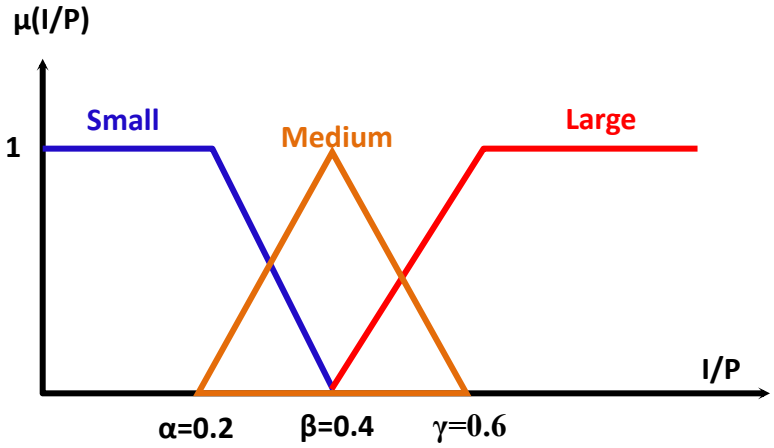

Figure 3: Membership function for any input.

3.5. Deep Learning

DL is a data mining technique that uses multiple layers to progressively extract higher-level features from the raw input [18]. Most modern DL models are based on ANN, specifically Convolutional Neural Networks (CNNs). The word "deep" in "deep learning" refers to the number of layers through which the data is transformed. Actually, DL process can learn which features to optimally place in which level on its own. DL has been shown to produce competitive results in medical application such as COVID-19 diagnosis [18]. The difference between ANN and DL is shown in figure 4.

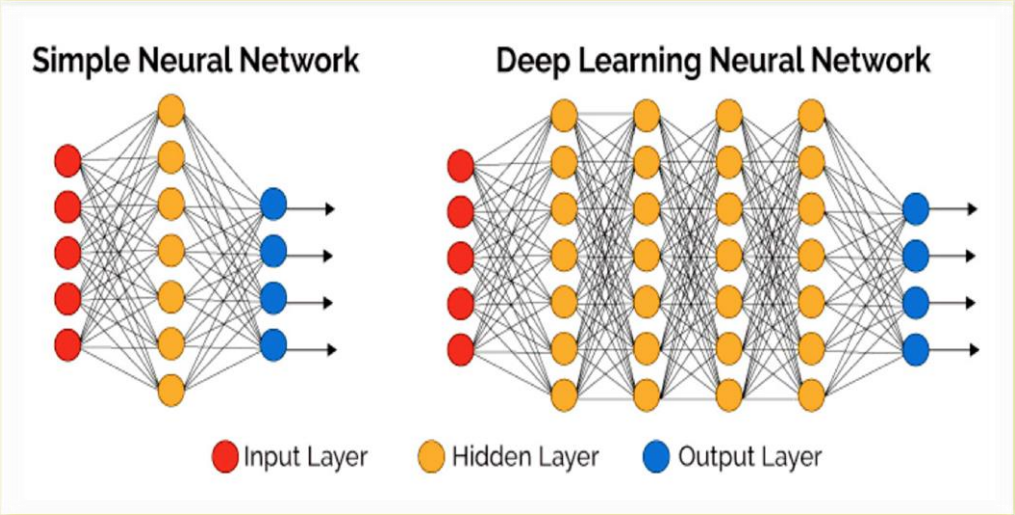

Figure 4: Difference between ANN and DL.

4. Important Role of IoT in COIVD-19

IoT technology plays an important role in different phases of various infectious diseases [19]. There is an essential need for patients to be connected with and monitored by their physicians proactively in different phases of COVID-19. In fact, IoT technology has main role in response to COVID-19 in three main phases; which are; (i) early diagnosis, (ii) quarantine time, and (iii) after recovery [19]. In early diagnosis phase, there is an essential need for faster diagnosis due to the high rate of contagiousness of COVID-19. That is because an asymptomatic patient can easily spread the virus to others. In quarantine time phase, IoT devices in this phase can monitor patients remotely with respect to their treatments and stay at home orders by the authorities [19]. Finally, after recovery phase, the chances of returning symptoms and potential infectivity can be high. Thus, social distancing should be implemented by deploying IoT devices to prevent that happening. These devices are bands and crowd monitoring devices. Many IoT devices including wearables, drones, robots, IoT buttons, and smartphone applications that are mainly utilized in the forefront of combating COVID-19 [19].

5. Conclusions

COVID-19 patients diagnosis is an important issue for disease cure and control. Thus, it is an important to find fast and accurate COVID-19 patients detection model that can quickly and accurately diagnose the patients. In this paper, COVD-19 problem was discussed and also many classification methods which can early detect COVID-19 patients were discussed. The effective role of IoT in COVID-19 was discussed as IoT was implemented in the infrastructure of the

medical system that leads to make the system more automated and enable the medical staff to monitor COVID-19 patients remotely.

6. References

[1] S. Zaim, J. Chong, V. Sankaranarayanan, and A. Harky, *"COVID-19 and Multiorgan Respons*e*,"* Current Problems in Cardiology, Elsevier, Volume 45, Issue 8, 2020, PP. 1-21.

[2] G. Rubin, C. Ryerson, L. Haramati, et al., *"The Role of Chest Imaging in Patient Management during the COVID-19 Pandemic: A Multinational Consensus Statement from the Fleischner Society,"* Radiology, https://doi.org/10.1148/radiol.2020201365, 2020, PP. 172-180.

[3] Y. Wang, H. Kang, X. Liu, and Z. Tong, *"Combination of RT-PCR testing and clinical features for diagnosis of COVID-19 facilitates management of SARS-CoV-2 outbreak,"* Journal of Medical Virology, Wiley, Volume 22, Issue 6, 2020, PP. 538-539.

[4] N. Mansour, A. Saleh, M. Badawy, and H. Ali, "*Accurate detection of Covid-19 patients based on Feature Correlated Naïve Bayes (FCNB) classification strategy*," Journal of Ambient Intelligence and Humanized Computing, Springer, https://link.springer.com/article/10.1007/s12652-020-02883-2, 2021, PP. 1-33.

[5] W. Shaban, A. Rabie, A. Saleh, and M. Abo-Elsoud, *"Detecting COVID-19 patients based on fuzzy inference engine and Deep Neural Network,"* Applied Soft Computing, Elsevier, Volume 99, 2021, PP. 1-19.

[6] W. Shaban, A. Rabie, A. Saleh, and M. Abo-Elsoud, *"A new COVID-19 Patients Detection Strategy (CPDS) based on hybrid feature selection and enhanced KNN classifier,"* Knowledge-Based Systems, Elsevier, Volume 205, 2020, PP.1-18.

[7] A. Rabie, A. Saleh, and K. Abo-Al-Ez, *"A New Strategy of Load Forecasting Technique for Smart Grids,"* International Journal of Modern Trends in Engineering and Research (IJMTER), Volume 2, Issue 12, 2015, PP.332-341.

[8] A. Saleh, A. Rabie, and K. Abo-Al-Ezb, *"A data mining based load forecasting strategy for smart electrical grids,"* Advanced Engineering Informatics, Elsevier, Volume 30, Issue 3, 2016, PP.422- 448.

[9] A. Rabie, S. Ali, H. Ali, and A. Saleh, *"A fog based load forecasting strategy for smart grids using big electrical data,"* Cluster Computing, Springer, Volume 22, Issue 1, 2019, PP.241- 270.

[10] A. Rabie, S. Ali, A. Saleh, and H. Ali, *"A new outlier rejection methodology for supporting load forecasting in smart grids based on big data,"* Cluster Computing, Springer, Volume 23, 2020, PP. 509-535.

[11] A. Rabie, S. Ali, A. Saleh, and H. Ali, *"A fog based load forecasting strategy based on multi-ensemble classification for smart grids,"* Journal of Ambient Intelligence and Humanized Computing, Springer, Volume 11, Issue 1, 2020, PP. 209-236.

[12] A. Rabie, A. Saleh, and H. Ali, *"Smart electrical grids based on cloud, IoT, and big data technologies: state of the art,"* Journal of Ambient Intelligence and Humanized Computing, Springer, https://doi.org/10.1007/s12652-020-02685- 6, 2020, PP. 1-32.

[13] A. Cortegiani, G. Ingoglia, M. Ippolito, A. Giarratano, and S. Einav, *"A systematic review on the efficacy and safety of chloroquine for the treatment of COVID-19,"* Journal of Critical Care, Elsevier, https://doi.org/10.1016/j.jcrc.2020.03.005, 2020, PP.1-5.

[14] F. Felice, A. Polimeni, and V. Valentini, *"The impact of Coronavirus (COVID-19) on head and neck cancer patients' care,"* Radiotherapy and Oncology, Elsevier, Volume 147, 2020, PP. 84-85.

[15] A. Narin, C. Kaya, and Z. Pamuk, *"Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks ,"* arXiv:2003.10849, 2020.

[16] V. Jagadeeswari, V. Subramaniyaswamy, R. Logesh, and V. Vijayakumar," *A study on medical Internet of Things and Big Data in personalized healthcare system*," Health Information Science and Systems, Springer, Volume 6, Issue 14, 2018, PP. 1-20.

[17] O. Rebrovs and G. Kuļešova, *"Comparative Analysis of Fuzzy Set Defuzzification Methods in the Context of Ecological Risk Assessment"*, Information Technology and Management Science, The Journal of Riga Technical University, Volume 20, Issue 1, 2017, PP. 25-29.

[18] T. Ozturk, M. Talo, E. Yildirim, et al., *"Automated detection of COVID-19 cases using deep neural networks with X-ray images,"* Computers in Biology and Medicine, Elsevier, Volume 121, 2020, PP. 1-11.

[19] M. Nasajpour, S. Pouriyeh, R. Parizi, et. al*.," Internet of Things for Current COVID-19 and Future Pandemics: an Exploratory Study,"* Journal of Healthcare Informatics Research, Springer, Volume 4, 2020, PP.325–364