International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Early-Stage Detection of Covid-19 Patient using ML Model: A Case Study

Pooja Mittal[1], Navita*[2]

Submitted: 18/10/2022 **Revised**: 22/12/2022 **Accepted**: 07/01/2023

Abstract: The Covid-19 pandemic has drastically changed the daily living style of human beings by astonishing the cultural, educational, regional, business, social, and marketing activities within a limited boundary. It also has impacted the healthcare system globally and provided a lot of burden on the healthcare system. The circumstances that arose due to such a pandemic require a vital solution to deal with it. In such a situation, most innovative technologies have grown up to find alternative solutions to track the situation that arises due to Covid-19. Among all innovative technologies, IoT can be counted as the best approach to deal with such a type of pandemic due to its associated features of transmitting data from any remote location without human intervention. Such type of technology has the capability of providing connectivity among various medical devices either in hospitals or other deliberate places to deal with such type of pandemic. First of all, this paper introduces the concept of IoT to deal with the circumstances of the Covid-19 pandemic. Along with that, a framework of a real-time Covid-19 patient monitoring system has been proposed in this paper that can be utilized in the future. The proposed framework helps in monitoring the symptoms of Covid-19 infected patients. On the basis of that model, a case study is done on Covid-19 symptom data by using different ML algorithms. The findings indicate that all algorithms achieved an accuracy of more than 80% and RFT achieved the highest accuracy of 92%. Based on these findings, we believe that these algorithms will produce efficient and precise outcomes when applied to real-time symptom data.

Keywords:*Covid-19, Internet of Things (IoT), Machine Learning (ML)*

1. Introduction

The novel Coronavirus of 2019 or Covid-19 is a global epidemic that has generated lots of panic throughout the whole world. Since the "1918 H1N1 influenza pandemic", the novel COVID-2019 has been reported the most destructive and hazardous epidemic. "Severe Acute Respiratory Syndrome Coronavirus Two (SARS-CoV-2)" is thought to be the major source of Covid -19 [1], [2]. It is believed that the first Covid-19 patient was recognized in Wuhan city of China in December 2019 [3].

According to the World Health Organization report, by April 10, 2020, there would have been 100,075 deaths and 15,225,252 case reports recorded. Therefore, it can be observed that COVID-2019 has expanded rapidly since the first of December 2020. COVID-2019 has so far extended to 172 countries. On January 30, 2020, a student who had recently returned from Wuhan was the observed patient as the first case of COVID-19 in India. After that, in February just two cases were reported. More number of cases emerged in March, and since the second part of April 2020, there has been an increase in the number of cases. As of June 9, 2020, 32 states and union territories had stated a total of 266 598 COVID-19 confirmed cases, according to the "Ministry of Health and Family Welfare (MoHFW)" [4]. Delhi, Gujarat, Tamil Nadu, and Maharashtra are the states where the majority of instances have been reported. As of now, the MoHFW has documented 7471 deaths caused by COVID-19.

The Warning sign of Covid-19 is vastly variable, going from none to life-threatening be considered symptomatic and asymptomatic. Common symptoms often observed in Covid-19 patients are cough, fever, headache, breathing difficulties, fatigue, and loss of taste and smell [5].

Unfortunately, there is still no effective treatment and vaccination yet. The creation of a successful will takes longer than a year, in part because the virus's nature has not yet been fully characterized [6]. At that time, the single way the world can combat this situation is to reduce its spread—by using strategies like social isolation, hand washing, and face masks. Through early detection (or prediction) of novel cases and monitoring of their progression, technology could, however, also aid in slowing its spread [7,8]. By following lockdown and proper strategies infection rate dropped in September 2021. But this is not the end of Covid-19, a 2nd wave started in March 2021 that was highly destructive in comparison to the first wave, and regions of the country experienced shortages of doctors, vaccines, oxygen cylinders, hospital beds, and other medical supplies [9]. By the end of April, India had the most recent and active cases worldwide. On April 30, 2021, it was the first nation to record more than 400,000 new cases in a 24-hour period. [14] In late August 2021, Soumya Swaminathan claimed that if such a situation continues in India, most people learn to live with the virus. According to experts, the virus may persist in India in some form rather than completely disappear. [16] India had 78,190 active cases as of December 23, 2021, the fewest in 573 days [17].

1 Maharshi Dayanand University, Rohtak, 124001, India
ORCID ID : 0000-0001-9746-6621
2 Maharshi Dayanand University, Rohtak, 124001, India
ORCID ID : 0000-0001-7880-8972
** Corresponding Author Email: navitamehra55@gmail.com*

In March 2022, this number dropped to 21,530. Since the start of Covid-19, WHO has worked to find an effective tool to various nations and launched the Covid-19 vaccine on 16 January 2021. India also started its vaccination program on 16 January 2021 with the "AstraZeneca vaccine (Covishield)" and the indigenous "Covaxin"[10][11]. But the vaccine was intended in order to protect against "severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus that causes coronavirus disease 2019 (COVID-19)". Still, there is no proper treatment and protection against covid-19. In near future, this virus may come in different forms and causes a very critical situation whose major impact would be on the health detect and treat cases that are most likely to survive. By doing this, they may more wisely use the limited supply of medical resources and treatments.

Machine Learning AI (Artificial Intelligence) and the Internet of Things (IoT) have emerged as the ground-breaking technology of the 21st century and found use in a variety of sectors, including autonomous systems, astronomical exploration, weather prediction, and health prediction.

This work contributes the following:

(i) This work performs a case study of the proposed model on the existing Covid-19 dataset and provided a framework that can be followed for early detection and monitoring of Covid-19 patients in an IoT Environment in future

(ii) This work also plays a significant role in understanding the nature of viruses by gathering and examining data with the help of ML.

(iii)The major objective of this work is to lower the death rates through early identification, monitoring of patients, and increased knowledge of the illness by utilizing the proposed model on an available dataset.

(iv) This study adopts various performance metrics to check the performance of the proposed model on gathered data patterns.

(v) The work also emphasizes the evaluation of various classification methods for the classification and diagnosis of disease.

The organization of this paper is as follows. Section 2 details about material and methods utilized for this work. Section 3 focuses on the experimental result and analysis performed on available datasets. Lastly, Section 4 concludes the work.

2. Materials And Methods

This section will outline the procedures and materials used to carry out the suggested work.

2.1. Proposed Model by utilizing the Concept of IoT

This section presents and explains our hypothetical IoT-based model that may be used to continuously track and detect (or anticipate) possible coronavirus cases in the future. The framework of our suggested IoT architecture, as shown in Fig. 1, is composed of three basic layers: the first layer is related to symptom data collection and uploading, the second layer is related to the processing of collected data using ML algorithms and the third layer relates to alert generation depending upon the analysis result.

2.1.1 Symptom Data Collection Layer

This layer relates to collect of the symptoms of patients using different wearable sensors. In previous studies, the most common symptoms related to covid-19 are identified depending on the real-time covid-19 dataset [13]. The most important sensors-based devices utilized for self or real-time monitoring of a patient's symptoms are non-contacting temperature measuring thermometers, pulse oximeters, etc. A real-life scenario in which a patient is monitored by utilizing different sensors is shown by Fig. 2. The most commonly identified symptoms are cough, fever, cough, shortness of breath, and sore throat. In the new era of technology, there exist lots of biosensors having the capability to detect such symptoms. For example, non-contacting temperature-based sensors are utilized for the identification of fever [14]. Cough identification at different age groups is possible by utilizing audio-based sensors [15]. Different motion-based and heart-rate sensors can be utilized to detect fatigue [16]. For the detection of shortness of breath oxygen-based devices can be utilized [17]. Data collected through all these sensors could be communicated Data collected through all these sensors could be communicated to some storage devices by utilizing effective communication technologies like Zigbee, lowan, Bluetooth and wi-fi, etc. [18]. The flow chart in which the proposed model work is shown by Fig 3.

2.1.2 Data Processing Layer

At this layer data collected from different wearable devices was collected and stored on some storage devices. For offering remote availability of patient data, it will be stored in the cloud, so that it would be accessed and analyzed from any location. Data uploaded from various IoT devices werefurther examined, to produce insights, and to work together to make more decisions. For this purpose, different ML/Deep Learning techniques were utilized and replaced the traditional approach of observation [19][20][21].

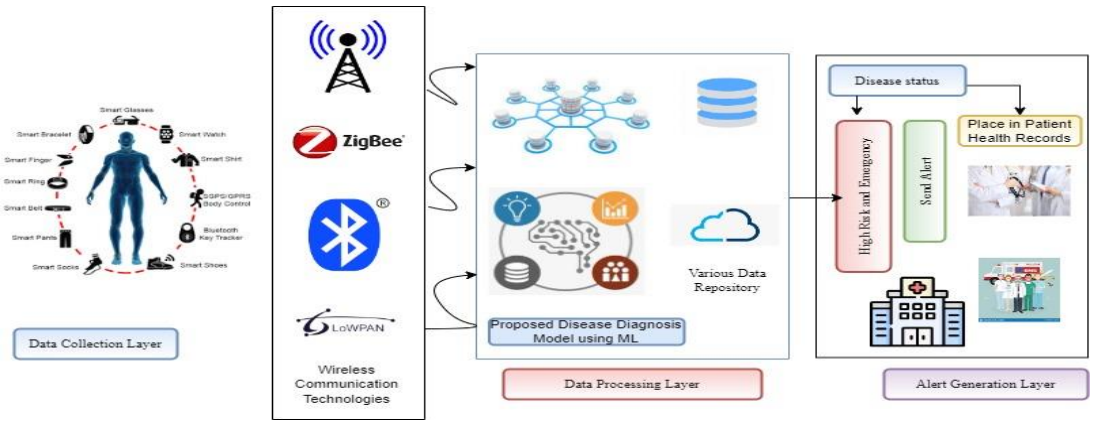

Fig. 1. IoT based Covid-19 Patient Monitoring Framework using different Wearable Sensors [12]

2.1.3 Alert Generation Layer

This layer relates to the generation of alerts to the patients, doctors, and family members in case of any critical situation arise due to changes in symptoms parameters.

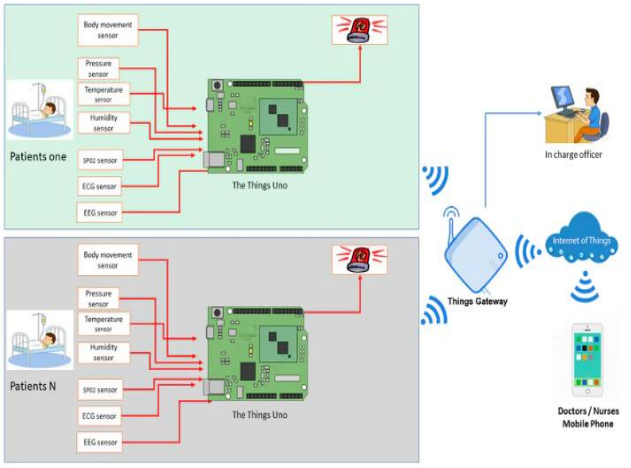

Fig. 2. Real -Life Scenario of Identification and Monitoring of Covid-19 Patient in IoT-based Environment [22]

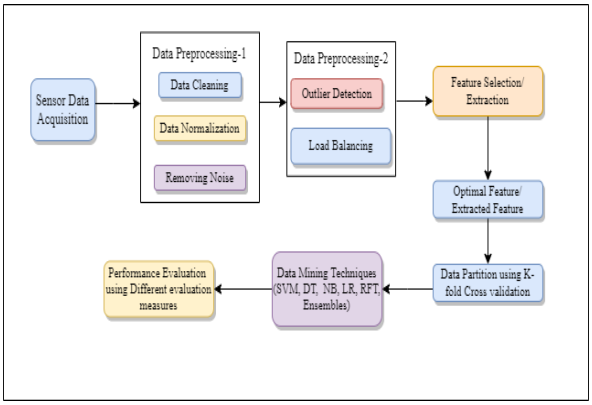

Fig. 3. Flow Chart for Proposed Disease Diagnosis Model

2.2. Data Set and its Implementation

The dataset and tool used to implement will be described in this section. Complete implementation of proposed model is done in Python programming by utilizing Jupyter notebook. Most commonly libraries like pandas, matplotlib, sns, and different classifier libraries available in Python are utilized for implementation purposes.

2.2.1 Data Acquisition

A Covid-19 symptom-based dataset is collected from different countries and combined to form a single dataset and is accessible from Kaggle "https://www.kaggle.com/iamhungundji/covid19-symptoms-checker". All the collected symptoms are based on a standard set of systems as prescribed by the "World Health Organization (WHO)" in order to determine whether the person has Covid or not. The symptom data set was obtained by utilizing different biosensors-based medical devices. The complete dataset has 27 attributes. The details about each attribute are represented in Table1.

Table 1: Data Set Description

Attribute No.	Attribute Name
1	Fever
2	Tiredness
3	Dry-Cough
4	Difficulty-in-Breathing
5	Sore-Throat
6	None _Symptom
7	Pains
8	Nasal-Congestion
9	Runny-Nose
10	Diarrhea
11	None_ Experiencing
12	Age_0-9
13	Age_10-19
14	Age_20-24
15	Age_25-59
16	Age_60+
17	Gender _Female
18	Gender_ Male
19	Gender_ Transgender
20	Severity_ Mild
21	Severity_ Moderate
22	Severity_ None
23	Severity_ Severe
24	Contact_ Don't-know
25	Contact_ No
26	Contact_ Yes
27	Country

2.2.2 Data Preprocessing

The existing dataset comprises attributes with distinct values. For instance, severity attributes contain four different records none, moderate, mild, and severe values. A symptom like Dry cough, fever, breathing, etc. contains binary values represented by pie chart in Fig. 4.

All ML algorithms expect information to take the shape of a vector number. So, in order to translate static data into separate numeric values for each characteristic, a Label Encoding technique is implemented. We have dropped these different severity columns and considered one target column. For further analysis we have considered only symptoms parameters as describedpresented by Fig. 5.

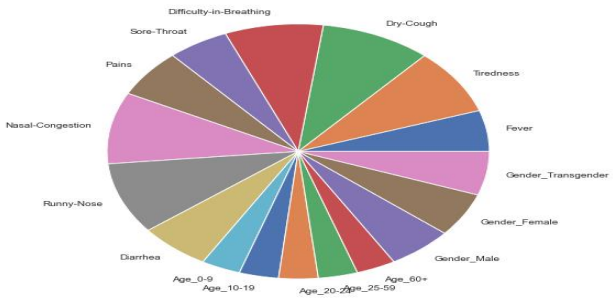

Fig. 4. A Pie Chart Representation of Different Symptoms of Covid-19

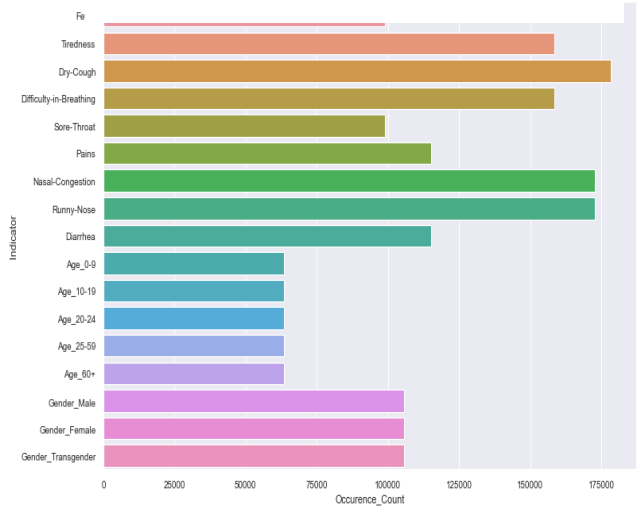

Fig. 5. Bar Chart Representation of Major Attributes after Data Preprocessing

2.2.3 Feature Selection

This is the most important steps to be followed for better prediction of results. There exist the most irrelevant features in the dataset that when passed to the model may take high computational time for the both testing and training phase. So, before passing all the features to a model, some feature selection mechanism must be followed for selecting efficient features. In this approach, we have followed PCA (Principal Component Analysis) as a feature selection technique. This is a statistical method followed to convert high dimensional data into small dimensional data by selecting the most important features on the basis of variance.

2.2.4 Machine Learning Techniques

This work mainly relates to the classification of Covid-19 patients' severity levels into four categories: none, mild, moderate, and severe. On basis of machine learning algorithms, the present state of an infected person is predicted. The subsequent ML algorithms were employed: Random Forest Tree (RFT), Naïve Bayes (NB), Decision Tree (DT), Logistic Regression (LR), AdaBoost, K-Nearest Neighbors (K-NN), etc. [23][24].

LR is mostly used to establish a connection between independent and dependent categorical variables. If the dependent variable contains only two values such as 0 and 1, true and false, and yes and no then it is called dual logistic regression and if it contains more than two values then it is called multinomial logistic regression. LR used a mathematical expression to predict transformation about any dependent variable given as

$$LR(p) = In(p/(1-p)) \quad (1)$$

p: the probability of 1 outcome, 1-p probability of 0 outcome

SVM is an efficient learning algorithm used mostly for classification problems and provides the best result in less computational time. This method is based on the principle of finding the optimal hyperplane which can easily classify the data in their related class. Naïve Bayes is a Bayes Theorem-based probabilistic classifier. All dataset instances are classified based on specific attributes or features. Bayes Theorem is given as:

$$P(A \mid B) = (P(B)P(A))/(P(B)) \quad (2)$$

K-NN is a very simple and easy-to-use Model used for both classifications as well as a regression problem. It is based on k Neighbour's approach in which classification of data is made based on similarity measures. It first checks the similarity between new data and existing data in the feature space and places the new data within a similar class of data.

Decision Tree: Due to its ease of use and ability to handle both continuous and categorical information, DT is very beneficial for classification jobs. Starting with a root node, a decision tree creates a tree-like structure. Subsequent node splitting is based on factors such as the Gini index, entropy, gain ratio, etc.

RFT is primarily used to handle complex problems and is built utilizing numerous decision trees. Both classification-based issues and regressions make heavy use of Random Forest. The bagging method was used by RF to produce results, and the final determination was based on the outcomes of the majority voting of each individual tree. This approach is mostly used to solve the decision tree's overfitting issue.

2.2.5 Evaluation Measures

Performance evaluation measures mainly deal with all performance measuring methods through which the performance of the model can be evaluated. The most commonly used evaluation measures are sensitivity, specificity, accuracy, F1_Score, etc.

3. Performance Result and Analysis

This section details the experimentation carried out on the collected dataset using different ML techniques. The result was analyzed by utilizing different evaluation measures.

3.1 Machine Learning Analysis

Table 3 shows the implementation results of different ML models in form of performance evaluation measures like Accuracy, Precision, Recall, and F1_score. A grid search CV method is followed to find the optimized hyperparameters to be passed to each technique. An optimized set of hyperparameters are depicted in Table 2.

Table 2: Optimal parameter by Grid Search Cross-Validation Method

Techniques	Optimized Parameter	Accuracy
RFT	Estimators=150	0.92
SVM	C=10, kernel=linear	0.89
K-NN	N_neighbours=7	0.872
DT	Max_features=5	0.90
NB	Solver=liblinear	0.85
LR	Var_ smoothing=0.23430358684	0.83

Table 2. Performance Evaluation of ML Model for Forecasting Severity level of Covid-19 Patient

Techniques	Accuracy	Precision	Recall	F1_Score
RFT	0.92	0.90	0.88	0.87
SVM	0.89	0.89	0.87	0.88
K-NN	0.872	0.85	0.84	0.84
DT	0.90	0.88	0.87	0.84
NB	0.85	0.83	0.83	0.82
LR	0.83	0.84	0.82	0.81

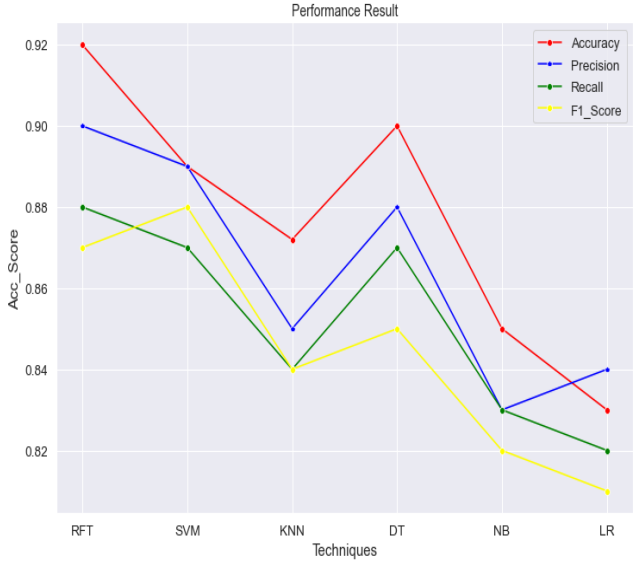

Fig.6: Experimental Result of Different ML Techniques

The graphical representation of different ML techniques is represented by Fig. 6. The result reveals that amongst all ML RFT achieved the highest accuracy rate of 92%.

4. Conclusion

One of the most popular technologies in the healthcare sector for controlling the Covid-19 pandemic epidemic is the Internet of Things. Smart gadgets that are IoT-enabled will support doctors, patients, health workers, and various health organizations to identify the person based on their symptoms and handle positive cases in a better and more efficient manner globally. IoT has given us the finest opportunity to quickly and effectively control the Covid-19 pandemic. This study examines how the suggested model performed using the Covid-19 symptom dataset. Six machine learning methods were tested in an experiment using a real COVID-19 dataset. SVM, RFT, NB, K-Nearest Neighbor (K-NN), Decision Table, and. All of these algorithms, with the exception of the LR, attained accuracy levels of more than 85%, according to the results. The four top methods would enable efficient and precise detection of possible COVID-19 situations. The model employing RFT approaches had the highest accuracy of all ML techniques, coming up at 92 percent.

Acknowledgment

I'm extremely grateful to Dr. Pooja Mittal for her input and assistance with the preparation of this paper

References

[1] J.Page., D.Hinshaw., B.McKay, "In Hunt for Covid-19 Origin, Patient Zero Points to Second Wuhan Market – The man with the first confirmed infection of the new coronavirus told the WHO team that his parents had shopped there," *The Wall Street Journal*, 26 Feb. 2021.

[2] C.Zimmer, "The Secret Life of a Coronavirus – An oily, 100-nanometer-wide bubble of genes has killed more than two million people and reshaped the world. Scientists don't quite know what to make of it,26 February 2021.

[3] M.A.Islam, "Prevalence and characteristics of fever in adult and paediatric patients with coronavirus disease 2019 (COVID-19): A systematic review and meta-analysis of 17515 patients," *PLOS ONE.*, vol. 16, no. 4, 2021.

[4] Ministry of Health and Welfare, Government of India. COVID-19 India. Accessed June 9, 2020.

[5] J.Siaya.Sania, M.A.Islam," Prevalence and Characteristics of Taste Disorders in Cases of COVID-19: A Meta-analysis of 29,349 Patients. Otolaryngology-Head and Neck Surgery, "vol. 165, issue 1, pp. 33-42, 2021 Jul, DOI: 10.1177/0194599820981018.

[6] New York Post, The Most Promising Coronavirus Breakthroughs So Far, from Vaccines to Treatments, 2020. April 8, https://nypost.com/2020/04/08/coronav virus-breakthroughs-how-close-are-we-to-a-vaccine/.

[7] M.P. Kelly," Digital technologies and disease prevention,"Am. *J. Prev. Med.*, vol. 51, no. 5, pp. 861-866, 2016. DOI: https://doi.org/10.1016/j.amepre.2016.06.012.

[8] P.M. Hlaing, T.R. Nopparatjamjomras, S. Nopparatjamjomras, "Digital technology for preventative health care in Myanmar,"*Digital Medicine*, vol.4, no. 3, pp. 117–121, 2018 https://doi.org/10.4103/digm.digm_25_18.

[9] Michael Safi "India's shocking surge in Covid cases follows baffling decline, " *The Guardian*. Retrieved 29 April 2021.

[10] "IndiaFightsCorona COVID-19". MyGov.in. Govt of India. 16 March 2020.

[11] Daily COVID-19 vaccine doses administered – India, Our World in Data. Retrieved 13 May 2021.

[12] Navita Mehra and Pooja Mittal, "Development of ML and IoT Enabled Disease Diagnosis Model for a Smart Healthcare System" *Int. J. of Comp. Sci.& Net. Sec.* (IJCSNS), vol. 22 No.7, July 2022, ISSN: 1738-7906.

[13] M.A. Alzubaidi, M. Otoom, N.Otoum, Y. Etoom, R.Banihani, "A novel computational method for assigning weights of importance to symptoms of COVID-19 patients,"*Artif Intell Med.* 2021 Feb;112:102018. doi: 10.1016/j.artmed.2021.102018.

[14] J. Medina, M. Espinilla, A.L. ´ García-Fern´ and, L. Martínez, "Intelligent multi-dose medication controller for fever: from wearable devices to remote dispensers,"*Comput. Electrical Eng.*vol.65, pp. 400–412, 2018.

[15] Y. Umayahara, Z. Soh, K. Sekikawa, T. Kawae, A. Otsuka, T. Tsuji, "A mobile cough strength evaluation device using cough sounds,"*Sensors, vol.* 18, 2018.

[16] D. Ichwana, R.Z. Ikhlas, S. Ekariani, "Heart rate monitoring system during physical exercise for fatigue warning using non-invasive wearable sensor, in: 2018,"*Int. Conf. on Info. Tech. Sys. and Innovation (ICITSI),* Bandung - Padang, Indonesia, pp. 497–502, 2018.

[17] B. Askarian, S.-H. Yoo, J.W. Chong, Novel image processing method for detecting strep throat (streptococcal pharyngitis) using smartphone, Sensors 19 (15) (2019) 3307.

[18] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, M. Ayyash, "Internet of things: a survey on enabling technologies, protocols, and applications,"*IEEE Commun*. Surveys Tutorials vol. 17, no. 4, pp. 2347–2376, 2015.

[19] E. Ahmed, I. Yaqoob, I.A.T. Hashem, I. Khan, A.I.A. Ahmed, M. Imran, A. V. Vasilakos, "The role of big data analytics in Internet of Things, "Comput. *Netw*., vol. 129, pp. 459–471, 2017.

[20] C. Mustafa, S. Askar, "Machine Learning for IoT HealthCare Applications: A Review", International Journal of Science and Business (IJSAB), vol. 5, no. 3, pp. 42-51, 2021.

[21] AAldahiri, B. Alrashed, W.Hussain,,"Trends in Using IoT with Machine Learning in Health Prediction System", *Forecasting*, vol. 3, pp. 181-207, 2021. DO: https://doi.org/10.3390/forecast301001

[22] Covid-19 Patient Monitoring Device based on LoRa - Arduino Project Hub

[23] H.Banaee, M.U.Ahmed., A.Loutfi.," Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges," *Sensors (Basel), vol.* 13, no. 12, pp.17472-17500, 2013.

[24] P.Dutta., S.Paul, and A. Kumar," Comparative analysis of various supervised machine learning techniques for the diagnosis of COVID-19," *Electronic Devices, Circuits, and Systems for Biom. App*., pp. 521–540, 2021.

[25] Z.Zhu., Z.Xingming, G. Tao, T.Dan., J.Chen. Li, I. Zhou, Z.Zhang, X. Zhou., D.Chen., H.Wen. and H. Cai.," Classification of COVID-19 by Compressed Chest CT Image through Deep Learning on a Large Patients Cohort,"*Interdisciplinary Sci.* pp.73-82, 2021.

[26] T.Yan, P.K.Wong, H.Ren, H. Wang., J.Wang, Y.Li." Automatic distinction between COVID-19 and common pneumonia using multi-scale convolutional neural network on chest CT scans,"*Chaos Solitons Fractals*, 2020.

[27] A. Darwish, A.E.Hassanien, M. Elhoseny, et al.," The impact of the hybrid platform of internet of things and cloud computing on healthcare systems: opportunities, challenges, and open problems," *J Ambient Intell Human Comput* 10, pp. 4151–4166 2019.