



ARTICLE

A Novel Edge-Assisted IoT-ML-Based Smart Healthcare Framework for COVID-19

Mahmood Hussain Mir^{1,*}, Sanjay Jamwal¹, Ummer Iqbal², Abolfazl Mehbodniya³, Julian Webber³ and Umar Hafiz Khan⁴

¹Department of Computer Sciences, Baba Ghulam Shah Badshah University, Rajouri, Jammu & Kashmir, India

²Department of Computer Sciences, National Institute of Electronics and Information Technology, Srinagar, Jammu & Kashmir, India

³Department of Electronics and Communication Engineering, Kuwait College of Science and Technology, Kuwait City, 20185145, Kuwait

⁴Department of Geriatric Medicine, Sher-i-Kashmir Institute of Medical Sciences, Srinagar, Jammu & Kashmir, India

*Corresponding Author: Mahmood Hussain Mir. Email: mahmoodhussain@bgsbu.ac.in

Received: 17 October 2022 Accepted: 23 March 2023 Published: 03 August 2023

ABSTRACT

The lack of modern technology in healthcare has led to the death of thousands of lives worldwide due to COVID-19 since its outbreak. The Internet of Things (IoT) along with other technologies like Machine Learning can revolutionize the traditional healthcare system. Instead of reactive healthcare systems, IoT technology combined with machine learning and edge computing can deliver proactive and preventive healthcare services. In this study, a novel healthcare edge-assisted framework has been proposed to detect and prognosticate the COVID-19 suspects in the initial phases to stop the transmission of coronavirus infection. The proposed framework is based on edge computing to provide personalized healthcare facilities with minimal latency, short response time, and optimal energy consumption. In this paper, the COVID-19 primary novel dataset has been used for experimental purposes employing various classification-based machine learning models. The proposed models were validated using k-cross-validation to ensure the consistency of models. Based on the experimental results, our proposed models have recorded good accuracies with highest of 97.767% by Support Vector Machine. According to the findings of experiments, the proposed conceptual model will aid in the early detection and prediction of COVID-19 suspects, as well as continuous monitoring of the patient in order to provide emergency care in case of medical volatile situation.

KEYWORDS

COVID-19; edge computing; framework; Internet of Things (IoT); machine learning (ML); network; symptoms

1 Introduction

Despite living in the twenty-first century, COVID-19 has claimed the lives of thousands of human beings due to the lack of an inadequate healthcare facilities. The world witnessed three waves from time to time since its outbreak in Wuhan City of China in November 2019 [1]. Coronavirus belongs



to the class of Coronaviridae that mostly infects the upper respiratory system of humans causing the illness from mild to severe [2]. Coronavirus is a positive-stranded RNA-based virus that replicates itself very fast and spread from human to human mostly through sneezing [3]. RNA-based viruses can easily get mutated and change their properties to make new variants. Coronavirus has mutated several times since its first report as an alpha variant [4]. Till now the world has witnessed many variants like alpha, beta, gamma, delta, and recent ones as Omicron [5]. COVID-19 is a novel life-threatening virus that has infected at least 42.6 Crore and has taken 58.9 Lakh lives worldwide so far [6]. COVID-19 has impacted almost every field such as the economy, trade, industries, education, etc., due to the shutdown [7]. India has imposed lockdowns thrice since November 2019, to decrease the mortality rate and spread of the virus [8]. Unfortunately, the world is battling to date for effective treatment and vaccine that is not available yet [9]. The nature of the coronavirus is not completely known because of the novelty of the virus and the limited availability of the data [10]. Currently, the only way to tackle the pandemic is to follow standard operating procedures (SOP) such as social distancing, face masks, and other methods to minimize the outbreak and mortality of COVID-19 [11]. Medical experts with the help of the government are giving immunity boosters to the general public so that they can resist the coronavirus infection. From the technological side, the technologies like IoT, AI, ML, and Deep Learning integrated with traditional healthcare can provide better solutions to tackle the COVID-19 deadlier pandemic [12].

Due to advancements in technology, the healthcare system needs to be renovated to cope with the speed of the modern era. The traditional healthcare system is a preventive system in which the first disease is diagnosed then action is taken. This type of healthcare system is not an efficient way to tackle the disease or any pandemic situation like COVID-19 [13]. Integrating modern technology in the traditional healthcare system will make it a preventive and proactive healthcare system. Detecting diseases in the initial stage is only possible using novel technologies like artificial intelligence, machine learning and computer vision. Skin disease can be diagnosed using image detection and classification techniques in its initial phases [14]. The healthcare system is a time-sensitive application in which the required service should be delivered in real-time to preserve the criticalness of an application [15]. The data collected in real-time should be processed as soon as possible to provide the required services in real-time [16]. Healthcare is a time-sensitive application area in the time frame of data and the age of data plays an important role in providing time-critical and timely services to patients [17]. Sensor technology advancements have enabled the development of new systems that can change the way humans are living. The Internet of Things (IoT) is a network of physical items (sensors, wearables, medical sensors, and so on) that can sense the environment and communicate its parameters [18]. IoT is a novel technology that can be integrated with other technologies to create autonomous systems, such as machine learning and artificial intelligence. Real-time health parameters can be collected via IoT and analyzed using machine learning models to determine the patient's status [19]. The data that is collected with the help of IoT devices, healthcare sensors, wearables, and etc. is analyzed in different places based on the requirements of the application. Healthcare is a time-sensitive application, the computation should be performed close enough to the data sources so that the vital sensitive information can be retrieved and sent to caregivers for necessary action in real-time. Blockchain based technology can be incorporated into IoT-based systems to secure the communication channels from illegitimate users [20]. IoT technology has proven to be the most advantageous technology for providing various solutions to tackle the COVID-19 pandemic. Tracing contacts of COVID-19-positive patients can be identified using IoT devices so that suspected persons can isolate themselves to break the chain of infection spreading [21]. The IoT can be beneficial in detecting and monitoring COVID-19 suspects to eliminate the spread of infection [22]. Data is collected using IoT devices, then

transferred via a gateway to the cloud for analysis and computed results are communicated back to take necessary actions. High-end analytics is performed by deploying different machine learning techniques to predict or forecast the future trend [23]. The cloud provides services on demand but it has some inherited drawbacks such as high latency, response time, jitter, and others [24]. The healthcare system is a time-sensitive application in which time plays an important role to make decisions. The IoT-based systems integrated with edge computing eliminate the problems of the cloud by delivering the services locally near the data source [25]. Edge computing is distributed environment in which the services such as computing, and storage being provided in one single hop [26].

1.1 Motivation and Novelty

The healthcare system globally collapsed due to the COVID-19 pandemic, as it was not ready for such a huge burst of patients that has been witnessed in the last two years. The healthcare system is not fully equipped with modern technology that can combat pandemic situations. Millions of people are dying every year due to inefficient healthcare facilities according to a WHO report¹. Half of the population of the world lacks access to basic and essential medical facilities. The integration of technology in the traditional healthcare system can somehow be modified to give healthcare services to more public in an efficient manner. COVID-19 has claimed thousands of lives in the last two years because of the unavailability of proper healthcare services. COVID-19 is a communicable disease, hence IoT-based technologies can be utilized to restrict the virus from spreading. IoT-based healthcare systems can provide better options for continually monitoring patients from any location. IoT, machine learning, and edge computing can be combined to create new proactive and preventive healthcare solutions. IoT is an internetwork of physical objects connected via communication technologies used for sensing and monitoring the environment. Machine learning is a subclass of artificial intelligence that is used for data analysis to get insights from the data. The prediction of the future can be made using machine learning models to take protective measures in advance to decrease the impact. the healthcare system is a time-sensitive application area in which the requests should be resolved in real-time to make intelligent decisions. Cloud computing suffers from various disadvantages like high latency, delayed response, jitter, and others, that are unacceptable in smart healthcare systems. Edge computing eliminates some of the disadvantages of cloud computing by locally providing the services in real-time. The IoT-based healthcare setups integrated with machine learning and edge computing will be more helpful to tackle a pandemic like the situation of COVID-19.

As the healthcare system is still in its infancy, a small contribution to this field will have a significant impact. The purpose of this research is to describe an innovative healthcare system that can detect and make accurate predictions about possible COVID-19 carriers at an early stage in order to stop the spread of the disease and reduce mortality rates. To collect symptomatic data in real-time, medical Internet of Things devices (M-IoT), sensors, and wearables are being utilized. The Internet of Things devices have three primary advantages over similar products, one of which is the ability to perform continuous monitoring from any location and at any time. Thirdly, the data can be collected within a particular time range, and secondly, they can be collected at regular intervals. On the edge of the network, the symptomatic data that is received from healthcare sensors are examined in order to reduce the problems that are caused by the cloud. Let's say that in the instance of COVID-19, real-time symptoms are gathered and processed on the edge in order to locate the potential suspect in a single hop. The edge has limited resources; thus, the power of the cloud is leveraged when it comes to the training of machine learning algorithms. The ML models are running on the edge after being trained on the cloud. The proposed framework makes contributions by: 1. Detecting and predicting the

¹<https://covid19.who.int/>

COVID-19 suspect in the initial phases; 2. Analyzing the symptomatic data on the edge of the network by running trained machine learning models; 3. Disease diagnosing and generating emergency alerts with respect to any untoward medical volatile situation; 4. Sending the analyzed data to the cloud for further analysis in order to better understand the disease; and 5. Maintaining the healthcare database in preparation for the future.

1.2 Research Contribution

The research contributions are enumerated below:

1. A novel edge-assisted IoT-based framework for early identification and prediction of COVID-19 suspects has been proposed.
2. A novel COVID-19 primary dataset has been used for experimental purposes employing various classification-based machine learning models. The proposed model has been trained using a novel primary symptomatic dataset collected from the Sher-i-Kashmir Institute of Medical Sciences (SKIMS), Srinagar, Jammu & Kashmir, India.
3. The proposed framework is validated by conducting an experimental study using various ML models viz: Support Vector Machine (SVM), Logistic Regression, Naïve Bayes, Neural Network, Decision Tree, and Random Forest. Accuracy, Precision, Recall, F-Score, Root Mean Square Error, and Area under Curve Score are among the performance measures used to validate the proposed models.
4. The proposed models were validated using k-cross-validation to ensure the consistency of the models. Based on experimental results, our proposed models have achieved well in terms of accuracy with the highest of 97.767% by Support Vector Machine.
5. Based on the findings of experiments, the proposed conceptual model will aid in the early detection and prediction of COVID-19 suspects, as well as continuous monitoring of the patient to provide emergency care in case of a medical volatile situation.

2 Existing Work

The Healthcare System is in its infancy and needs to be renovated to cope with the speed of modern technology. Numerous IoT-based healthcare setups have been proposed for various diseases. IoT is a system made up of a network of things connected by communication technologies. Healthcare is an IoT application where IoT devices are deployed to collect and monitor patient data remotely. Haggi et al. have systematically reviewed the IoT in healthcare of 146 articles of five years up to 2020. Authors have categorized the literature on IoT in healthcare into five categories. The categories are divided based on hardware and software, such as Sensors, Communication, Security, Resources, and Applications. The limitations in each field have been highlighted. Lastly, the authors have detailed the future scope and challenges in the healthcare IoT [27]. Habibzadeh et al. have surveyed the up-to-date literature regarding the Internet of Things in healthcare. They have highlighted various application areas of IoT healthcare, such as cardiac monitoring, sleep monitoring, BP monitoring, etc. Apart from this, they have detailed a general architecture of the cloud-based healthcare framework. The framework is comprised of several layers, such as the sensing layer, cloud layer, and analytics layer. Lastly, the paper concludes by highlighting some future issues that must be taken into consideration to make IoT-based healthcare systems more powerful [28]. Bhardwaj et al. have proposed the IoT cloud-based smart COVID-19 monitoring system. The proposed model is comprised of sensors for monitoring the oxygen level, temperature, pulse, and heart rate of a person. The data collected by sensors is sent by edge gateway to the cloud for analysis and alerts are generated in case of emergency [29].

Mondal et al. have examined current approaches for combating the COVID-19 epidemic. The state of art technologies such as IoT, machine learning, artificial intelligence, and others are helping to combat COVID-19's influence during the epidemic. IoT with other technology like artificial intelligence could have a great impact in detecting, predicting, monitoring, and mitigating the outbreak of COVID-19 infection. Finally, the researchers have explored the impact of the technologies in the post-pandemic era [30]. Verma et al. offered a cloud IoT-based mobile healthcare architecture for disease diagnosis and monitoring. ML techniques are used to classify the condition into four severity levels. The accuracy, sensitivity, specificity, and f-measure are all used to compare the performance of machine learning models [31]. Sood et al. presented an IoT-fog-based smart framework for identifying and controlling heart attacks. The authors are collecting the data using sensors on remote sites and analyzing it in fog using an artificial neural network. The information is shared with the caregivers and responders through the cloud. Various ML techniques have been used on different datasets to measure the performance of the proposed system [32]. Verma et al. introduced an IoT cloud-based student monitoring mobile healthcare platform that uses health parameters to compute disease severity. Machine learning techniques were utilized to test the proposed system using a dataset of 182 suspected persons. The ability of the proposed machine learning techniques is computed by comparing the various measures like accuracy, sensitivity, specificity, and response time [33]. Bukhari et al. have proposed different machine learning techniques to predict the vaccine for the Zika virus (ZIKV) T-cell epitope. The T-cell epitope-based vaccines have high efficacy and can be developed in less time than vivo-based approaches. The researchers used the IEDB (Immune Epitope Database and Analysis Resource) repository dataset, which contains 3519 sequences, 1762 of which are epitopes and 1757 of which are not. The results have been analyzed using distinct performance measures like accuracy, specificity, sensitivity, precision, F-score, and Gini coefficient [34]. Kavitha et al. have proposed a real-time context-aware framework for healthcare monitoring based on a neural network. The researchers have used machine learning techniques for extracting of features and for classification purposes. Apart from this, the authors have also used optimization techniques to get the best optimal solution. Based on various performance measures such as accuracy, response time, etc. comparative analyses have been conducted [35]. Nasajpour et al. have examined the importance of IoT technologies in preventing the COVID-19 epidemic. The authors have identified IoT-based system architectures, apps, platforms, and other features that can help diminish the impact of the COVID-19 outbreak [36]. Singh et al. have evaluated the implications of IoT technology during the COVID-19 era. The effectiveness of IoT in remotely monitoring the COVID-19 patient to reduce the spread rate was reviewed by the authors. IoT is a prospective technology that can collect the suspect's symptoms in real time and deliver effective treatment based on the symptoms [37]. Ajaz et al. have explained the potential of IoT and ML in tackling the COVID-19 outbreak. Researchers have developed a six-layer architecture and have elaborated the working of each layer with respect to COVID-19. Apart from this the paper also introduces the concept of unmanned aerial vehicle (UAV) for the effective contact tracing [38]. Ahmadi et al. have conducted extensive research on IoT-fog-based healthcare systems. Healthcare is a time-critical application area and it needs responses in real-time so fog computing in a distributed environment is an idle solution to preserve the criticalness of healthcare [39]. Verma et al. have presented an IoT-fog-based monitoring framework in a smart home environment. The authors used multiple machine-learning approaches to validate the framework using a dataset of 67 patients. The model's performance was assessed using a variety of metrics, including accuracy and response time [40]. Nguyen et al. have detailed out the impact of AI (Artificial Intelligence) techniques that can tackle pandemic situations. A lot of application areas of AI have been discussed such as text mining, image classification, natural language processing, and data analysis. The authors have also discussed the open databases that are spacing the COVID-19 datasets. Lastly, the paper identifies

research gaps and presents future directions to provide solutions [41]. Yang et al. have presented a framework for the COVID-19 monitoring system based on Epidemic Prevention and Control Center (EPCC). The framework internally uses the mixed game-based Age of Information (AoI) optimization technique. The authors have used AI bots on the edge of the network and have proposed algorithms for channel selection [42]. Elbasi et al. have presented an overview of IoT and machine learning techniques that can collectively be employed in countering the COVID-19 pandemic. The authors have detailed the classification as well as clustering techniques of machine learning concerning COVID-19 [43]. Otoom et al. proposed a COVID-19 identification and monitoring architecture based on the Internet of Things. The proposed framework uses IoT for data collection and analyzes incoming data by applying various ML techniques. The system is comprised of various components such as a data collection unit, data analytics submodule, cloud system, etc. The researchers have proposed various ML models to identify the COVID-19 suspect. The authors have used the dataset of 1476×7 records out of which 854 are confirmed positive and 622 are not confirmed. Various performance measures, including accuracy, precision, and recall, have been used to validate the machine learning model's performance. The five machine learning models have achieved above 90% accuracy in predicting the potential suspect [44]. Imad et al. presented a systematic review of IoT-based ML and DL models to diagnose and predict the COVID-19 suspect. The authors have also identified issues and presented a summary for future directions [45]. Castiglione et al. have presented the role of IoT in countering COVID-19 and have developed a framework for identifying and observing COVID-19 patients. The proposed framework based on cloud technology made up of four modules: data gathering module, data analytics module, health center, and healthcare professionals. The framework has been validated using different machine learning techniques [46]. Ashraf et al. presented a smart system for monitoring the patient vital signs such as blood pressure, fever, etc., remotely based on edge computing. The proposed system generates warning signals in case of an emergency by detecting irregularities in parameters [47]. Mir et al. have used several machine learning approaches to develop an IoT cloud-assisted system for identifying and predicting COVID-19 suspects. Authors have validated the proposed framework by proposing various ML algorithms, and the results have been compared with previous works [48]. Greco et al. presented a state of art literature on IoT-based medical systems from centralized technology to distributed technology. The survey presents the issues of cloud technology with respect to the healthcare system. The cloud has inherited disadvantages like delay, response time, latency, and security issues so shifting from the cloud to the edge is an important move that is to be taken [49]. Rahman et al. have presented a medical-IoT edge computing-based framework to tackle the COVID-19 pandemic. Edge computing has various advantages over cloud computing such as low latency, high response time, and improved security and privacy. The authors of this research article have created an edge-based system. to collect the system of suspects and analyze them on the edge to generate emergency alerts in real-time [50]. Dewanto et al. have presented a heart rate monitoring system based on edge computing. The proposed architecture is comprised of several modules such as a data acquisition module, analysis module and user system. The results have been generated such as in terms of time, latency, memory usage, etc. using simulation methods [51]. Malik et al. have proposed different ML techniques to analyse the COVID-19 symptomatic data. Authors have used SVM, logistic regression, k-neighbor, decision tree and naïve bayes to determine the impact of COVID-19 illness. In most of the models they have achieved above 92% in terms of accuracy but naïve bayes and decision tree outperform by achieving 93.7% accuracy [52]. Mohammed et al. have devised a novel IoT-based smart helmet to identify the COVID-19 suspect. They have taken a temperature as an attribute because it is the most popular symptom of COVID-19 suspects [53]. Shabaz et al. have developed a network model based on link prediction technique, by which the authors study the different comorbidities of patient. Suppose if a patient dies who was suffering from pneumonia and was also COVID-19 positive. In this

case the actual cause of death is unknown whether a patient dies due to COVID-19 or pneumonia. The paper uses the link prediction technique to predict the future possible disease of the patient [54].

The healthcare system is still in its infancy lot of work is needed to be done to make it more intelligent to cope with the modern pace. Table 1 below summarizes the findings of the literature. The table is comprised of several rows and columns, in which rows represent the various latest works and columns represent the tools and techniques for smart healthcare.

Table 1: Summary of findings

Reference	Framework used	Technology used	Techniques adopted	Healthcare system
[29]	Cloud	IoT-based sensors	N/A	COVID-19 monitoring system
[31]	Cloud	IoT-based sensors	Various ML techniques	Disease diagnosis system
[32]	Fog-Cloud	IoT devices	Neural network	Identification system
[33]	Cloud	IoT devices	Basic ML techniques	Monitoring system
[44]	Cloud	IoT devices	Various ML techniques	COVID-19 identification and monitoring system
[50]	Edge-Cloud	IoT devices	ML/DL techniques	COVID-19 detection system

3 Materials and Methods

3.1 Internet of Things for Smart Healthcare

The Internet of Things (IoT) is an internetwork of connected physical items with onboard communication, computation, and other technologies to form a network of smart objects. IoT technology provides smart services to users anytime, anyplace to ease their lives. IoT systems provide a better environment and smarter applications in terms of reduced cost, time, and energy. IoT refers to a network of interconnected things that can perceive data and transmit it to others so that actions can be taken without the need for human interaction. There are three layers to the IoT architecture: perception, network, and application layer [55]. The sensing layer is comprised of sensors for data collection, and actuators for taking actions based on sensed data. The sensors are very small in size so they limited computation power and are battery-powered. Sensors are mainly made up for data collection purposes and have inbuilt communication technology to send the data to other devices for computations. BLE, near-field communication (NFC), ZigBee, Bluetooth, IEEE 802.15.4, RFID, and other low-power communication technologies are designed for low-power sensors [56]. The network layer is meant for send the data for analysis and it is also responsible for providing solutions to heterogeneous objects. The IoT devices are from different vendors, there is no common standard of communication of these objects, and the transport layer provides these services. Besides these services transport layer provides other high-level services like resolving the scalability issues, service discoveries, and security & privacy of IoT devices. IoT devices are generating huge data, and to get insights from the data, ML and DL techniques or algorithms can be used to get better possible results [57]. Various computing technologies can be used for analyzing the IoT data like cloud computing, fog computing,

or edge computing, based on the need and nature of an application area [58]. Cloud computing provides high-level computing power in a centralized fashion. The cloud computing architecture suffers from various disadvantages like high latency, high response time, and other network-related issues. Cloud computing is not an ideal solution for time-sensitive applications like healthcare, autonomous vehicle, robotics, traffic control, and others. Fog computing is another architecture that provides the services like scheduling servers, resolving scalability issues, and providing data filtering services. Fog computing provides services such as computation, storage, and network in a limited manner. Edge computing is a distributed architecture in which services are provided locally in a decentralized fashion. The data that is gathered from sensors are processed on local edge servers to minimize the cons of cloud technology like latency, response time, and other issues. Data is stored locally to enhance data security & privacy, aggregation & filtering, and to provide time-sensitive information in real-time.

3.2 Machine Learning for Smart Healthcare

Machine learning is the subject of computer science and a subset of AI, that allows machines to learn explicitly without having to be programmed. By 2025, there will be 27 billion connected devices and a tremendous amount of data will be generating these devices [59]. To tackle this data machine learning will provide better solutions to get insights from the data. Machine learning models need data to predict future events by analyzing historical events. Machine learning models can replace traditional processes with automated systems that uses statistically derived actions [60]. Machine learning is classified into three different types based on input data: supervised, unsupervised, and reinforcement. The healthcare problems are both related to classification or clustering either the problem is to classify it into different levels or discover hidden patterns. In supervised learning, an algorithm takes an input vector with a predefined label to train the model to enable it to label the new data. The applications where the output labels are finite and discrete sets are known as classification problems and where output labels are continuous are known as regression problems. In unsupervised learning, a model is fed with only input vectors to detect different patterns on its own and to make different possible clusters. The next incoming data is put into that cluster to which its similarities are close enough, these problems are known as clustering. Reinforcement learning is gaining experience from the previous action taken [61]. Classification refers to the labeling of the incoming input data into a class based on the training dataset. There are various classification machine learning techniques such as K-Nearest Neighbor (KNN), Naïve Bayes, and Support Vector Machine (SVM) [62]. Regression is supervised machine learning in which the output vector is a continuous variable such as the salary of a person. There are various regression techniques the simplest one is linear regression. Clustering refers to grouping the unlabeled data into a different possible number of groups. The most basic clustering technique is K-Means clustering in which the input vector is divided into k-clusters with different properties.

3.3 Computing Technologies for Smart Healthcare

IoT-based applications are producing a lot of data, to process the data an efficient computing framework is to be selected based on the nature of the application. There are various types of applications such as data storage, entertainment, and social networks that use the cloud for data processing. For some applications, data should be processed near to data source to give fast responses. Cloud computing architecture is a centralized platform where data centers receive the data, and analyze and process it according to the need of an application. Cloud computing provides high computing power and other high-end resources on-demand [63]. Mainly three types of services are provided by cloud computing they are as under. Infrastructure as a Service (IaaS): where the

infrastructure is taken as service from the cloud-like hardware, servers, or network. Software as a Service (SaaS): in this, the software is taken as a service, it is a distributed software system and is hosted by service providers to users through the internet. Platform as a Service (PaaS): in this service, the services are provided on rent or based on pay per use. For some applications cloud computing is not an idle choice for processing, because of its inherited cons like latency, response time, and other issues. Cisco coined another computing environment named “fog computing”, which is the layer below the cloud environment [64]. Fog computing provides some low-level computations such as scheduling of processes, limited computations, data filtering, and storage services for cloud data centers. Fog computing is beneficial in providing application-specific services by data filtering below the cloud to minimize the issues of the cloud computing environment [65]. Some of the applications cannot tolerate the issues of cloud computing like latency, response time, bandwidth, and other issues. A smart healthcare system is one of them in which time plays an important role. Healthcare is a time-sensitive application in which the required services should be serviced in real-time to preserve the criticalness of the application. Edge computing is a layer above the data source in an edge of the network to provide low-level services in real-time [66]. It is a distributed computing environment that provides limited services such as storage, processing, computation, and other services are provided within a local network. Edge computing provides services locally with low latency, high response time, and other services like time-sensitive information in real-time. Edge computing is an idle solution for time-sensitive applications like healthcare, autonomous vehicle, traffic signaling, etc.

3.4 Methodology Adopted for Data Collection

Coronavirus disease (COVID-19) is an infection that affects a person’s upper respiratory tract. COVID-19 was initially reported in November 2019 in Wuhan, China, and has since spread to practically every corner of the globe. COVID-19 has been declared a global pandemic and medical health emergency by the World Health Organization (WHO). The healthcare system was not ready for such a medical explosion, that COVID-19 has caused in the world. The healthcare system collapsed due to the huge number of fatalities and the shortfall of medical facilities for patients in every part of the globe. The hospitals were full of COVID-19-positive patients and people were dying due to the shortage of ventilators, oxygen concentrators, and other essential medical items. To combat this pandemic traditional healthcare systems are not enough to tackle an emergency like one COVID-19. Researchers from every field are working by leaps and bounds to cope with the current deadly situation. WHO has made this an open problem so that everyone can contribute to this underdeveloped field. The COVID-19 data is open and is available on different repositories like Google, WHO, CDC, NIH, COVID, and other databases & government sites. The available datasets are simple deaths, positives, and other metadata. Our research is aimed to develop a preventive and proactive IoT-ML-based framework, that will predict and identify the COVID-19 suspect in its initial stages. Based on sensors and other IoT devices used to collect symptoms from suspected COVID-19 patients, our model will make predictions about the suspects. To train our model, the dataset should be symptomatic that are primary for a COVID-19 suspect like Fever, Temperature, Cough, Fatigue, Breathlessness, Oxygen level, and other parameters. The primary symptoms are identified by WHO², CDC³, NIH⁴, and Medical Institutes and published by researchers. The symptomatic dataset is not available, to the best of our knowledge because of the novelty of coronavirus disease. To make the proposed model productive and usable the real novel primary data has been acquired from the Sher-i-Kashmir Institute of Medical Sciences (SKIMS), Srinagar, Jammu & Kashmir, India with due approval and proper ethical process. The

²https://www.who.int/health-topics/coronavirus#tab=tab_3

³<https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html>

⁴<https://covid19.nih.gov/covid-19-topics>

SKIMS⁵ Institute is one of the premier institutes, and of national importance that has served COVID-19 during the pandemic. The list of symptoms given by WHO, CDC, and NIH with other symptoms that were known to be present in patients visiting for consultations to SKIMS has been drafted in the presence of a team of doctors from the COVID-19 management department. According to the draft the symptoms were collected from both IN patients as well as from OUT patients with the help of doctors. Our proposed system is based on machine learning, to train an ML model abundant data is needed so that after deployment it should predict and classify accurately. The methodology adopted in the process of data collection and data preprocessing to get the required symptomatic dataset is depicted in Fig. 1.

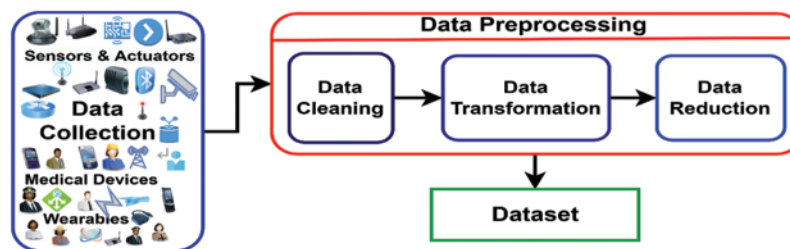


Figure 1: Preparation of dataset

3.5 Symptomatic Dataset Formation

The next phase is data preprocessing, which comes after data preparation. Data preprocessing is comprised of many sub-steps, in which the dataset is made ready for the model for the training phase. The first step of data preprocessing is data cleaning. The process of detecting the values of erroneous values and eliminating or correcting them is known as data cleansing. In our collected dataset, the analysis is made to check the missing or wrong values to correct them using various statistical methods. The missing or incorrect values were removed by deleting the rows of the dataset to make it complete for our model. The second step is data transformation also known as normalization, in which data is brought to one single format so that it can easily be interpreted by a machine learning model. Suppose in our case, the fever, pulse, oxygen level, etc. as attributes were collected in numerical form, and other symptoms of the dataset were collected in categorical form. To make the data acceptable for the machine learning model, it must be transformed into a single format. The attributes fever, pulse, oxygen level, etc. were converted to their categorical equivalent values to keep the semantics unchanged. The conversion was made after consulting the team of experts i.e., doctors of the institute. The third step is data reduction, which is the process of reducing the size of the dataset either by dimensionality reduction techniques or by feature engineering to optimize the use of storage and time of the system. In the case of our collected dataset, some of the attributes were dropped or merged keeping the originality of the data intact. Like shortness of breath with breathlessness, fever with temperature, etc. After applying the above steps, the dataset is now made ready for training the model with the following attributes as shown in Table 2. The dataset contains 7015 rows and 16 columns with the binary value either 1 (one) or 0 (zero). The column value one (1) indicates that the symptom is present, while zero (0) indicates that it is not present. The class label for each row is also in binary format either 1 or 0, in which 1 represents the COVID-19 positive and 0 represents the COVID-19 negative.

⁵https://www.skims.ac.in/index.php?option=com_content&view=category&id=141:statistics&Itemid=275&layout=default

Table 2: Attributes of dataset

Fever	Cough	Rhinitis	Breathlessness
Fatigue	Loss of Appetite/Taste	Oxygen saturation	Vomiting
Sore throat	Myalgia	Severity of illness	Contact/Travel History
Diarrhea	Comorbidity	Survival	Result

4 Proposed Edge-Assisted Healthcare System

Traditional healthcare systems are cloud-based, in which data is sent for processing to the cloud via the gateway, which is near the data source. Cloud computing has ruled the market during the last two decades, in this technology processing is done in a centralized fashion. There are various cons associated with this technology, such as latency, jitter, response time, and other related issues that are unbearable to time-critical applications. To eliminate these cons, somehow, an edge-assisted system has been proposed. In this proposed system, an edge computing technology has been incorporated on the network’s edge near the data source. Edge computing is a computing paradigm in which processing is done close to a data source, i.e., near the sensors, to eliminate the cons of cloud computing. The proposed system comprises two parts: the proposed architecture and the second is proposed framework. The first part of the proposed system describes the system’s working in a layered fashion, and the second with the help of different modules. The pictorial view of the proposed architecture is presented in Fig. 2: Layered Architecture of Proposed System and the pictorial presentation of the proposed framework is formulated in Fig. 3: Proposed Edge Assisted Smart Healthcare Framework.

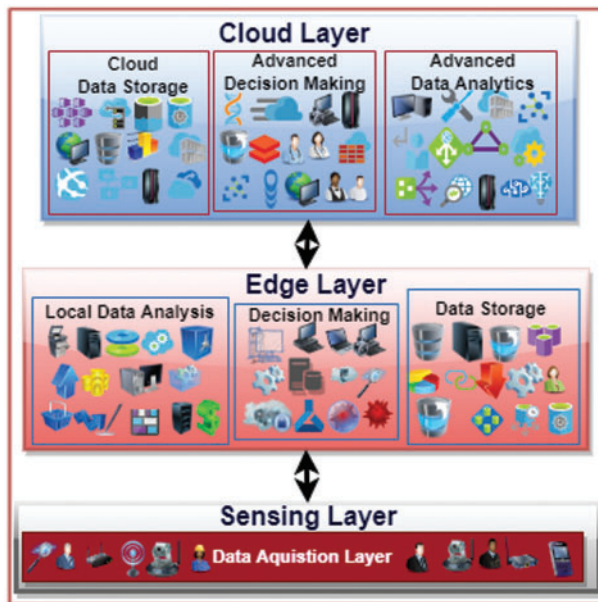


Figure 2: Layered architecture of proposed system

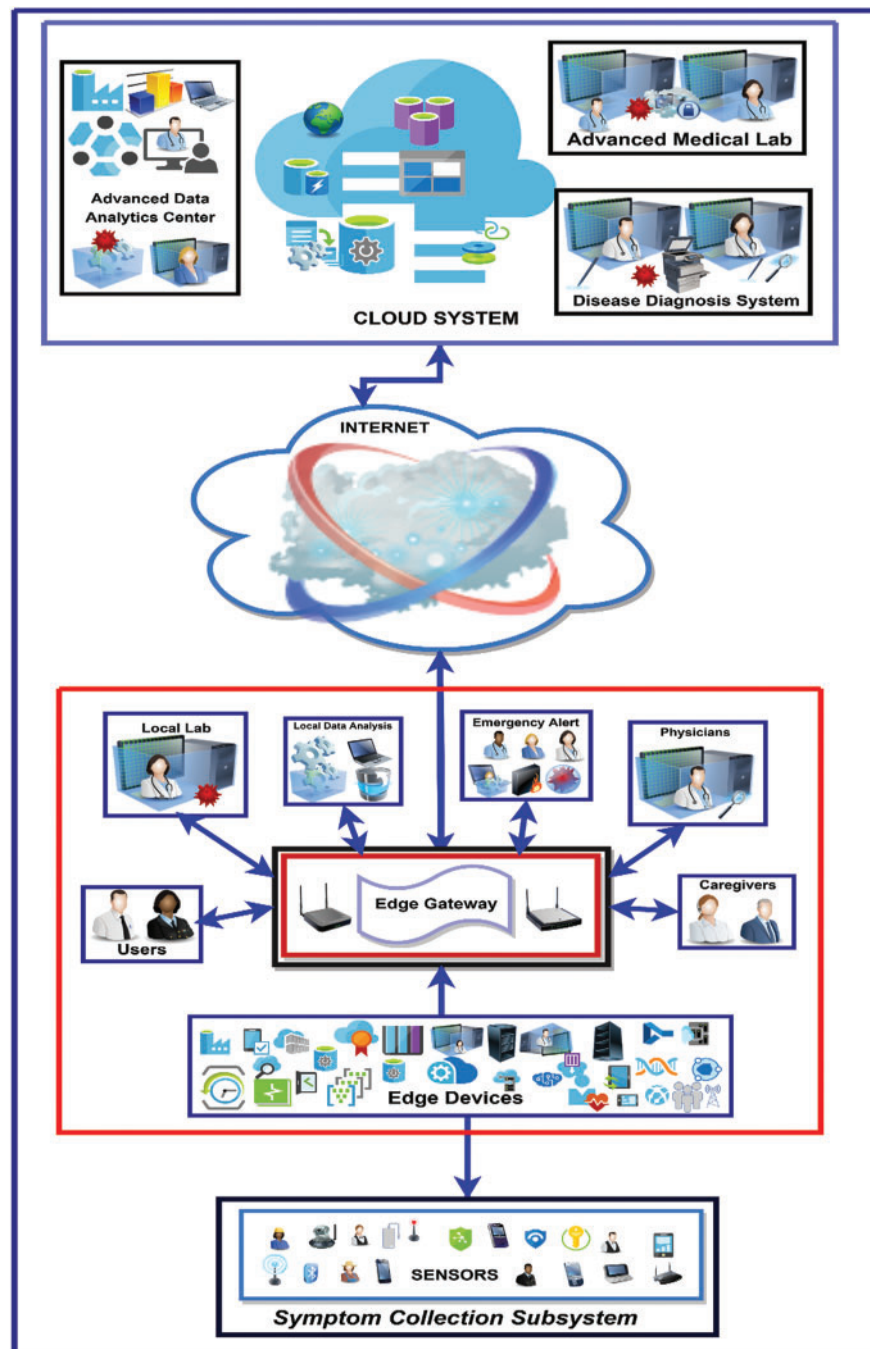


Figure 3: Proposed edge-assisted smart healthcare framework

4.1 Proposed Architecture

This section explains the proposed system's IoT-based edge-cloud three-layer architecture. The proposed architecture is divided into three logically connected layers: Sensing Layer (Perception Layer), Edge Layer, and Cloud Layer. The data is generated by sensors and wearables deployed in

the sensing layer of the proposed architecture. The sensing layer is meant for collecting data (i.e., symptoms). The sensors that are used in our case are a temperature sensor, pulse oximeter sensor, inertia sensor, audio sensor, blood pressure sensor, and other bio-medical sensors like EEG, ECG, etc. Runny nose, loss of appetite, vomiting, and other non-sensor-based parameters are taken from apps like Aarogya Setu. Other parameters such as travel history, contact history, etc., will be taken from a person's phone location register. This layer is only made for collecting symptomatic data and sending it layer next to it. Sensors are battery-powered miniature devices, so it is better to do any processing there. The second layer is Edge Layer: in this layer, our system has low-powered systems capable of doing low-level computations and running trained machine-learning models. In this layer, the alert is generated if the suspected has values of certain parameters high or low based on the required level. Suppose a person has a fever, a warning is generated, and an alert is sent to caregivers and responders. After collecting the other parameters, the trained machine model will be run on the collected parameters to further analyze the suspect for the possible presence of COVID-19 in a suspect. The edge layer is further divided into three parts: the local data analysis layer, the local decision-making layer, and the local storage layer. In this layer, the incoming symptoms are locally analyzed on edge, decisions are made, alerts are generated for a suspected, and local storage is also on the edge layer of the proposed system. The cloud layer, the top layer of the proposed three-layer architecture, is providing high-end computations, like advanced data analytics, advanced decision-making, and cloud storage. The high-end ML models are trained on this data to draw better insights. The data received from the edge is in aggregated form because some of the logic is also implemented on the edge layer. The edge-based architecture has many advantages over cloud-based architecture. The Healthcare system is a time-sensitive application, it needs real-time responses, so the cloud is not a good option. Edge computing is a novel technique that overcomes the drawbacks of cloud computing. Because the data is generated near the edge, it is preferable to process it near the data source to get real-time findings. The cloud layer provides high-end services like storage and other health experts use the stored data to analyze the disease better. The proposed architecture is presented in [Fig. 2](#).

4.2 Proposed Framework

IoT-based systems are preventive and proactive rather than reactive as are traditional healthcare systems. This section explains the IoT-based Edge-assisted proposed framework. The framework is based on Edge Computing, so the cons of Cloud Computing are eliminated. The proposed framework's main goal is to identify and predict the COVID-19 suspects in its initial phases. The framework can also be used to limit the spread of disease by detecting a suspect in real-time and sending the suspect's data for further investigation. This framework is unique in that it generates alerts at the network's edge to avoid delays and latency between the cloud and the user. The framework is comprised of three different subsystems: symptom collection subsystem, user-centric subsystem, and cloud subsystem as shown in [Fig. 3](#).

4.2.1 System Collection Subsystem

IoT technology has opened new doors for various application areas, smart healthcare is one of them. With the help of IoT, a person can easily be monitored from anywhere, anytime. In our proposed system, the system collection subsystem is responsible for collecting data, i.e., symptoms. The collected symptoms are of two types based on the collection technology used: sensor-based and application-based. Numerous sensors and wearables are deployed to collect real-time symptoms, and others are collected from applications. The sensors are forwarding this data to another subsystem above it for

analysis. The sensors are low-powered devices with onboard communication technology such as BLE, ZigBee, Bluetooth, and other low-power communication technologies.

4.2.2 User-Centric Subsystem

The second important module of our proposed system is the user-centric subsystem, which comprises several sub-modules. The data received from the subsystem below it through different sensors and applications are analyzed. Various services like data aggregation, data security, data redundancy, and other services are provided by this module. Most of the logic of the proposed system is implemented in this module, the data sent by the sensors are checked for any severity in real-time. In the case of COVID-19, the first parameter that is checked is body temperature, the temperature sensor's reading is checked by a machine learning algorithm model deployed on edge of the network. If the temperature is above the normal level, the alert is generated and other essential parameters are taken to check the possibility of COVID-19 suspects. Similarly, the proposed system works with COVID-19-positive patients in case of emergency. Suppose a COVID-19-positive patient is isolated or home quarantined, our proposed will keep track of oxygen saturation, if it falls below the threshold value, an emergency alert will be generated and sent to different participating parties. The novelty of the proposed system is monitoring and getting the current status of the COVID-19 suspected from the local database i.e., from edge anytime in real-time without contacting the cloud services. Considering the importance of time criticality in the healthcare systems, the proposed framework at its core incorporates edge as a base technology, which helps in providing decisions in real-time. Keeping all the cons of the cloud in consideration, the proposed system is based on edge computing to eliminate the cons of the cloud.

Edge Gateway: This sub-module is used for communication with an outside network and with the sensors and wearables. Due to the heterogeneity of devices and different communication technologies used by different sensors and wearables, there is a need for a standard device that can translate to and from different technologies. An edge gateway is used to communicate with other subsystems in its vicinities such as users, caregivers, local data analysis, and others. The edge gateway uses low-power communication technologies with the module below it and other high-power communication technologies such as cellular networks etc., through the internet with the cloud.

Users: The general public comes under the category of users. Our proposed system collects the symptoms as input from users and compares the results with stored values to see if COVID-19 is present in the first phase. The proposed system checks different parameters with predefined values, such as fever, oxygen saturation, pulse, blood pressure, etc. If these parameters are in the normal range, the person is treated as normal, otherwise, COVID-19 suspect. If the person is a COVID-19 suspect, the alert is generated, and the suspect is sent for a medical test such as RAT, or RT/PCR. The COVID-19-positive patients are monitored by keeping track of different parameters for any medical emergency. Suppose an oxygen level of a COVID-19 person dips below a predefined value or threshold value, an emergency alert is sent to caregivers, doctors, and other medical facilities to provide services in real-time. The proposed system can both eliminate the spread of infection by early detecting the suspect and minimize the fatality rate by providing the services immediately in case of emergency.

Local Data Analysis: The data that is received from the below subsystems are analyzed in this subsystem at the edge of the system. The machine learning models that are trained in the cloud are deployed on the edge of the network and are running to check the severeness of the user. The Bayesian methods work well for a small amount of data, so different Bayesian machine-learning models are used in our proposed system. Suppose, if a user has a fever, i.e., above a threshold value, the suspect is sent

for medical examinations to get better insights. In case of any emergency like low oxygen, the alert can be generated in real-time to provide the needed medical services. The edge subsystem of the proposed system can be used for different services such as data aggregation, data security, data storage, etc. In this sub-module, the data is stored in local databases, i.e., data is local to the system and synchronized with the cloud system for further analytics.

Caregivers and Physicians: The healthcare system is incomplete without caregivers; it is a generalized term for medical facilities and non-medical facilities. The medical staff like doctors, nurses, and ambulance services come under medical facilities, and caretakers (family members) come under non-medical facilities. The caregivers monitor the patient continuously, indirectly (IoT), or directly (physicians & caretakers). The IoT technology is a potential technology that can be mingled with traditional healthcare systems to tackle the COVID-19 outbreak.

Medical Lab: The suspect data sent from the edge gateway is processed here in this module. This module is made up of medical lab technologists like biochemistry, microbiology, etc. If the suspect is COVID-19 positive, the data is sent back to the edge gateway, where the alert is generated, and the local database is updated. If a person is negative, he may be put under surveillance or be home quarantined.

Emergency Alert: Human life is the most valuable on earth, so it is a fundamental right to provide high-end services to the public. The healthcare system is mainly related to human lives, so the healthcare system should be developed enough to tackle any pandemic. Our proposed system is a combination of different technologies like IoT, Machine Learning, and computing technology (Edge/Cloud). The emergency alert is a subsystem of our proposed system by which alerts are generated in case of a medical emergency. This module is generating two types of alerts one is first when a person with high temperature, high or low pulse, etc. came under the scanner and when an under surveillance COVID-19 patient's oxygen level dips below the threshold value. An emergency alert is generated by the proposed system in real-time to tackle the situation.

4.2.3 Cloud Subsystem

The central processing system where everything is processed on demand is cloud computing. Cloud computing is a buzzword of the twenty-first century, where processing is done in a centralized manner. Cloud computing provides different services on demand like infrastructure as a service (IaaS), platform as a service (PaaS), software as a service (SaaS), and other services. The cloud system is a powerful system where high-end computations and analytics are done using the available services. Our proposed system is connected through the internet, i.e., cellular networks with the cloud system and other components. The data analyzed by the subsystem below it is sent here for other analysis to better understand the nature of the disease. The cloud system is used for storing the data and training the machine learning models on the updated data received from the edge system. The data from the cloud is accessed by different participating agencies which are part of the system.

Advanced-Data Analytics Center: The advanced-data analytics center is a sub-module of our proposed system, in which high-end machine learning models are trained repeatedly to modify the model. The parameters of a person taken from the module below are analyzed for other possible comorbidities such as hypertension, kidney disease, lung disease, etc. If so, information is sent to the edge gateway to update the patient's local database and emergency alerts are generated if needed. The machine learning models currently running on the edge of the network are synchronized with those running on the advanced-data analytics center. The accuracy of both subsystems increases with time, and our proposed system will be more accurate with time and data. The medical database of the cloud subsystem is also updated with the results generated during predictions.

Disease Diagnosis System: Our healthcare system is not technology-based; our proposed system is a step forward toward it. Our eyes will see intelligent healthcare systems where bots will soon replace doctors with high accuracy in diagnosing the disease. All medical procedures will be done through an automatic medical machine where there will be no chance of errors. Till now as of our knowledge, no medical expert system is available, so still, medical experts are needed. Our proposed system has different modules; the disease diagnosis system is one among them. In this module, well-known experts in the field are checking the parameters of patients and prescribe medicines. The health physicians are checking the other symptoms of the patients by collecting the data from the advanced medical lab module. The data is analysed for the possible presence of other diseases such as comorbidities like hypertension, kidney, etc., or levels of the infection. Similarly, in the case of COVID-19, the parameters are checked based on severity like mild, moderate, severe, and critical based on a patient's clinical parameters. The symptoms of individuals are analyzed to understand the virus's nature better so that an effective vaccine can be developed.

Advanced Medical Lab: The high-level medical procedures of the patients are checked in this module, like the other microbiological and biochemistry procedures. The edge module is not so powerful to do whole procedures on a local level so the offloaded requests from the edge module via the cloud are handled by this module.

4.3 Data Flow of Proposed Framework

The most important part of the proposed system is showing the flow of data in a pictorial representation to better understand the proposed framework. Fig. 4, gives a graphical representation of the data flow of the proposed system. In our proposed system, the data goes from lower layers to upper layers, and from upper layers, information flows to lower layers. The data flow from the data acquisition module to the local data analysis module and then to the high-end data analytics module.

The data flow of the proposed system is elaborated following steps:

1. The sensors, wearables, and other applications are used as data generators. These sensors collect the symptoms such as fever, pulse, BP, oxygen saturation, runny nose, dyspnea, myalgia, and other parameters through smartphones like travel history, etc., in real-time. The sensors, wearables, and smartphones collectively make a network of body-worn sensors known as a body area network (BAN). The collected symptoms from the sensors are then analyzed.
2. The data, i.e., symptoms received from step 1 is analyzed with the help of trained machine learning models running on the edge of the network. Based on the results of algorithms the detection and prediction are made on whether the suspect is COVID-19 or not. The main goal of this sub-module is to send time-sensitive information to caregivers and responders in real-time. The incoming data from step 1, is locally stored on edge of the network. The analyzed data is then sent to the next module for further analysis and machine learning models are updated accordingly.
3. The data received from step 2, is further analyzed by high-end machine learning models to get better insights from the received data. The machine learning models are continuously updated with data and are synchronized with models deployed on the network edge.
4. The information is stored in the cloud and is shared or accessed by participating parties like healthcare professionals and healthcare centers. The patients are then sent for further clinical tests to check further comorbidities like hypertension, lung disease, kidney disease, etc. If so, the database in both places is updated accordingly. The patients are sent for medical treatment

and are monitored and put under medical observation. In this way, the spread of coronavirus infection may be eliminated and the nature of the disease may be better understood.

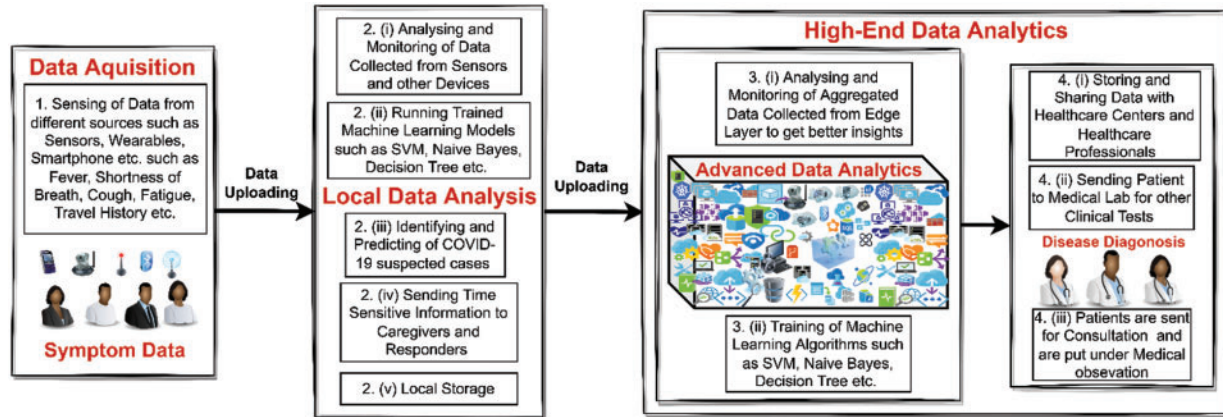


Figure 4: Pictorial representation of data flow of proposed system

5 Experimental Setup

5.1 Methodology Adopted

The section describes the overall methodology adopted and the working of a proposed framework in detail. According to the proposed architecture and framework in the above paragraphs, the framework is divided into three different networks. The three different networks are Medical Sensor Network (MSN), Medical Edge Network (MEN), and Medical Cloud Network (MCN). MSN is responsible for data collection (symptoms) by different deployed sensors. MEN receive’s the data from MSN for local analysis and generates emergency alerts. MCN is receiving the symptoms from MEN for further analytics, training the ML models, and further analyzing the symptoms. The pictorial presentation of the overall setup is depicted in Fig. 5.

The working of each module concerning layered architecture is elaborated in the following steps:

1. Medical Sensor Network (MSN): MSN is a network of medical sensors, such as a temperature sensor, pulse sensor, oxygen sensor, breath sensor, inertial sensor, airflow sensor, audio sensor, motion sensor, ECG sensor, other devices & wearables. The sensors collect the data (symptoms) of a suspect and send the values to the edge gateway. In the case of COVID-19, the temperature sensor sends the temperature of the suspect and the oxygen sensor sends the oxygen saturation of the suspect. The edge gateway is providing services locally like computation, analysis, aggregation, and storage. The parameters of the sensor reading are checked in the edge layer with the available values. The sensors are low-power and resource-constrained devices, so offloading is necessary to increase the life the sensors. The communication between the sensors and the edge gateway is through 802.15 standards.

2. Medical Edge Network (MEN): MEN is a middle layer between MSN & MCN and acts as a gateway between these two layers. The edge gateway translates the protocols of MEN when communicating to MCN and vice versa. The edge layer provides various services such as computation, local analysis, storage, and aggregation in real-time. The edge layer is close enough, i.e., in the proximity of the data source, to provide time-sensitive information to caregivers and responders. The edge layer is comprised of many high-level devices where computation and analysis can be done. For example, if the temperature of the COVID-19 suspect is above the predefined value an alert will be generated, and other parameters of the suspect are collected. If a suspect is identified as COVID-19 positive, it is then

monitored continuously. Suppose the oxygen level of a person dips below a certain level an emergency alert will be generated and the caregivers and responders will be notified accordingly. In our proposed framework the machine models are trained on the cloud and are running on edge to eliminate the cons of the cloud. Healthcare is a time-critical application as it does not bear high response, high latency, and jitter, so the edge provides the solution to overcome these problems. The edge gateway uses cellular technology to communicate with the cloud network. Two types of alerts are generated in this layer one is when a suspect has a high temperature, low or high pulse, low oxygen saturation, and difficulty breathing. In this case, an early alert is generated to check the suspect for the presence of COVID-19 infection. The procedure and working for an early sign for the possible suspect to generate an early alert is described in Algorithm 1. The second is when the suspect is confirmed as COVID-19 positive and is on continuous monitoring support for vital parameters. If the patient has a dip in oxygen saturation and shortness of breath, an emergency alert is generated, and the signal is sent to all participating parties for emergency care. The patient needs to be hospitalized as soon as possible to provide the necessary support. The working procedure for a medical emergency needed for a patient to generate an emergency alert is depicted in Algorithm 2.

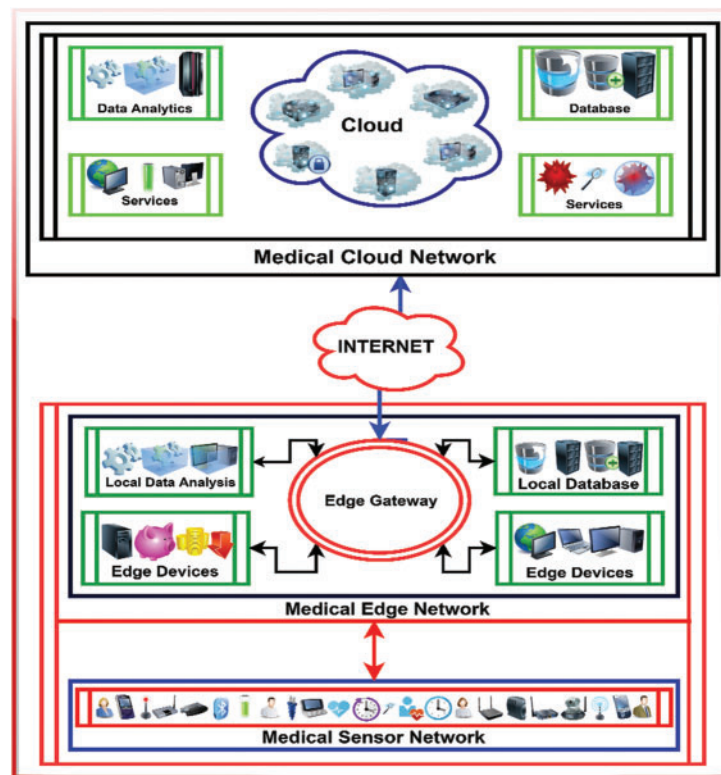


Figure 5: Logical view and working of framework

Algorithm 1: To Predict the COVID-19 Suspect at Edge Layer

INPUT: *Symptoms of Suspect;*

OUTPUT: *Whether a Suspect is COVID-19 Positive or Negative;*

Step 1: *Select the Primary Symptoms of Context, i.e., Fever, Pulse, O₂ Saturation, and Breath;*

(Continued)

Algorithm 1: (continued)

3 Step 2: *Compare each Symptom value with the stored value:*
 a. *If the value is not in the Normal Range;*
 Then the person may be a COVID-19 suspect: and GOTO Step 3;
 b. *Else Person may not be a suspect GOTO Step 5;*
 Step 3: *Generate an Early Alert and send it to responders and caregivers:*
 a. *Collect all other parameters of the suspect;*
 b. *Send the collected parameters for analysis;*
 Step 4: *If the Suspect is COVID Positive;*
 a. *Update the database;*
 b. *Send the Information to Caregivers and Responders;*
 Step 5: **Exit;**

Algorithm 2: To Predict the Severity of COVID-19 Patients at Edge Layer

INPUT: *Vital Signs of Patient;***OUTPUT:** *Whether a COVID-19 Positive needs Emergency Support or Not;*Step 1: *Select the vital Symptoms of Context, i.e., Fever, Oxygen Saturation, and Breath;*Step 2: *Compare each Symptom value with the stored value:*a. *If the value is not in the Normal Range;*

Then the COVID-19 patient needs emergency support: and GOTO Step 3;

b. *Else Patient is Stable GOTO Step 4;*Step 3: *Generate an Emergency Alert and send it to responders and caregivers;*a. *Collect all other parameters of the suspect;*b. *Monitor the Patient Continuously;*Step 4: **Exit;**

3. Medical Cloud Network (MCN): MCN is the last layer of our proposed architecture and is comprised of high-end medical servers. This layer is responsible for providing high-level services computation, analytics, storage, and others. The communication between the edge and cloud layer goes through edge gateways using cellular networks like 3G, 4G, 5G, and LTE. The requests that MSN generates are handled by MEN and those which are not available on edge are sent to the MCN. MCN responds to the requests of MEN and a response is sent to MSN. The high-level machine learning algorithms are running on the data received from MEN to get better insights and better understand the nature of the disease. The ML models are trained on MCN and deployed to MEN. The edge layer is not powerful enough, so machine learning models can be trained on MEN. MCN provides all other high-level services and stores the data for future use. Cloud computing is a centralized computing paradigm in which computation is done in the cloud in a centralized fashion. There are many disadvantages of this fashion like response time, latency, and jitter, which are not tolerable by healthcare applications. Healthcare is a time-critical application in which high responses, latency, and jitter are not acceptable. To overcome this and maintain healthcare's criticalness, our proposed system has used an edge layer to eliminate these problems.

The main focus of our proposed system is on edge of the network, i.e., the Medical Edge Network (MEN). The main logic is implemented in this module of the proposed system. The data that is acquired through the perception layer from different sources such as sensors, wearables, and medical IoT devices is handled by the edge layer. The processing of data and ML models are running to generate

results. The time-sensitive information is sent to caregivers and responders as emergency alerts. The detection and prediction of the potential suspect are made through ML models in the early stages. The other services are also provided by the edge layer such as data aggregation, data compression, local storage, etc. for fast communication. The data is sent to the cloud for further high-end analytics and for getting other high-level services. The proposed model will minimize the communication latency, response time and delays occurred in cloud-based systems. The data that is sent to the cloud from the edge is aggregated and compressed to minimize the network traffic and bandwidth requirement. Other services can also be provided such as data security to ensure the privacy of the individual's healthcare data.

5.2 Machine Learning-Based Detection & Prediction Model

This work uses the dataset preprocessed in [Subsections 3.4](#) and [3.5](#) of [Section 3](#) to build a prediction model for our proposed system. Predictive analytics are useful in almost all fields, especially in modern smart healthcare systems where the data is generated from multiple sensors. In traditional systems, the models were built on historical data and by expert suggestions, associations between variables are established. These data models are static and of limited use for fast-changing and irregular data. On the other hand, IoT-based systems are producing a lot of data in a dynamic state continuously and in real-time, so traditional data models would not be a good choice to follow. The machine learning models can learn explicitly without being programmed. Since our proposed model is predictive and the dataset collected is used to test the model for the prediction of COVID-19 suspects. The gathered dataset is symptomatic and is categorical data with the class label as positive (1) and negative (0). The work of this proposed model is to predict the COVID-19 suspects in the initial stage to minimize the transmission of coronavirus infection. Our dataset is labeled and the problem is based on the classification of COVID-19 as positive or negative, so our primary focus is on supervised machine learning techniques. During model building the dataset is split into two parts; one training part and another testing part. The training part is mostly above half of the dataset and the testing part is below half. The training part is used to train the build on different instances and the test part is used to examine the performance of the proposed model. Based on the working of the ML algorithms there are various categories of algorithms. In our study different ML techniques are used to build a perfect model.

Support Vector Machine (SVM) is a non-probabilistic binary classifier best separates the two classes with the maximum margin by drawing the hyperplane. The model is then used to predict the class label for new unseen input data to which class it belongs. In our case, the proposed model is used to predict the class label based on the symptoms collected in either positive or negative classes. SVM works by creating a decision boundary that can best segregate m-dimensional space into different classes. The hyperparameter settings adopted during the training and testing phase of the model for some of the important parameters like, $C = 1.0$, Kernel = 'rbf', Cache = 200 and Random_State = 'None'.

Logistic Regression is the most significant supervised machine learning technique based on logistic function. Logistic regression is the upper version of linear regression and is used for regression problems whereas logistic regression is used for classification problems. Logistic regression predicts the output as a categorical value (Yes or No) of 0 or 1 but in probabilistic form. Logistic regression works by fitting the S-shaped curve based on a function that can predict two maximum values. The parameter settings adopted for the solver = 'lbfgs', max_iter = 100, and verbose = 0 during the training and testing of the model.

Naïve Bayes is a probabilistic supervised machine learning technique based on the Bayes theorem used for classification problems. Naïve Bayes is a fast and easy machine-learning technique that can predict the class label of the data. Naïve Bayes is based on assumption that the attributes have no relation and are independent of each other so it is unable to draw the relationships among the attributes. There are various types of naïve bayes models like Gaussian, Multinomial and Bernoulli naïve bayes. The values adopted for the parameters like $\alpha = 1.0$, `force_alpha = 'warn'`, `fit_prior = True`, and `class_prior = None`. The hyperparameter setting is adopted during the training and testing of the model according to the problem to get the best results.

Neural Network belongs to the family of artificial neural networks (ANN) modeled to mimic the human brain. The neural networks are used for classifications as well as clustering problems but it is mainly used for supervised learning. Neural networks work by learning the relationships among the attributes by using a series of inbuilt connections, i.e., neurons are connected with others to classify things. Neural networks are mainly made up of three layers: input layer, hidden layer, and output layer. There are various types of neural networks like modular NN, Feedforward NN, Recurrent NN, Convolutional NN, etc. The attribute setting adopted during the training and testing of the proposed model for `activation_function = 'relu'`, `learning_rate = 'constant'` and `random_state = 10`. Based on these settings, we have got the best results for our problem of COVID-19 disease classification.

A decision Tree is a supervised machine learning technique that is used for both classifications as well as for regression problems. Internally it organizes the attributes in the form of a tree-based structure, where internal nodes are representing the attributes, links are decisions made and leaf nodes represent class labels. For selecting the best root various selection measures are followed, Information Gain and Gini Index are most commonly used. The hyperparameter values for the decision tree adopted during the training and testing phase are as, `criterion = 'gini'`, `max_depth = 4` and `splitter = 'best'`.

Random Forest is a supervised machine learning technique used for both classification and regression problems. Random forest internally works by generating many decision trees and takes an average of all to predict the accuracy. The random forest is based on the concept of ensemble learning, in which multiple classifiers are used to solve complex problems. Decision trees suffer from a problem of overfitting during the training phase that is eliminated in random forests. The attribute settings adopted during the training and testing phase of the model for some of the important parameters are, `n_estimators = 100`, `criterion = 'gini'`, and `max_features = 'sqrt'`. Hyperparameter tuning is an important step in building the optimal model for any problem. The hyperparameters are to be tailored according to the need of the problem. In our case, to classify the suspect either as COVID-19 Positive or COVID-19 Negative, we have tried different settings and lastly saved the best one.

The work is done using Python 3.9 and in Anaconda Environment. The novel primary dataset collected from SKIMS, Srinagar, Jammu & Kashmir, India, is used to build the models using different algorithms. The statistical analysis of the data is performed in python to get better knowledge about the nature of the dataset. [Table 3](#), highlights the main statistical findings of the prepared dataset. The dataset contains 7015 rows and 16 columns with binary values either 1 or 0. The statistical analysis of the dataset shows that the dataset contains 3112 as negative and 3903 instances as positive. Apart from this, the average mean is 0.399567 and the average standard deviation is 0.258997.

During our study, the dataset has been divided into ratio of 70:30 in which 70% of the data is used to train the model and 30% is used to test the model. The novel primary dataset of COVID-19 symptoms contains a total of 7015×16 instances. The training set contains the 4910 and the test set contains the 2105 instances of the dataset based 70:30 ratio. As our problem is a classification problem,

the suspect is either COVID-19 positive or negative so the binary classification method is used, in which 0 represents negative and 1 represents positive. Fig. 6 displays the train, test, and overall accuracy of the proposed models.

Table 3: Statistical analysis of dataset

Total number of instances	Total number of rows	Negative instances
7015	16	3112
Positive instances	Average mean	Average standard deviation
3903	0.399567	0.258997

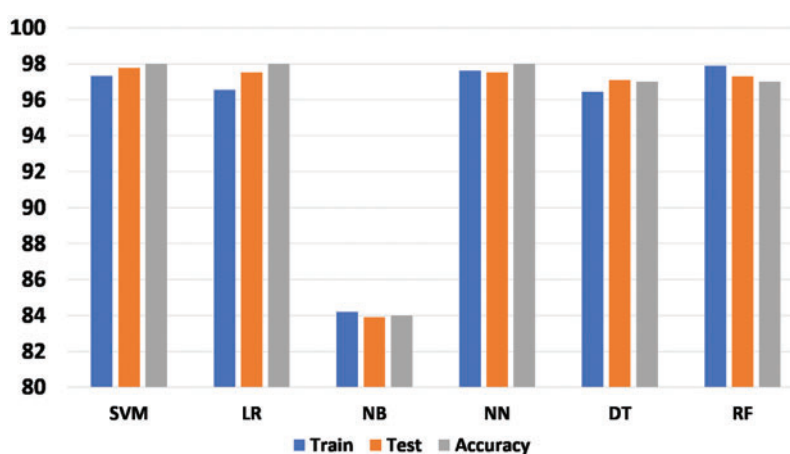


Figure 6: Accuracy of proposed machine learning models

The performance of the developed models is generally measured in terms of the accuracy achieved by the model. During our research work, six machine learning models were proposed namely, support vector machine, logistic regression, naïve bayes, neural network, decision tree, and random forest. The models were compared using three different accuracies such as train accuracy, test accuracy, and overall accuracy. The training accuracy is defined as the accuracy achieved by the model during the training phase to perfectly classify the suspects into either a positive class or a negative class. Test accuracy on the other hand is the ability of the developed to perfectly classify the suspects into either positive or negative classes. Accuracy is generally the same as the test accuracy of the proposed model to check the performance of the model. The support vector machine, logistic regression, and neural network have achieved 98% accuracy, whereas the decision tree and random forest have achieved 97% accuracy. The Naïve bayes model did not perform well on this data while it achieved only 84% accuracy. During the training phase, the support vector machine achieved 97.31, logistic regression 96.56%, naïve bayes 84.20%, neural network 97.62%, decision tree 96.44%, and random forest achieved 97.88%. The above discussion makes it clear that most of the proposed machine learning models performed well in terms of accuracy except naïve bayes which performed least on the novel dataset.

6 Performance and Model Evaluation

The models that are built in the above section are validated through various performance measures. Model evaluation is an important part of the development process of a model. The model evaluation process helps us to select the best-fit model that describes our data. The future predictions that would make our model are evaluated in the process of model evaluation. As our problem is a binary classification of the COVID-19 suspects, there are four types of possible outcomes. These four possible outcomes make a confusion matrix of the model represented in Fig. 7.

True Label	Positive	True Positives (TP)	False Positives (FP)
	Negative	False Negatives (FN)	True Negatives (TN)
		Positive	Negative
		Predicted Label	

Figure 7: Diagrammatic representation of confusion matrix

6.1 Various Performance Measures

A confusion matrix is a 2×2 (two rows, two columns) matrix representation, that provides the figure of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). The confusion matrix provides a graphical representation of the accuracy of the model, and how much our model is accurate.

- True positive (TP) is the number of instances that are predicted positively by the model and are also positive in actual. Suppose a person is COVID-19 positive, our model has also predicted the suspect as COVID-19 positive.
- True negative (TN) is the number of instances that are negative in actuality, and our model has also predicted them as negative. In our case of COVID-19, if a suspect is negative and the model has put the suspect in the negative class.
- False positive (FP) is the number of instances that are negative in actual but our model has predicted them as positive. Let us suppose a COVID-19 case if a suspect is negative but our model predicted it as positive.
- False negative (FN) is the number of instances that are labeled as negative by our model but in actuality, they are positive. Suppose a person is COVID-19 positive, our model has predicted it as negative.

Furthermore, the performance evaluation of the above machine learning models is evaluated through the different performance measures such as Accuracy, Precision, Recall (Sensitivity), F-Score, Specificity, Area Under Curve Score (AUC-Score), and Root Mean Square Error (RMSE). Based on the above short definitions the performance measures are computed as follows:

- The accuracy of a model is a basic and important performance-measuring metric that is used to check the performance of a machine learning model. Mathematically, it is computed as the total number of correctly predicted instances divided by the total number of instances. The mathematical representation of accuracy is denoted as below:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

- Precision is the measure of the exactness of the machine learning model. Mathematically it is represented in the equation below:

$$Precision = \frac{TP}{TP + FP}$$

- The recall is referred to as the true positive rate (TPR) of a machine learning model and is also known as sensitivity. A mathematical representation of the recall is shown below:

$$Recall/Sensitivity = \frac{TP}{TP + FN}$$

- F-Score is the harmonic mean of the recall and the precision and mathematical representation of the equation being given below:

$$F-Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

- Specificity is referred to as the true negative rate (TNR) of a machine learning model. Mathematically it is represented as follows:

$$Specificity = \frac{TN}{TN + FP}$$

- RMSE is a performance measure and is computed as the difference between the predicted instances of the model and the total number of instances. A mathematical representation of the equation is given below:

$$RMSE = \sqrt{\frac{FP + FN}{TP + TN + FP + FN}}$$

- AUC-Score is the region under the ROC Curve and the value is always between 0 and 1. The proposed model is good if its value is close to 1. Mathematically it is represented as follows:

$$AUC = \frac{1}{2} * \frac{TP}{TP + FN} + \frac{TN}{TN + FP}$$

6.2 Model Validation Techniques Adopted

Model Validation is a technique of evaluating the performance of the model by authenticating that the model is performing for what it was intended to. Suppose in the case of a COVID-19 suspect, it is predicting the right person as a COVID-19 suspect based on the symptoms. The logic behind the model evaluation is that the model is trained with one part (mostly large) of the data and one part (small part) is used for testing the performance of the model on new instances. There are various techniques to evaluate the performance of the model, the hold-out method and k-cross-validation are the most frequently used. The hold-out cross-validation is the simplest method in which the dataset is

divided into two sets usually in the ratio of 70:30 depending upon the size of the data. 70 percent of the data is used for training the model and 30 percent is used to test the performance of the model. But this method may result in overfitting, in which the model performs well during the training phase but drops its performance in the testing phase. To overcome this problem another technique is k-cross-validation. In k-cross-validation the dataset is divided into k equal parts, k-1 parts are used for training the model and the kth part is used to test the model’s performance. This procedure is followed k times and shuffling of data each time. Generally, the $k = 5$ or $k = 10$, as these values do not suffer from high bias or low variance. In our case, the 10-cross validation has been used to check the performance of the model proposed. The dataset is divided into k-folds ($k = 10$), k-1 folds are used for training, the kth fold is used for testing and this process is repeated k times. The procedure of k-fold-validation employed in our research is represented in Fig. 8.

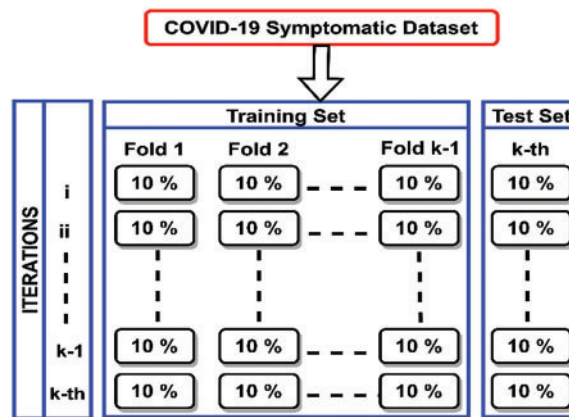


Figure 8: Diagrammatic representation of adopted k-fold cross-validation

7 Results and Discussions

The performance of the proposed machine learning models is elaborated in the below sections. As our problem is a binary classification of disease, i.e., COVID-19 positive or negative. The most basic and important performance measures by which our proposed models are verified are Accuracy, Precision, Recall, F-Score, Root mean Square Error and Area Under Curve Score. Apart from these performance measures Confusion Matrices and Receiver Operating Characteristic Curves are also shown diagrammatically. Lastly, the k-cross-validation of proposed machine models is carried out to check the model’s robustness.

7.1 Confusion Matrices of Various Proposed Models

The confusion matrices of the machine learning models applied to the novel primary symptomatic dataset are represented in Fig. 9. The diagonal elements of the matrices represent truly classified scores whereas non-diagonal elements represent misclassified scores of the proposed models.

The confusion matrices of the proposed machine learning models above are summarized in Table 4. The table displays the outcome of the matrices of the proposed models in terms of true positives, true negatives, false positives, and false negatives. The confusion matrix represents the instances into four quadrants in the case of binary classification. As our problem is to classify the suspects into either COVID-19 positive or negative based on the symptoms. During the experimentation, the hold-out method has been used in which the dataset has been divided into unequal parts 70:30. The 70% of the data is used for training the proposed machine learning model and 30% for testing the performance

of the model in terms of various performance measures. In our study, the primary novel dataset based on the symptoms of the patients collected from the SKIMS, Srinagar, J&K, India has been used. The dataset contains 7015 rows and 16 columns with each column having a binary number either 1 or 0 as a value. The class label is also in binary value, as our problem is to classify the suspects as either COVID-19 positive or negative. 1 represents the presence and 0 represents the absence of the symptom similarly, for class label 1 represents the COVID-19 positive, and 0 represents the COVID-19 negative. After analyzing the confusion matrices of proposed machine learning models almost all have performed well except the naïve bayes.

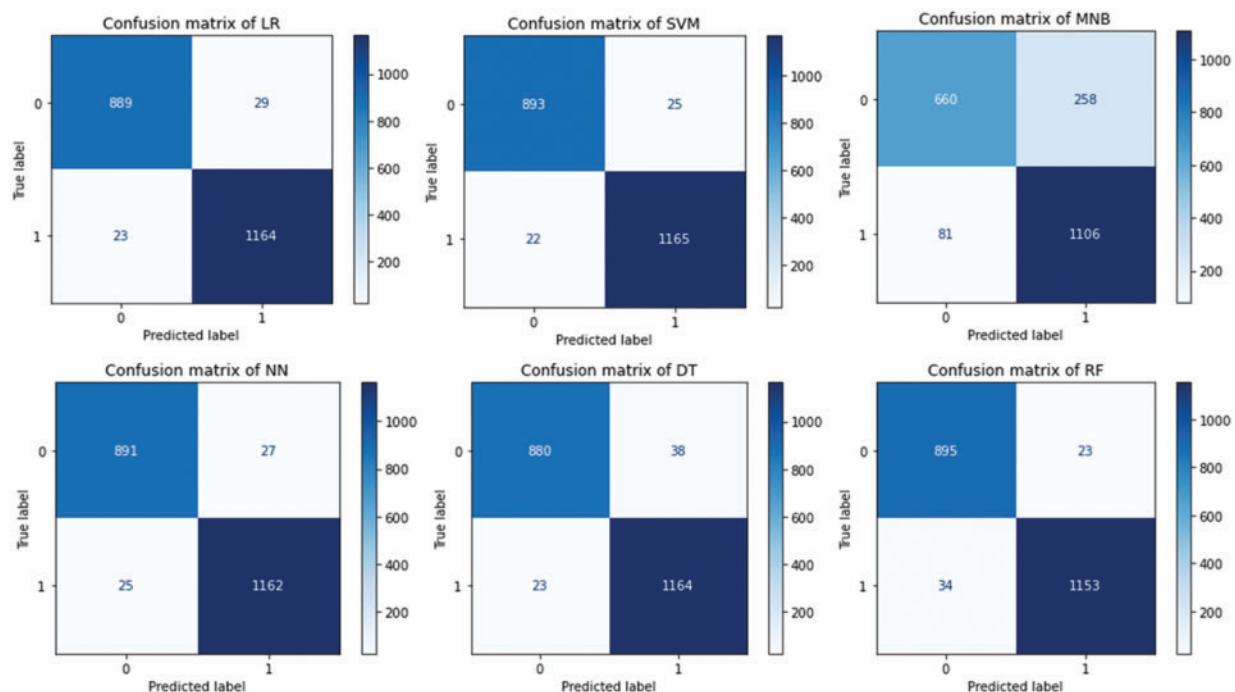


Figure 9: Confusion matrices proposed machine learning models

Table 4: Summary of figures of confusion matrices

Machine learning model	True positive (TP)	True negative (TN)	False positive (FP)	False negative (FN)	Support
Support vector machine	1165	893	22	25	2105
Logistic regression	1164	889	23	29	2105
Naïve bayes	1106	660	81	258	2105
Neural network	1162	891	25	27	2105
Decision tree	1164	880	23	38	2105
Random forest	1153	895	34	23	2105

The table above displays each proposed machine learning model concerning its four quadrants of matrix and the support. From the above table, it is clear that the dataset used for experimentation purposes is balanced, so there are less chances of biased prediction. As in our proposed machine

learning models, 70% of the data is used for training and 30% for testing the performance of the model. The dataset contains 7015 instances in total which implies 4910 instances of the dataset are being used for training the machine learning model and 2105 instances are used for testing the performance of the model. The above table represents the correctly classified instances and misclassified instances into four columns. Our problem is a binary classification of suspects either COVID-19 positive or negative, so the entries of the table will be true positive & true negative (correctly classified) and false positive & false negative (misclassified). The model has performed well if it has less (~ 0) number of false positives and false negatives, i.e., a smaller number of misclassified instances. In our case, almost all proposed machine learning models have performed well except the naïve bayes which have made more misclassifications. In the case of disease, both the values, i.e., false positives and false negatives should be low (~ 0) to minimize the impact of the disease. But in the case of COVID-19, the model should have a low i.e., ~ 0 number of false negatives to minimize the fatality and spread of the infection. Suppose our model predicts COVID-19-positive persons as negative, it will infect many others and degrades the performance of the proposed model. The above concept can be further understood with the concept of errors made by the model. There are two types of errors first is a Type-I error and the second is a Type-II error. Type-I error means when the model misclassifies a person and wrongly puts him in a positive class. Suppose in the case of COVID-19, a person is classified as positive but in actuality, he is negative, this will not impact so much on our proposed model much. On the other hand, if the person is misclassified as negative, but in actuality he is positive, it will make a bad impact on the proposed model. This is known as Type-II error, Type-II error is more dangerous in our case of COVID-19.

7.2 Various Performance Measures Used for Comparing the Proposed Models

The proposed machine learning models are evaluated using different performance measures such as Accuracy, Precision, Recall, F-Score, RMSE, and AUC area. Table 5 displays the results of each of the applied machine-learning techniques with respect to essential performance measures.

Table 5: Summary of results of performance measures

Machine learning model	Accuracy	Precision	Recall	F-Score	AUC-Score	RMSE
Support vector machine	97.767	0.98	0.98	0.98	0.98	14.94
Logistic regression	97.530	0.98	0.97	0.97	0.98	15.72
Naïve bayes	83.895	0.85	0.83	0.83	0.83	40.13
Neural network	97.530	0.97	0.97	0.97	0.98	15.72
Decision tree	97.102	0.97	0.97	0.97	0.97	17.02
Random forest	97.292	0.97	0.97	0.97	0.97	16.02

The proposed models are checked or tested using different measures that are listed in the table above. Our problem is related to the healthcare domain to detect and predict the COVID-19 suspect in the early stage to minimize the fatality and spread of the coronavirus infection. Accuracy is a performance measure to check the performance of the proposed model, from the above table almost all of the proposed models have achieved above 97% accuracy but naïve bayes performed least with 83.895% accuracy.

Root Mean Square Error (RMSE) is the standard deviation of errors that are made by the proposed machine learning models in making the predictions. The lower the values of RMSE the better the model is, SVM achieved the least value of 14.94, logistic regression and neural network got 15.72, random forest 16.02, decision tree 17.02 and the highest score of 40.13 by naïve bayes. In terms of precision, recall, f-score, and AUC-score almost all the proposed models have achieved a good score

except naïve bayes which performed the least. F-score is the best measure for any application area, it gives overall information about the proposed model in terms of classification. F-score is a combination of precision and recalls as a single metric to check the performance of the model. When the cost is associated with the false negatives (FN) the recall is the best measure to check the performance of the model. Otherwise, if the cost of the problem is associated with false positives (FP), precision is the metric to check the performance of the model. In our study of the detection and prediction of COVID-19 suspects, the cost of the problem is associated with the false negatives (FN) recall is the best measure to check the performance of the model.

7.3 ROC Curves of Various Proposed Models

Receiver Operating Characteristic Curve (ROC) is another important performance measure by which the efficiency of the machine learning model is checked. The ROC curve is shown in the graph with the true positive rate on the x-axis and the false positive rate on the y-axis on all thresholds. The classifier is better if the curve tends to be close to the upper left corner of the graph, otherwise degrades the efficiency of the model. Fig. 10 shows the ROC curves of the proposed machine-learning models.

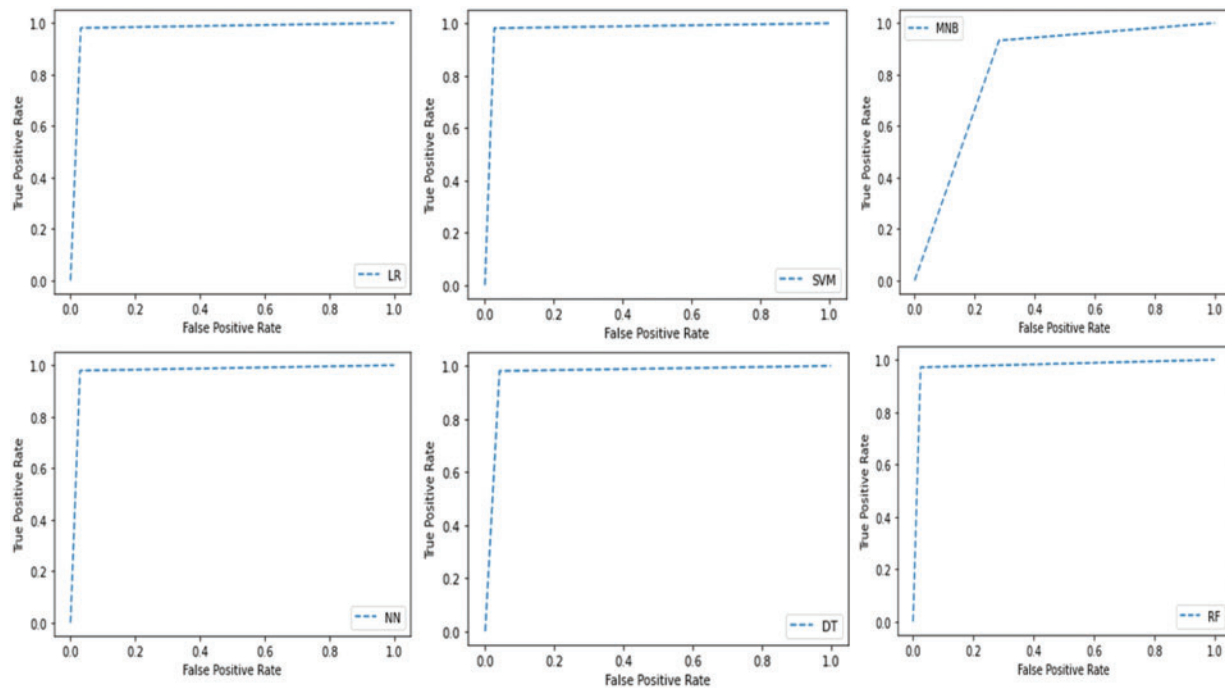


Figure 10: ROC curves of proposed ML models

7.4 K-Cross Validation of Various Proposed Models

The most important thing to check the reliability and consistency of the proposed model is by using cross-validation. In cross-validation, the proposed models are checked for the problems of overfitting and underfitting that are occurred in hold-out validation. Overfitting is a problem of training the model with a lot of data and it tends to learn new things from noise. Underfitting on the other hand occurs when there are fewer data instances to train the model or when the model does not fit the data. Both conditions are degrading the performance of the model in terms of accuracy and other performance measures. In our research, the 10-fold cross-validation ($k = 10$) has been used

to check the robustness of the proposed machine learning models. The accuracies of all 10 iterations have been recorded with the mean accuracy of each iteration listed in [Table 6](#).

Table 6: K-Fold (10) cross-validation results

ML model	Folds										Mean accuracy
	Accuracy										
	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	
Support vector machine	96.01	97.15	98.01	98.15	96.58	96.58	98.15	96.29	97.57	98.00	97.25
Logistic regression	95.73	96.58	97.58	97.44	96.44	95.86	97.43	96.01	97.72	98.00	96.89
Naïve bayes	85.91	82.48	85.04	83.33	83.62	83.02	83.31	84.31	84.74	85.73	84.15
Neural network	96.01	97.01	97.44	98.01	96.44	96.15	97.18	96.15	97.29	98.15	97.04
Decision tree	95.58	96.72	97.15	98.14	96.87	96.01	97.43	95.72	96.72	97.15	96.75
Random forest	95.73	96.58	97.15	98.15	97.01	96.58	97.72	95.86	96.58	97.86	96.92

The model validation is an important step to check the reliability and consistency of the model and whether the model is free from overfitting and underfitting problems. During our work, the models are validated using k-cross-validation. The accuracy graph is shown diagrammatically of all 10 iterations in [Fig. 11](#). From [Table 5](#) and [Fig. 11](#), it is clearly visible that each proposed machine learning model has shown consistency in all ten accuracies. The accuracies are varying at most three decimal numbers for almost each proposed machine learning model. The lowest accuracy of SVM is 96.01%, the highest is 98.15% and the mean accuracy is 97.25%. Similar differences are in the accuracies of the rest of the proposed machine learning models. [Fig. 11](#) displays the 10-fold cross-validation accuracies of various proposed machine learning models.

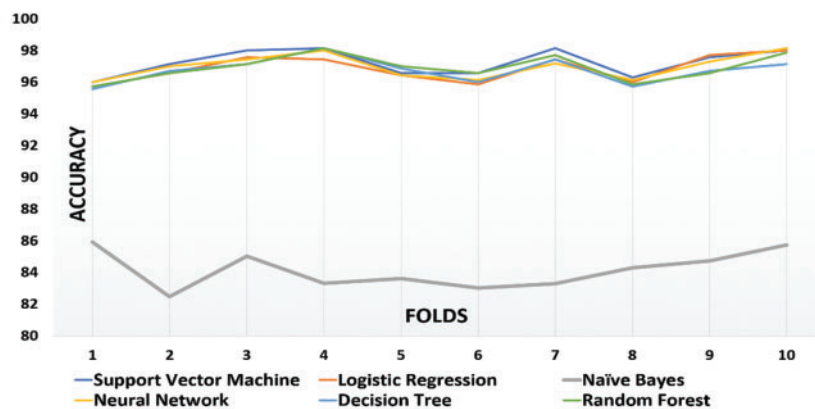


Figure 11: Graphical representation of the 10-fold cross validation accuracies

The graph above is drawn for the accuracies of proposed machine learning techniques against the 10-fold. From the figure above it is clear that almost all the proposed techniques have achieved accuracy above 96%, and the naïve bayes have achieved only around 84%. SVM and neural networks have achieved around 97% mean accuracy, whereas, decision trees and random forests have achieved around 96.5%. From the above discussion, it is clear that SVM and neural networks have performed well on our novel dataset.

7.5 Comparative Analysis

7.5.1 Comparative Study of Different Proposed ML Models Based on Accuracies

The [Table 7](#) displays the accuracies of various proposed machine learning models with respect to two validation methods. The two validation techniques used in our work are hold-out validation and k-cross validation, for hold-out validation the ratio of 70:30 is used and for k-cross validation, $k = 10$ is used. The hold-out validation is dependent on the train-test split that cannot be changed for all. On the other hand, during the cross-validation method, the model is trained on each part and tested on each part. The table compares the accuracies achieved each in hold-out validation and k-cross validation. During our study we have used $k = 10$, which implies the dataset is divided into ten-folds (10), from that nine-folds are used for training and one-fold for testing. The process is repeated ten times to give each fold a chance for training as well as for testing. This process requires a lot of time and resources to build a model, but it will indicate how better predictions can be made on unseen instances of the data. From the above discussion and [Table 6](#), it is clear that our models have performed well in both the hold-out and cross-validation methods. The differences in hold-out accuracies and 10-fold accuracies are minor for each proposed model so it is clear that our proposed models are not biased.

Table 7: Comparison of hold-out and K-fold accuracies

Machine learning model	Hold-out accuracy	10-fold mean accuracy
Support vector machine	97.767	97.249
Logistic regression	97.530	96.878
Naïve bayes	83.895	84.148
Neural network	97.530	97.035
Decision tree	97.102	96.750
Random forest	97.292	96.921

The two accuracies such as hold-out accuracy and 10-fold mean accuracy of each proposed model given above in the table being represented in the bar chart format. The SVM has achieved 97.77% accuracy in the hold-out method and 97.23% accuracy in the cross-validation method. Similarly, all proposed methods have achieved good in terms of both the accuracies except naïve bayes which has achieved the least. [Fig. 12](#) gives a graphical representation of the hold-out accuracies and cross-validation accuracies.

[Fig. 13](#), a line graph is a diagrammatical representation of the hold-out and cross-validation accuracies of various proposed machine-learning models. From [Figs. 12](#) and [13](#), it is clear that the models built on the novel symptomatic dataset, are robust. The two accuracies have minor differences so the proposed models are reliable and consistent and have not suffered from any bias.

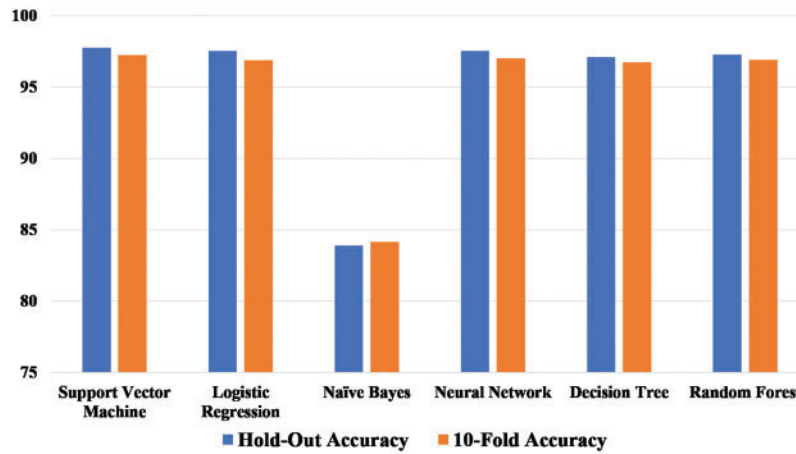


Figure 12: Comparison of hold-out and cross-validation accuracies

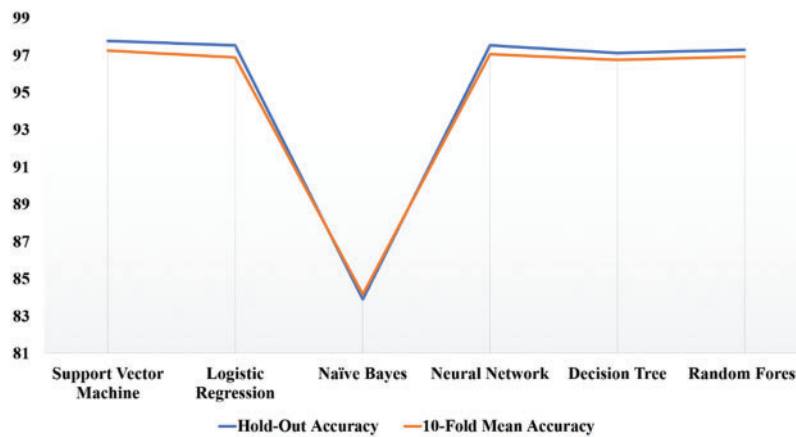


Figure 13: Graphical representation of hold-out and cross-validation accuracies

7.5.2 Comparative Analysis of the Proposed Work with the Existing Work

The work that is being presented is examined in conjunction with articles that have been published in the past on methods that are based on machine learning. For this comparative analysis, we selected the most recent papers that were also the most pertinent. Our proposed research is based on symptomatic data in order to anticipate and diagnose COVID-19 in its earliest phases, hence reducing the infection’s spread and the number of fatalities. In addition to the aforementioned aims, one of our key goals is to develop a rapid and inexpensive approach to recognizing and preventing the spread of COVID-19. The proposed system employs an edge-based Internet of Things (IoT) environment to communicate time-sensitive data efficiently and inexpensively. In this work, a variety of machine learning and ensemble learning models are presented, and their performance is compared to that of previous research in related fields. The findings of our study, which are based on an environment at the edge of the Internet of Things (IoT), have an accuracy rate of 97.767 percent. The work that has been proposed, to the best of our knowledge, has reached the highest achievable level of accuracy. Only a handful of the publications on COVID-19 identification are based on symptoms, whereas the vast bulk of the research on COVID-19 identification is based on CT scans. [Table 8](#) displays the most

up-to-date findings from research on COVID-19 based on a variety of criteria. The dataset that was used, the methods that were applied, the framework that was applied, and the valuable information regarding the accuracy that each obtained are all listed in separate columns in the table.

Table 8: Comparative analysis of proposed work with existing work

Authors	Technique	Dataset	Computing framework	Accuracy
[44]	WEKA Tool various algorithms	1476 × 7 records 854 positives and 622 negatives	IoT cloud-based system	92%
[67]	Deep learning (CNN)	618 CT samples	N/A	86.7%
[68]	Gradient-boosting with decision-tree	51,831 tested individuals (of whom 4769 were confirmed to have COVID-19)	N/A	95%
[69]	Machine learning (LR, SVM, NN, KNN)	Medical history, demographic data, as well as COVID-19-related information	N/A	93.75%
[70]	Transfer-learning model	1065 CT images of pathogen-confirmed COVID-19	N/A	79.3%
Proposed	Machine learning (support vector machine, logistic regression, naïve bayes, neural network, decision tree and random forest)	Primary symptomatic data collected from SKIMS 7015 × 16 with 3112 Negative instances and 3903 Positive instances	IoT edge-based system	97.767%

8 Conclusion

The COVID-19 pandemic has taken thousands of precious lives since 2019. The combination of IoT with other novel technologies such as ML, AI, and edge computing can combat the COVID-19 pandemic. The novel healthcare systems can provide time-sensitive information to caregivers and responders in real-time is the need of an hour. Traditional healthcare systems are reactive systems that are providing healthcare services to the users only after diagnosing a specific disease. In contrast to reactive healthcare systems, the article proposes a novel healthcare framework that is proactive and preventive. A novel healthcare edge-assisted framework is proposed in this study to detect and predict the COVID-19 suspect in the early stages to break off the spread of coronavirus infection and reduce the mortality rate. The framework is based on edge technology to provide localized healthcare facilities with minimal latency, fast response time, and optimal energy consumption. The framework has mainly three components: symptom collection module, user-centric module, and cloud module. The proposed framework uses sensors, wearables, and IoT devices to collect the data and apply machine learning models to detect and predict COVID-19 suspects. The framework is validated using the novel symptomatic dataset collected from the SKIMS, Srinagar, Jammu & Kashmir, India. During our study, six machine learning models have been proposed namely, support vector machine, logistic regression, naïve bayes, neural network, decision tree, and random forest. Various performance

metrics, including accuracy, precision, recall, F-score, AUC-score, and RMSE, have been used to assess how well the proposed models perform. The 97% and above accuracy is achieved by almost all proposed models except naïve bayes has achieved only 84%. In our case study, the proposed models are validated using k-cross-validation to check the reliability and consistency of the model. The proposed novel edge-assisted framework will help in identifying and predicting the COVID-19 suspect in the early stages and monitor the patient continuously. Apart from the above, the framework will provide time-sensitive information to caregivers and responders to provide emergency support. In the future, the large dataset will be used to better comprehend the nature of the disease to come up with more specific solutions. Our focus would be to propose ensemble models like Gradient Boosting, Bagging, and XGBoost algorithms. Furthermore, after getting access to a large dataset, deep learning techniques would be implemented to enhance the performance of the model.

Acknowledgement: The authors would like to thank the SKIMS (Sher-i-Kashmir Institute of Medical Sciences) for permitting us to collect the COVID-19 data from various departments.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Khan, M., Syed, F. A., Hamad, Z. A., Tahir, M. N., Saif, S. et al. (2021). COVID-19: A global challenge with old history, epidemiology and progress so far. *Molecules*, 26(1), 39. <https://doi.org/10.3390/molecules26010039>
2. Al-rohaimi, A. H., Al, F. (2020). Novel SARS-CoV-2 outbreak and COVID19 disease; a systemic review on the global pandemic. *Genes & Diseases*, 7(4), 491–501. <https://doi.org/10.1016/j.gendis.2020.06.004>
3. Jha, D. K., Shah, D. S., Gurram, S., Amin, P. D. (2020). COVID-19: A brief overview of diagnosis and treatment. *Journal of Innovations in Pharmaceutical and Biological Sciences*, 7(3), 21–38. <https://jipbs.com/index.php/journal/article/view/391>
4. Dejnirattisai, W., Huo, J., Zhou, D., Schreiber, G., Stuart, D. I. et al. (2022). SARS-CoV-2 Omicron-B.1.1.529 leads to widespread escape from neutralizing antibody responses. *Cell*, 185(3), 467–484. <https://doi.org/10.1016/j.cell.2021.12.046>
5. Singhal, T. (2022). The emergence of omicron: Challenging times are here again! *Indian Journal of Pediatrics*, 89, 490–496. <https://doi.org/10.1007/s12098-022-04077-4>
6. Aslam, B., Javed, A. R., Chakraborty, C., Nebhen, J., Raqib, S. et al. (2021). Blockchain and ANFIS empowered IoMT application for privacy preserved contact tracing in COVID-19 pandemic. *Personal and Ubiquitous Computing*, 65(1), 102589. <https://doi.org/10.1007/s00779-021-01596-3>
7. Gokul, K. (2021). Employment challenges and trends in India's rural labour market with the Impact of MGNREGA. *Social Science Journal for Advanced Research*, 1(1), 18–24. <https://doi.org/10.31033/SSJAR/2021010104>
8. Ambika, S., Basappa, U., Singh, A., Gonugade, V., Tholiya, R. (2020). Impact of social lockdown due to COVID-19 on environmental and health risk indices in India. *Environmental Research*, 196, 110932. <https://doi.org/10.1016/j.envres.2021.110932>
9. Sahu, K. K., Kumar, R. (2020). Current perspective on pandemic of COVID-19 in the United States. *Journal of Family Medicine and Primary Care*, 9(4), 1784–1791. https://doi.org/10.4103/jfmpe.jfmpe_424_20

10. Gregory, L. W., Xiong, D., Lu, Z., Joseph, A. Z., John, S. et al. (2021). Pandemic velocity: Forecasting COVID-19 in the US with a machine learning & Bayesian time series compartmental model. *PLoS Computational Biology*, 17(3), 1008837. <https://doi.org/10.1371/journal.pcbi.1008837>
11. Khalid, A., Ali, S. (2020). COVID-19 and its challenges for the healthcare system in Pakistan. *Asian Bioethics Review*, 12(4), 551–564. <https://doi.org/10.1007/s41649-020-00139-x>
12. Rajesh, R. S. M. (2021). Application of machine learning (ML) and Internet of Things (IoT) in healthcare to predict and tackle pandemic situation. *Distributed and Parallel Databases*, 40, 887. <https://doi.org/10.1007/s10619-021-07358-7>
13. Mahalle, P. N., Sonawane, S. S. (2021). Internet of Things in healthcare. *Springer Briefs in Applied Sciences and Technology*, 13–25. https://doi.org/10.1007/978-981-33-6460-8_2
14. Shetty, B., Fernandes, R., Rodrigues, A. P., Chengoden, R., Bhattacharya, S. et al. (2022). Skin lesion classification of dermoscopic images using machine learning and convolutional neural network. *Scientific Reports*, 12(1), 1–11. <https://doi.org/10.1038/s41598-022-22644-9>
15. Sood, S. K., Sood, V., Mahajan, I., Sahil (2022). Fog-cloud assisted IoT-based hierarchical approach for controlling dengue infection. *The Computer Journal*, 65(1), 67–79. <https://doi.org/10.1093/comjnl/bxaa005>
16. Yang, Y., Wang, W., Liu, L., Dev, K., Qureshi, N. M. F. (2022). AoI optimization in the UAV-aided traffic monitoring network under attack: A stackelberg game viewpoint. *IEEE Transactions on Intelligent Transportation Systems*, 24(1), 1–10. <https://doi.org/10.1109/TITS.2022.3157394>
17. Wang, W., Srivastava, G., Lin, J. C. W., Yang, Y., Alazab, M. et al. (2022). Data freshness optimization under CAA in the UAV-Aided MECN: A potential game perspective. *IEEE Transactions on Intelligent Transportation Systems*. <https://doi.org/10.1109/TITS.2022.3226493>
18. Siriwardhana, Y., de Alwis, C., Gur, G., Ylianttila, M., Liyanage, M. (2020). The fight against the COVID-19 pandemic with 5G technologies. *IEEE Engineering Management Review*, 48(3), 72–84. <https://doi.org/10.1109/EMR.2020.3017451>
19. Chakraborty, C., Abougreen, A. N. (2021). Intelligent Internet of Things and advanced machine learning techniques for COVID-19. *EAI Endorsed Transactions on Pervasive Health and Technology*, 7(26), 1–14. <https://doi.org/10.4108/eai.28-1-2021.168505>
20. Wang, W., Yang, Y., Yin, Z., Dev, K., Zhou, X. et al. (2022). BSIF: Blockchain-based secure, interactive, and fair mobile crowdsensing. *IEEE Journal on Selected Areas in Communications*, 40(12), 3452–3469. <https://doi.org/10.1109/JSAC.2022.3213306>
21. Garg, L., Chukwu, E., Nasser, N., Chakraborty, C., Garg, G. (2020). Anonymity preserving IoT-based COVID-19 and other infectious disease contact tracing model. *IEEE Access*, 8, 159402–159414. <https://doi.org/10.1109/ACCESS.2020.3020513>
22. Dong, Y., Yao, Y. D. (2021). IoT platform for COVID-19 prevention and control: A survey. *IEEE Access*, 9, 49929–49941. <https://doi.org/10.1109/ACCESS.2021.3068276>
23. Dash, S., Chakraborty, C., Giri, S. K., Pani, S. K., Frnda, J. (2021). BIFM: Big-data driven intelligent forecasting model for COVID-19. *IEEE Access*, 9, 97505–97517. <https://doi.org/10.1109/ACCESS.2021.3094658>
24. Mir, M. H., Ravindran, D. (2017). LETISA: Latency optimal edge computing technique for IoT based smart applications. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 2(4), 688–694. <http://ijsrcseit.com/paper/CSEIT1724178.pdf>
25. Li, H., Ota, K., Dong, M. (2018). Learning IoT in edge: Deep learning for the Internet of Things with edge computing. *IEEE Network*, 32(1), 96–101. <https://doi.org/10.1109/MNET.2018.1700202>
26. Hartmann, M., Hashmi, U. S., Imran, A. (2019). Edge computing in smart health care systems: Review, challenges, and research directions. *Transactions on Emerging Telecommunications Technologies*, 33(3), e3710. <https://doi.org/10.1002/ett.3710>

27. Haghi, K. M., Madanipour, M., Nikravan, M., Asghari, P., Mahdipour, E. (2021). A systematic review of IoT in healthcare: Applications, techniques, and trends. *Journal of Network and Computer Applications*, 192, 103164. <https://doi.org/10.1016/j.jnca.2021.103164>
28. Habibzadeh, H., Dinesh, K., Rajabi Shishvan, O., Boggio-Dandry, A., Sharma, G. et al. (2020). A survey of Healthcare Internet of Things (HIoT): A clinical perspective. *IEEE Internet of Things Journal*, 7(1), 53–71. <https://doi.org/10.1109/JIOT.2019.2946359>
29. Bhardwaj, V., Joshi, R., Gaur, A. M. (2022). IoT-based smart health monitoring system for COVID-19. *SN Computer Science*, 3(2), 1–11. <https://doi.org/10.1007/s42979-022-01015-1>
30. Mondal, S., Mitra, P. (2022). The role of emerging technologies to fight against COVID-19 pandemic: An exploratory review. *Transactions of the Indian National Academy of Engineering*, 7, 157–174. <https://doi.org/10.1007/s41403-022-00322-6>
31. Verma, P., Sood, S. K. (2018). Cloud-centric IoT based disease diagnosis healthcare framework. *Journal of Parallel and Distributed Computing*, 116, 27–38. <https://doi.org/10.1016/j.jpdc.2017.11.018>
32. Sood, S. K., Mahajan, I. (2019). IoT-fog-based healthcare framework to identify and control hypertension attack. *IEEE Internet of Things Journal*, 6(2), 1920–1927. <https://doi.org/10.1109/JIOT.2018.2871630>
33. Verma, P., Sood, S. K., Kalra, S. (2018). Cloud-centric IoT based student healthcare monitoring framework. *Journal of Ambient Intelligence and Humanized Computing*, 9(5), 1293–1309. <https://doi.org/10.1007/s12652-017-0520-6>
34. Bukhari, S. N. H., Jain, A., Haq, E., Khder, M. A., Neware, R. et al. (2021). Machine learning-based ensemble model for zika virus t-cell epitope prediction. *Journal of Healthcare Engineering*, 2021, 1–10. <https://doi.org/10.1155/2021/9591670>
35. Kavitha, D., Ravikumar, S. (2021). IOT and context-aware learning-based optimal neural network model for real-time health monitoring. *Transactions on Emerging Telecommunications Technologies*, 32(1), 1–19. <https://doi.org/10.1002/ett.4132>
36. Nasajpour, M., Pouriye, S., Parizi, R. M., Dorodchi, M., Valero, M. et al. (2020). Internet of things for current COVID-19 and future pandemics: An exploratory study. *Journal of Healthcare Informatics Research*, 4(4), 325–364. <https://doi.org/10.1007/s41666-020-00080-6>
37. Singh, R. P., Javaid, M., Haleem, A., Suman, R. (2020). Internet of Things (IoT) applications to fight against COVID-19 pandemic. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, 14(4), 521–524. <https://doi.org/10.1016/j.dsx.2020.04.041>
38. Ajaz, F., Naseem, M., Sharma, S., Shabaz, M., Dhiman, G. (2021). COVID-19: Challenges and its technological solutions using IoT. *Current Medical Imaging Formerly Current Medical Imaging Reviews*, 18(2), 113–123. <https://doi.org/10.2174/18756603MTE0fMjEE3>
39. Ahmadi, Z., Haghi Kashani, M., Nikravan, M., Mahdipour, E. (2021). Fog-based healthcare systems: A systematic review. *Multimedia Tools and Applications*, 80, 36361–36400. <https://doi.org/10.1007/s11042-021-11227-x>
40. Verma, P., Sood, S. K. (2018). Fog assisted-IoT enabled patient health monitoring in smart homes. *IEEE Internet of Things Journal*, 5(3), 1789–1796. <https://doi.org/10.1109/JIOT.2018.2803201>
41. Nguyen, T. T., Nguyen, Q. V. H., Nguyen, D. T., Hsu, E. B., Yang, S. et al. (2020). Artificial intelligence in the battle against coronavirus (COVID-19): A survey and future research directions. *arXiv:2008.07343*. <http://arxiv.org/abs/2008.07343>.
42. Yang, Y., Wang, W., Yin, Z., Xu, R., Zhou, X. et al. (2022). Mixed game-based AoI optimization for combating COVID-19 with AI Bots. *IEEE Journal on Selected Areas in Communications*, 40(11), 3122–3138. <https://doi.org/10.1109/JSAC.2022.3215508>
43. Elbasi, E., Mathew, S., Topcu, A. E., Abdelbaki, W. (2021). A survey on machine learning and Internet of Things for COVID-19. *2021 IEEE world AI IoT congress*, pp. 115–120. <https://doi.org/10.1109/AIIoT52608.2021.9454241>

44. Otoom, M., Otoum, N., Alzubaidi, M. A., Etoom, Y. (2020). An IoT-based framework for early identification and monitoring of COVID-19 cases. *Biomedical Signal Processing and Control*, 62, 102149.
45. Imad, M., Hussain, A., Hassan, M. A., Butt, Z., Sahar, N. U. (2022). IoT based machine learning and deep learning platform for COVID-19 prevention and control: A systematic review. *Lecture Notes on Data Engineering and Communications Technologies*, 105, 523–536. <https://doi.org/10.1007/978-3-030-90618-4>
46. Castiglione, A., Umer, M., Sadiq, S., Obaidat, M. S., Vijayakumar, P. (2021). The role of Internet of Things to control the outbreak of COVID-19 pandemic. *IEEE Internet of Things Journal*, 8(21), 16072–16082. <https://doi.org/10.1109/JIOT.2021.3070306>
47. Ashraf, M. U., Hannan, A., Cheema, S. M., Ali, Z., Jambi, K. M. et al. (2020). Detection and tracking contagion using IoT-Edge technologies: Confronting COVID-19 pandemic. *2nd International Conference on Electrical, Communication and Computer Engineering*, pp. 1–6. Istanbul, Turkey. <https://doi.org/10.1109/ICECCE49384.2020.9179284>
48. Mir, M. H., Jamwal, S., Mehbodniya, A., Garg, T., Iqbal, U. et al. (2022). IoT-enabled framework for early detection and prediction of COVID-19 suspects by leveraging machine learning in cloud. *Journal of Healthcare Engineering*, 2022, 7713939. <https://doi.org/10.1155/2022/7713939>
49. Greco, L., Percannella, G., Ritrovato, P., Tortorella, F., Vento, M. (2020). Trends in IoT based solutions for health care: Moving AI to the edge. *Pattern Recognition Letters*, 135, 346–353. <https://doi.org/10.1016/j.patrec.2020.05.016>
50. Rahman, M. A., Shamim Hossain, M. (2021). An Internet-of-Medical-Things-enabled edge computing framework for tackling COVID-19. *IEEE Internet of Things Journal*, 8(21), 15847–15854. <https://doi.org/10.1109/JIOT.2021.3051080>
51. Dewanto, S., Alexandra, M., Surantha, N. (2020). Heart rate monitoring with smart wearables using edge computing. *International Journal of Advanced Computer Science and Applications*, 11(3), 141–148. <https://doi.org/10.14569/issn.2156-5570>
52. Malik, M., Iqbal, M. W., Shahzad, S. K., Mushtaq, M. T., Naqvi, M. R. et al. (2022). Determination of COVID-19 patients using machine learning algorithms. *Intelligent Automation and Soft Computing*, 31(1), 207–222. <https://doi.org/10.32604/iasec.2022.018753>
53. Mohammed, M. N., Syamsudin, H., Al-Zubaidi, S., Sairah, A. K., Ramli, R. et al. (2020). Novel COVID-19 detection and diagnosis system using IoT based smart helmet. *International Journal of Psychosocial Rehabilitation*, 24(7), 2296–2303. <https://doi.org/10.37200/IJPR/V24I7/PR270221>
54. Shabaz, M., Garg, U. (2022). Predicting future diseases based on existing health status using link prediction. *World Journal of Engineering*, 19(1), 29–32. <https://doi.org/10.1108/WJE-10-2020-0533>
55. Islam, S. M. R., Kwak, D., Kabir, M. H., Hossain, M., Kwak, K. S. (2015). The Internet of Things for health care: A comprehensive survey. *IEEE Access*, 3, 678–708. <https://doi.org/10.1109/ACCESS.2015.2437951>
56. Iqbal, U., Mir, A. H. (2020). Secure and scalable access control protocol for IoT environment. *Internet of Things*, 12(4), 100291. <https://doi.org/10.1016/j.iot.2020.100291>
57. Mahdavinejad, M. S., Rezvan, M., Berekatain, M., Adibi, P., Barnaghi, P. et al. (2017). Machine learning for Internet of Things data analysis. *Digital Communications and Networks*, 4(3), 161–175. <https://doi.org/https://doi.org/10.1016/j.dcan.2017.10.002>.
58. Francis, T. (2018). A comparison of cloud execution mechanisms fog, edge, and clone cloud computing. *International Journal of Electrical and Computer Engineering (IJECE)*, 8(6), 4646–4653. <https://doi.org/https://doi.org/10.11591/ijece.v8i6.pp4646-4653>
59. Ghasempour, A. (2019). Internet of Things in smart grid: Architecture, applications, services, key technologies, and challenges. *Inventions*, 4(1), 22. <https://doi.org/10.3390/inventions4010022>
60. Li, W., Chai, Y., Khan, F., Jan, S. R. U., Verma, S. et al. (2021). A comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system. *Mobile Networks and Applications*, 26(1), 234–252. <https://doi.org/10.1007/s11036-020-01700-6>

61. Khanday, A. M. U. D., Rabani, S. T., Khan, Q. R., Rouf, N., Mohi Ud Din, M. (2020). Machine learning based approaches for detecting COVID-19 using clinical text data. *International Journal of Information Technology*, 12(3), 731–739. <https://doi.org/10.1007/s41870-020-00495-9>
62. Mir, M. H., Jamwal, S., Islam, S., Khan, Q. R. (2021). Machine learning techniques and computing technologies for IoT based smart healthcare (COVID-19 Case Study). *3rd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, pp. 2089–2095. <https://doi.org/10.1109/ICAC3N53548.2021.9725648>
63. Darwish, A., Hassanien, A. E., Elhoseny, M., Sangaiah, A. K., Muhammad, K. (2019). The impact of the hybrid platform of Internet of Things and cloud computing on healthcare systems: Opportunities, challenges, and open problems. *Journal of Ambient Intelligence and Humanized Computing*, 10(10), 4151–4166. <https://doi.org/10.1007/s12652-017-0659-1>
64. Cisco (2015). Fog computing and the Internet of Things: Extend the cloud to where the things are. *Cisco*, 1–6. www.cisco.com
65. Islam, S., Jamwal, S., Mir, M. H. (2021). Leveraging fog computing for smart Internet of Thing scrop monitoring farming in COVID-19 era. *Annals of the Romanian Society for Cell Biology*, 25(6), 10410–10420.
66. Chen, M., Li, W., Hao, Y., Qian, Y., Humar, I. (2018). Edge cognitive computing based smart healthcare system. *Future Generation Computer Systems*, 86, 403–411. <https://doi.org/10.1016/j.future.2018.03.054>
67. Xu, X., Jiang, X., Ma, C., Du, P., Li, X. et al. (2020). A deep learning system to screen novel coronavirus disease 2019 pneumonia. *Engineering*, 6(10), 1122–1129. <https://doi.org/10.1016/j.eng.2020.04.010>
68. Zoabi, Y., Deri-Rozov, S., Shomron, N. (2021). Machine learning-based prediction of COVID-19 diagnosis based on symptoms. *npj Digital Medicine*, 4(1), 1–5. <https://doi.org/10.1038/s41746-020-00372-6>
69. Quiroz-Juárez, M. A., Torres-Gómez, A., Hoyo-Ulloa, I., León-Montiel, R. D. J., U'Ren, A. B. (2021). Identification of high-risk COVID-19 patients using machine learning. *PLoS One*, 16(9), 1–21. <https://doi.org/10.1371/journal.pone.0257234>
70. Wang, S., Kang, B., Ma, J., Zeng, X., Xiao, M. et al. (2021). A deep learning algorithm using CT images to screen for Corona virus disease (COVID-19). *European Radiology*, 31(8), 6096–6104. <https://doi.org/10.1007/s00330-021-07715-1>