iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 Issue: XI Month of publication: November 2020

DOI: https://doi.org/10.22214/ijraset.2020.32165

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue XI Nov 2020- Available at www.ijraset.com

Collection, Detection, and Data Analysis of SARS COVID-19 Biomarkers

Nitish Bhardwaj[1], Sakshi Sinha[2], Naman Soni[3], Mayur Shankar[4], Vasu Verma[5]
[1]*Software Engineer at LTI, Mumbai, India*
[2]*System Engineer at TCS, Chennai, India*
[3, 4, 5]*B.Tech in Electronics and Communication Engineering, MNIT Jaipur, India*

Abstract: COVID-19 is an infectious disease caused by the SARS-CoV-2 virus, was declared a pandemic by the World Health Organisation (WHO) in March 2020. By the first week of October 2020, more than 35.7 million people have tested positive worldwide. Infections have been spreading rapidly and tremendous efforts are being made to fight against the disease. Due to the burden being faced by the healthcare system we intend to propose a system with which we can categorize the level of attention needed to patients from medical staff and help in the prediction of COVID-19 susceptibility through data collection from our proposed handheld device. In this paper, we attempt to collect the various COVID-19 biomarkers readings and analyze data with data science techniques, where we define data science broadly to encompass the various methods and tools—including those from machine learning (ML), modeling, simulation, statistics, data visualization—that can be used to store, process, and extract insights from data. With the help of data, we intend to predict the COVID-19 susceptibility and health monitoring of its patients. As well as we collected data from public datasets, local hospitals, and repositories that can be used for further work to track COVID-19 symptoms, spread, and mitigation strategies.

I. IMPACT STATEMENT

Data science, broadly defined, will play a crucial role in the global response to the COVID-19 pandemic. Our proposed system will facilitate the rapid engagement of data science with the breadth of the ongoing research work. In particular, it will reduce the burden being faced by the healthcare system and identify the major challenges involved, promising directions for further work, and important community resources. Given the interdisciplinary nature of the challenge, this proposed system will help healthcare to have a unified system for all the patients where data on identification, monitoring, and quarantining can be easily available for healthcare systems.

II. INTRODUCTION

A whopping 4.4 lakh people are currently in quarantine across the country as authorities made concerted efforts to mitigate the spread of the novel coronavirus. Suspected COVID-19 patients, asymptomatic and mild-positive cases, and high-risk contacts of infected people are being kept in quarantine to check the virus spread. COVID-19 patients are allowed home quarantine and are being constantly monitored by experts through telephone or video calls. Though health departments of all states are strained to the limits to contain the crisis, the lapses in screening, monitoring and lack of a proper mechanism to send people into home quarantine is another challenge altogether adding to fears of a full-blown possibility of community transmission.

The Indian government has introduced an app called Arogya setu. It has been developed to keep track of coronavirus patients and safeguard the public. It offers a simple user interface and uses bluetooth technology for most of its fundamental features. It does the job of safeguarding people from the virus to the maximum potency of the digital space.

The receptiveness across multiple languages is a definite boon to India’s diverse audience. It also offers a self-testing chat space, which can soothe the minds of hypochondriacs.

Since the growing number of positive patients are choosing home care over hospitalization we would like to introduce an online solution for improving and optimizing the efforts spent on tracking the progress of every patient in-home quarantine. Based on a few biomarkers, our model aims to detect any upcoming symptoms which can help address the COVID-19 challenges faced by the patients. These biomarkers i.e. Spo2 level, heart rate, cough, difficulty in breathing, chest pain, body temperature, rate of respiration contribute to present a biometric analysis which in turn detects the early onset of Corona. It also presents a holistic view of the ongoing symptoms of such patients to the appointed officials, making the process of monitoring relatively faster and easier than ever before.

Our proposed integration with the existing system is to increase the viability of care and attention needed by the vast population of infected people facilitating convenience. We used flutter for the development of our prototype, which predicts the susceptibility of the patients registered through the Arogya setu app using the data entailed by the biomarkers. By using the k-Nearest neighbor algorithm (Machine learning) the susceptibility rate is predicted**.** This paper builds upon recent reviews and prospective papers, to help systematize existing resources and support the algorithm (Machine learning) the susceptibility rate is predicted**.** This research paper reposes upon the recent reviews and paper's perspective , to help systematize existing resources and support the research community in building solutions to the COVID-19 pandemic. We have attempted to be extensive, however, in a rapidly-evolving field such as this, it is not possible to aim for exhaustiveness. Nonetheless, we hope that our work will provide a useful introduction to the field for researchers interested in this area. research community in building solutions to the COVID-19 pandemic. We hope that our work will provide a useful introduction to the field for researchers interested in this area.

III. ARCHITECTURE

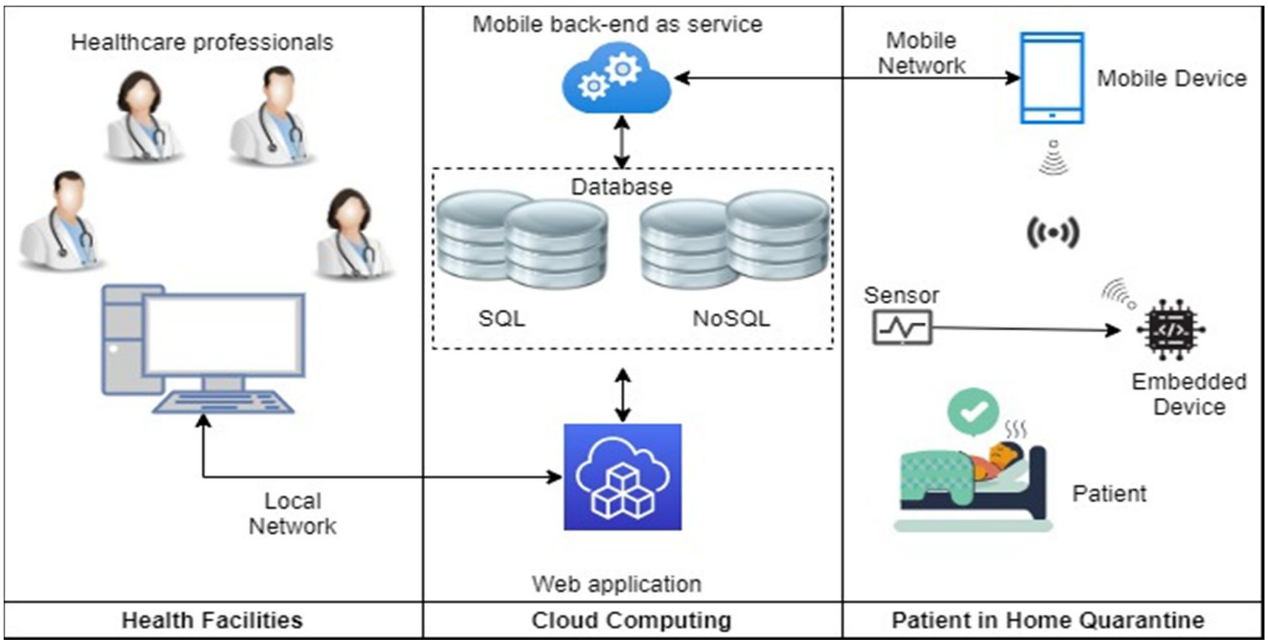


Fig. 1 Proposed Application Architecture

We have divided the application into three parts, one where the healthcare can monitor and get a prompt response from the system about the patients who need attention. In the second part where the data is collected and processed, analyzed, conclusions are made with the help of data science techniques. In the third part, the patient data is collected and sent to the server.

Healthcare professions can have regular monitoring of the patients who are home quarantined. This helps the hospital staff to have less burden and focus on the patients who need the most attention. Our application using machine learning automatically shows the patient at risk whose symptoms have shown a worsening condition, this can be seen in the healthcare systems. The data is available of the patient's various biomarkers and how he is feeling.

We have developed a mobile application using flutter which collects the data from the embedded device. Embedded device has sensors to record Sp02 , body temperature , heart rate and respiratory rate. These record are send further in the data analytics model from mobile application and helps in the prediction of the person's susceptibility rate and helps the infected person get the necessary attention from the healthcare professional. This helps people who are home quarantined to have a monitoring system so that in the case of one's degrading health healthcare professions can swiftly take the necessary actions.

All the magic happens in the cloud computation where the collected data from the sensors are sent in the database. This Database details can be accessed by the healthcare professions. With the help of Data science it is analysed which patient is recovering , how is his recovery rate , how people nearby his location are recovering ,how the virus is spreading in different location , what are the main bio marks which shows spikes in the infected people. This data can be further used for many research and more parameters can be added to strengthen the model.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue XI Nov 2020- Available at www.ijraset.com

IV. PROTOTYPE FOR IMPLEMENTATION

A. Requirements

1) Hardware

a) Heartbeat Measuring Sensor Module -1

b) Bolt IoT Wifi Module -1

c) Arduino UNO -1

2) Software

a) Bolt SMS Cloud

b) Plugins for Email

3) Other Requirements

a) In case using at 240v, A 4 channel 5v relay module is suggested

B. Working

The Bolt IoT Wifi module is connected to the arduino which has the circuit schematic as shown below. It has the connection with Heartbeat Measuring Module(This can be replaced with the App Data and a M3 microcontroller). Now the sensor module continuously monitors the test case(patient). The sensor detects the amount of infrared light being blocked by the finger and accordingly calculates the abnormal pulse. As it detects the abnormal levels as specified by the random data modelling using the K-nearest neighbours. It generates an alert that is sent to the cloud, this cloud has a novel API url that can trigger actions. This is used to generate a SMTP query over an email plugin which can be used to send the concerned authorities about the potential emergency and thus make arrangements for the same.

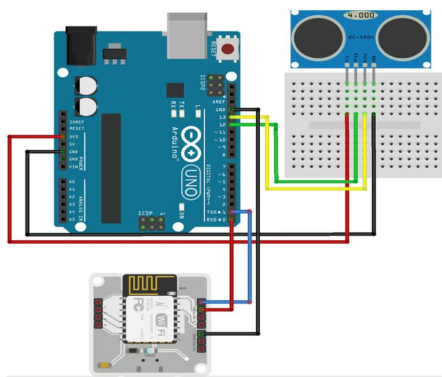


Fig. 2 In the diagram, blue module is the heart beat sensor

C. Working of Heartbeat Sensor

The Sensor consists of a sensor along with a unique circuit controlling functionality. The sensor Heartbeat Sensor has a LED (infrared) and a Light sensitive Diode, The Circuit has an IC-741(Op-Amp) and some parts that provide signal to the MCU. The working of the Sensor can be thought of better if we use the schematic and take a look at its circuit diagram.

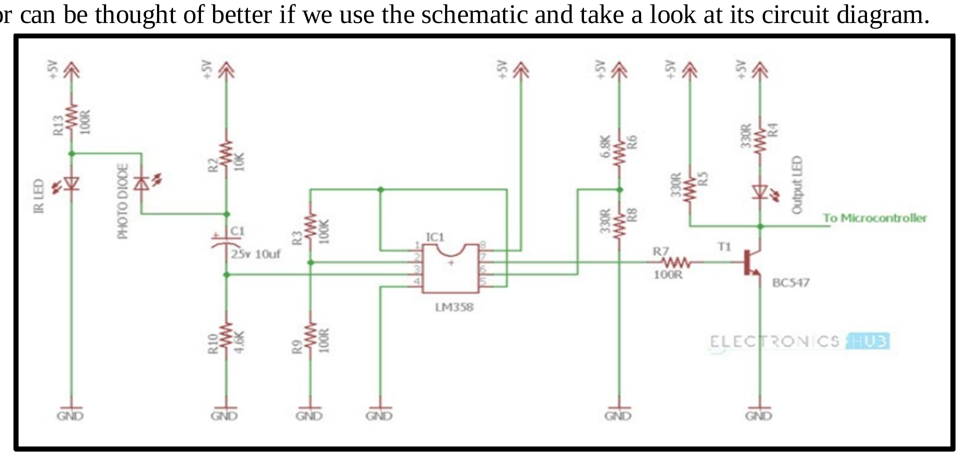


Fig. 2.1 In the diagram, the schematic of the used heartbeat sensor module is shown

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue XI Nov 2020- Available at www.ijraset.com

The light sensitive diode's output is given to the non – inverting input of the 741 IC via capacitor, which blocks the DC part of the signal. The first 741 IC is acting here as a non – inverting amplifier with an amplification factor of 1001.The output of the first op – amp is supplied to one of the inputs to the second 741 IC, which is nothing but a comparator. The output of the second 741 IC initiates a transistor, from which the signal is fed to a MCU like Arduino/M4. The Op – amp used in this circuit is LM358/ however 741Ic is interchangeable. It has two op – amps on the same circuit . Also, the transistor is a BC547. An LED, which is connected to a transistor, will blink when it detects a pulse.

D. Code sent on Bolt Cloud

In this the code on the arduino is shown, it gets the MAX_IRQ_PULSE from the K-MAP data prediction model, that can find out the next possible outcomes based on statistical analysis. This is fetched to the arduino, the limit is further set up by the critical Sp02 levels for the COVID-19 danger. As soon as the limit is reached or predicted by the model, it immediately turns the PIN of the arduino HIGH that is connected to the BOLT IoT module. This then runs the code on cloud running the SMTP query which triggers the alert.

```
unsigned long MAX_IRQ_PULSE = 0;
int p = 0;
int val = 0;
int Latest_p = LOW;
unsigned long oldMillis = 0;

void setup() {
 pinMode(2, INPUT);
 Serial.begin(9600);
 }

void loop() {
p = digitalRead(2);
if (p != Latest_p) {
  Latest_p = p;
  if (p == HIGH) MAX_IRQ_PULSE++;
}

// print and reset MAX_IRQ_PULSE every seconds
if ( millis() - oldMillis >= 10000 )
{
  oldMillis = millis();
  val = MAX_IRQ_PULSE * 6;
  if (MAX_IRQ_PULSE > 1)
  Serial.println(val);
  MAX_IRQ_PULSE = 0;
}
}
```

Fig. 3 Code pushed on arduino for abnormal Spo2 detection from the sensor

V. DATASET AND RESOURCES

The prevalent symptoms of novel coronavirus are rise in the body temperature and weakness. It is also accompanied by dry cough and certain increment in the cough levels of the lungs. It was also observed and as noted by the WHO itself that certain patients also fall prey to nasal congestions and diarrhoea. These are mild initially but get worse with time. It is also interesting to note that more than 30% of patients do not show major symptoms and are characterised as asymptomatic ones.

Around only 16.6% COVID-19 patients become critical and develop respiratory issues. This is even less in countries like India, Sri-Lanka and Bangladesh Older people, and those with prior medical problems like high heart problems or diabetes, are more prone to such critical respiratory issues. Thus, for the remaining people this prototype can prove significant in early detection of the flu and thus getting its treatment.

We have checked certain datasets from hospitals and incorporated them for our random datasets. These include Spo2 levels, WBC levels, Respiratory rate, cough, tiredness, breathing difficulty. We collected the data of infected and non-infected patients from hospitalized and home quarantined people, creating a dataset.

VI. COVID-19 CASE DATA

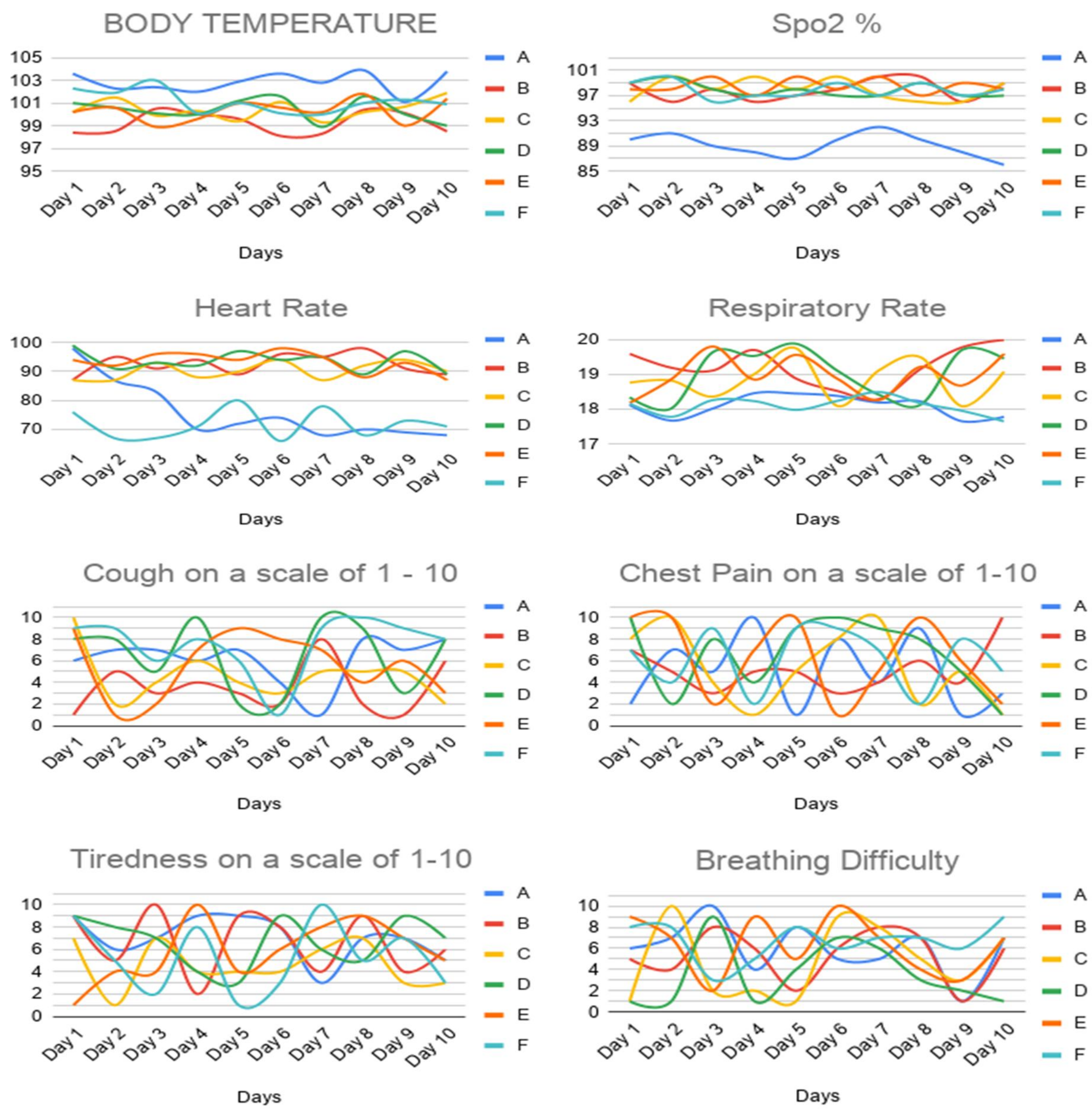


Fig. 4 Graph illustrating each patient's biomarker level over the span of 10 days

In the above graphs: A, B, C, D, E, and F are the patients, and as we can see that patient A is at severe risk and must be immediately hospitalized. Cough, Chest pain, tiredness, and breathing difficulty are measured on a scale of 1-10 (10 being the worst). Also, patient F can also be at risk and our model has predicted that he will be at risk in the upcoming days and must be hospitalized as a precautionary measure.

VII. ALGORITHM

A. Classification Problem

We are dealing with this as a classification problem as our aim is to predict whether the person is at risk or not. There are several algorithms available but we think the k-Nearest Neighbours will be most suitable for our needs. For testing purposes, we will be using the sklearn libraries and for implementation, PyTorch will be used.

The k-Nearest Neighbors algorithm or KNN for short is a machine learning classification or regression algorithm. The k in the name stands for the number of nearest neighbors that are used for matching as can be seen in Fig.5. There are many similarity techniques that can be used to calculate similarity between the neighbors like Euclidean distance, Cosine similarity etc.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue XI Nov 2020- Available at www.ijraset.com

Euclidean distance is most commonly used. Once the neighbors are discovered, the summary predictions are often made by returning the foremost common outcome or taking the typical . As such, KNN are often used for classification or regression problems. There is no model to talk of aside from holding the whole training dataset. Because no work is completed until a prediction is required, KNN is usually mentioned as a lazy learning method.

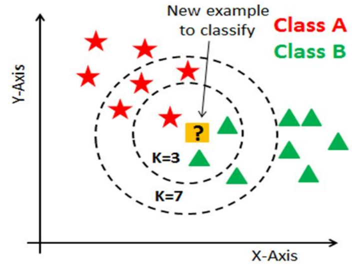


Fig. 5 For K = 3 the new example has two green triangles and one red star as its neighbors, so it is classified as a green triangle. But for K = 7 there are 3 green triangles and 4 red stars as seven closest neighbors so the algorithm classifies it as a red star

Some drawbacks of k-Nearest Neighbours are as follows:

1) Accuracy depends on the standard of the info
2) With large data, the prediction stage could be slow
3) Sensitive to the size of the info and irrelevant features
4) Require high memory – got to store all of the training data
5) Given that it stores all of the training, it are often computationally expensive

B. *Training and Modelling*

This is the most challenging aspect of this project as no data is available on the internet regarding Covid Bio Markers. We have collected data from the reports of our friends and family members who got tested and used them for our study. Some of the data is randomly generated with the range of the different biomarkers available on the internet. After training, we can predict whether a person is at risk or not, and if he is at risk, he should call for hospitalization from the app.

The main drawback of this model is that it has been trained on a small dataset, we have kept the dataset balanced and unbiased to minimize the effects of a compact dataset. Also, some of the data may not be according to the real-world data as it is randomly generated but as we have made the data according to the range this drawback is also cut down.

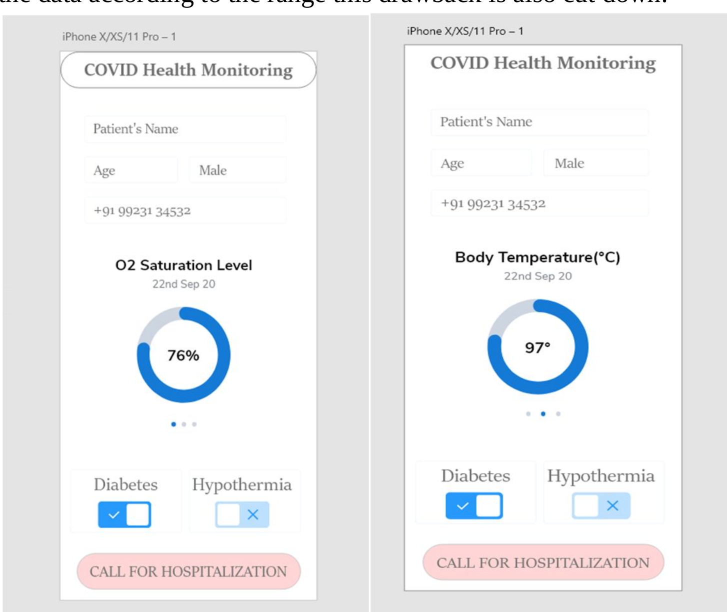


Fig. 6 UI Of the App

C. Patient Prioritization

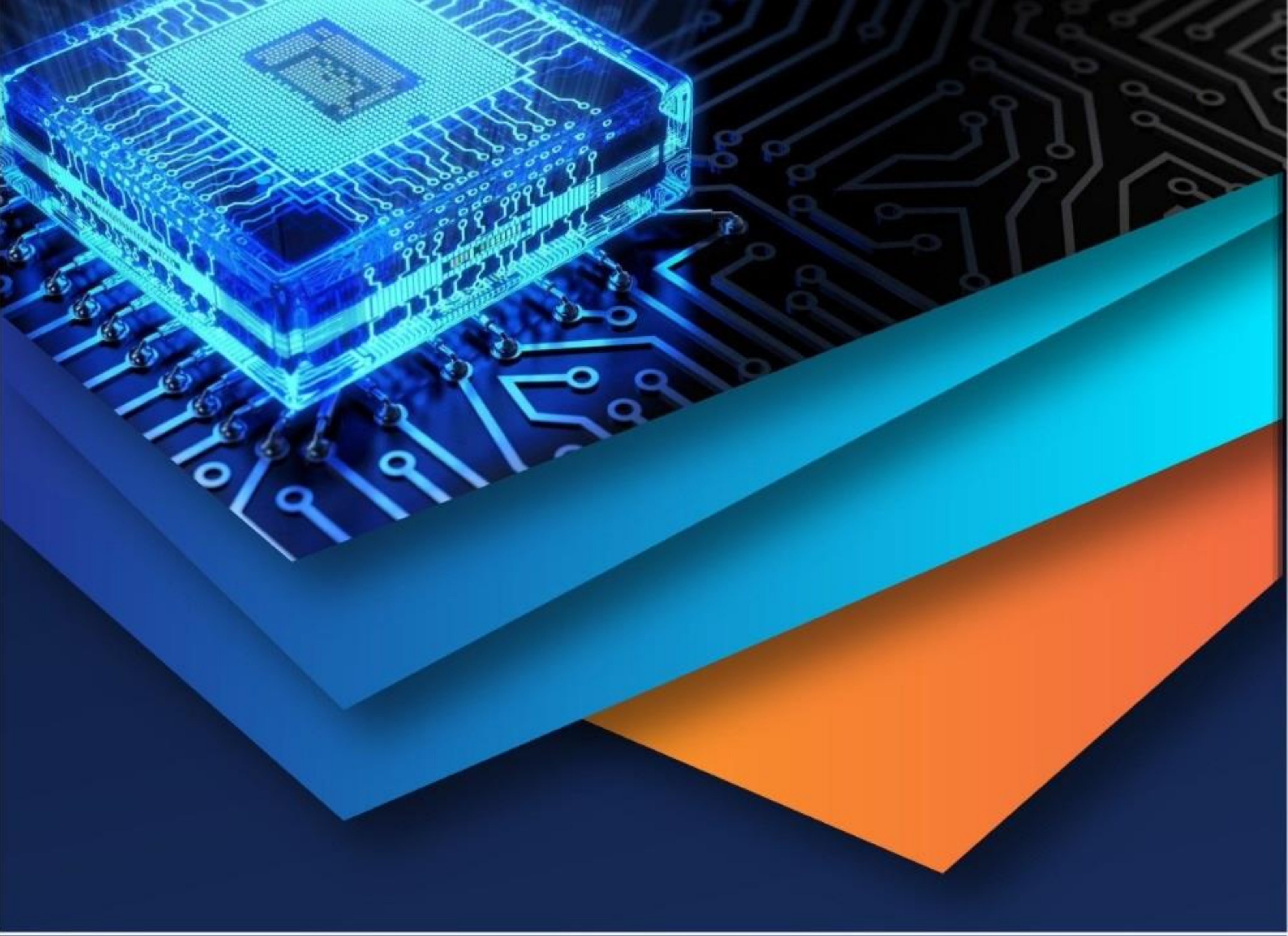
Waiting lists should be managed as fairly as possible to make sure that patients with greater or more urgent needs receive services first. The technique of making a priority list of patients according to how much they require medical attention is known as Patient Prioritization. Although there are many tools used for this purpose in healthcare services, the growing literature on this subject is focused mainly on urgent situations. Evidence has not been synthesized with reference to all the non-emergency services. In our paper, patients are prioritized in accordance with different factors like age, different diseases associated with them, and whether they are at a risk or not. So, in case of lack of hospital beds or facilities patients at more risk will be provided with the best facilities.

VIII. CHALLENGES

Data science systems typically learn and improve as a larger amount of data is gathered over time. Ideally, the dataset should be of high fidelity and voluminous. For most of the used test cases here, extensive labelled datasets are not yet available publically, e.g., for biomarkers analysis. Although there are a few publically available datasets for required biomarkers, these are small datasets as compared to the requirements of deep learning models. For example, in the case of biomedical data, sample sizes range from a few up to double digits patients. The lack of measured data is frequently due to the distributed nature of many data sources. For example, electronic healthcare records are often segregated on a national, state, or even hospital level. A key challenge is therefore, union these sources, and overcoming practical differences across each source, e.g., regarding schemas. Thus, superior and more automated approaches to data mining, data wrangling etc. may be crucial in attaining fast and reliable outcomes. Common standards and international collaboration will help. Beyond these challenges regarding data availability, there are also major challenges within the data itself. The time-critical nature of this research has caused hurdles in development of certain types of high-quality dataset. For instance, by the time data is collected, curated and annotated it can become obsolete. Due to such circumstances, COVID-19 datasets and their causal interpretations often contain poorly quantified biases. For example, daily infection rates in Italy exhibit few similarities to those in Japan. Training models on unrepresentative datasets will lead to poor and inaccurate outcomes. Considering techniques such as transfer learning could allow models to be specialized with regional characteristics, the fast-moving nature of the problem can make it problematic to perform informed model selection and parameterization. A key challenge is to devise analytical approaches that can work with these data limitations.

REFERENCES

[1] GQ India: https://www.gqindia.com/get-smart/content/aaroyga-setu-app-heres-everything-you-need-to-know#:~:text=The%20drawbacks%20of%20the%20Aarogya,could%20prove%20to%20be%20costly
[2] K-nearest neighbors: https://machinelearningmastery.com/tutorial-to-implement-k-nearest-neighbors-in-python-from-scratch/#:~:text=The%20k%2DNearest%20Neighbors%20algorithm%20or%20KNN%20for%20short%20is,a%20summarized%20prediction%20is%20made. https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn
[3] Leveraging Data Science To Combat COVID-19: A Comprehensive Review https://ieeexplore.ieee.org/document/9184922
[4] Patient Prioritization: https://systematicreviewsjournal.biomedcentral.com/articles/10.1186/s13643-019-0992-x#:~:text=Patient%20prioritization%20refers%20to%20the,in%20the%20delivery%20of%20care.
[5] https://www.electronicshub.org/heartbeat-sensor-using-arduino-heart-rate-monitor/



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY