2021

IoT pulse oximetry model for early detection of COVID-19

Bonyo, Lesley
School of Computing and Engineering Science
Strathmore University

IoT Pulse Oximetry Model for Early Detection of COVID-19

BY LESLEY BONYO

081306

A Thesis Submitted to the School of Computing and Engineering Sciences in partial fulfillment of the requirements for the award of Master of Science in Information Technology at Strathmore University

Declaration and Approval

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Lesley Auma Bonyo

24/09/2021

Approval

The thesis of Lesley Bonyo was reviewed and approved by the following:

Dr. Joseph Orero, PhD.

Senior Lecturer, School of Computing and Engineering Sciences,

Strathmore University.

Dr Bernard Shibwabo, PhD.

Director, Directorate of Graduate Studies,

Strathmore University.

Abstract

Coronavirus disease 2019 (COVID-19), which was declared a pandemic by the World Health Organization, is a respiratory illness caused by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2). With no specific treatment against SARS-CoV-2, early detection of COVID-19 is vital to effective tracking and management of the disease. For this reason, several diagnostic strategies have been implemented to identify COVID-19 infection, to test for past infection and immune response. These include molecular tests such as RT-PCR, antibody tests and medical image analysis. While the RT-PCR is the gold standard test for confirming the COVID-19 infection, it requires specialized labs and is time consuming. As an alternative, Chest X-Ray and CT images using deep learning algorithms have been used. However, because of harmful radiation doses these approaches cannot be relied on for patients' screening. Hence, there is a need for a less expensive, more accessible, and faster detection model to identify COVID-19 disease. Physiological data such as temperature and oxygen saturation can aid in COVID-19 detection and monitoring of COVID-19 patients. The symptoms for a person who is indicative of Covid-19 include shortness of breath, abnormal heartbeat, and abnormalities in lung function like the symptoms of pneumonia. Further, there is a target oxygen saturation range for patients with COVID-19 recommended by the National Institutes of Health. Such data can be continuously collected to monitor health of individuals using Internet of Things (IoT). The research tested various machine learning algorithms and implemented a low cost IoT based system with a KNN model which produced the best results. The KNN model, based on monitored oxygen saturation levels, heart rate and other COVID-19 symptoms, made predictions of person's health based on their possibilities for COVID-19 infection an accuracy of 66.67 percent. This can aid in early detection of COVID-19 symptoms to influence early testing of individuals and to assist hospitals in remote monitoring of symptoms in patients who have contracted the virus.

Keywords:

COVID-19, IoT, Pulse Oximeter, KNN

Table of Contents

Abbreviations/Acronyms

COVID-19	Coronavirus disease 2019
ELISA	Enzyme-linked immunosorbent assay
IoT	Internet of Things
KNN	K-Nearest Neighbor
MCU	Microcontroller unit
RAD	Rapid Application Development
RT-PCR	Reverse-Transcription Polymerase Chain Reaction
SPO2	Saturated Percentage of Oxygen
WHO	World Health Organization

List of Figures

List of Tables

Acknowledgements

First and foremost, I am grateful to God for enabling me to conduct this research. I would also like to acknowledge my supervisor, Dr. Joseph Orero for his constant guidance and support during this research. My sincere gratitude to the School of Computing and Engineering Sciences for the dedication shown in imparting knowledge and skills needed to succeed in this field. I also wish to express my gratitude and appreciation to Dr. Vincent Omwenga for his guidance during the research. Lastly, I thank my colleagues for constantly motivating and supporting me during the study.

Dedication

I would like to dedicate this thesis to my mum, dad, and siblings for their great support during my study.

Chapter 1: Introduction

1.1 Background of the Study

Coronavirus disease 2019 (COVID-19) is a respiratory illness caused by severe acute respiratory syndrome coronavirus-2 also referred to as SARS-CoV-2 (Narzullaev et al., 2020). SARS-CoV-2 is a virus that attacks the respiratory system and causes ailments such as cough, fever, fatigue, and breathlessness. Distinguishing characteristics of the virus include its extremely contagious nature and an incubation period of around 1 to 14 days (Chamola et al., 2020). During this period, a person infected by the virus, may or may not show any symptoms at all. Therefore, infected people with no symptoms may unknowingly contribute to the spread of SARS-CoV-2. The virus has since spread like wildfire throughout the world (Kang et al., 2020). As of March 23, 2021, there have been 122,536,880 confirmed cases of COVID-19 and 2,703,780 deaths globally(World Health Organization, 2020)(WHO, 2020b).

The rapid rise has heavily been contributed by the easy mobility across countries. This has prompted the need for immediate countermeasures to curb the catastrophic effects of the COVID-19 outbreak. In March 2020, the World Health Organization declared the COVID-19 outbreak a pandemic (WHO, 2020a). Since then, governments of different countries have been forced to close their borders and impose lockdowns to help curb further spread of the virus. With no specific treatment or vaccine against the SARS-CoV-2, several preventive measures for individuals to comply with have been set up by the World Health Organization. This is to help prevent further spread of SARS-CoV-2 virus. These include social distancing, increase in awareness of hygiene, staying at home when it is not necessary to go out, frequently washing hands, wearing masks around other people, among other measures(Chamola et al., 2020).

With these crucial times, rapid and accurate diagnosis of COVID-19 suspected cases plays a key role in timely treatment and quarantine. Several diagnostic strategies have been implemented to identify COVID-19 infection, to test for past COVID-19 infection and immune response. First, widespread molecular testing is being conducted daily to detect signs of active infections. They involve taking a sample from the upper respiratory track of an individual with a cotton swab. The sample then undergoes a polymerase chain reaction (PCR) test to detect signs of the virus's genetic material (Arpaci et al., 2021). While the RT-PCR is the gold standard test for confirming the COVID-19 infection, it is limited as it is time-consuming, and need specialized labs. Additionally, they have logistic and utility limitations while applying them to low-resource settings (Erdem et al., 2021a). As substitute, past research works have focused on Chest CT images and Chest X-Ray images using deep learning algorithms. However, because of high radiation doses they are not suitable to be used for patients' screening. There is therefore need for a faster and more accessible diagnostic model to identify the positive and negative cases of COVID-19 (Arpaci et al., 2021).

To enhance COVID-19 diagnosis and counter these challenges, several technologies are being developed to help in the early identification of potential COVID-19 cases (Cabitza et al., 2020). First, Smart devices have been used for symptomatic detection of COVID-19 in individuals. These devices integrated with sensors collect enormous amount of data of human physiology every day(Erdem et al., 2021a). This data helps in analyzing various symptoms to help in COVID-19 diagnosis. Secondly, in Medical Image Analysis, computer vision has been applied to supplement RT-PCR testing for COVID-19 diagnosis by classifying COVID-induced pneumonia from chest X-rays and CT scans(Shorten et al., 2021). Lastly, numerous voice samples of individuals have been analyzed to come up with voice detection systems for screening of COVID-19. These applications then predict whether a person has symptoms relating to COVID-19. One such system, developed by a team of researchers from Carnegie Mellon University, detects COVID-19 in individuals based on their voice (Chamola et al., 2020).

Early detection of COVID-19 is key for effective clinical management and tracking of the disease. Various clinical studies on COVID-19 have demonstrated how physiological data such as heart rate, temperature, and oxygen saturation can aid in COVID-19 diagnosis and monitoring of COVID-19 patients. The symptoms for a person who is indicative of Covid-19 include shortness of breath, abnormal heartbeat, and abnormalities in lung function like the symptoms of pneumonia (Hidayat et al., 2020). The current target blood oxygen level range as recommended by the National Institutes of Health for patients with COVID-19 is 92–96% (Shenoy et al., 2020)(Miron-Alexe, 2020)**.** Such data can be continuously collected to monitor the health of individuals using the Internet of Things (IoT) which is an interconnection of physical devices and the Internet (Kumar et al., 2020). This can be achieved by employing various sensors which collect data for analysis. This research proposes a low-cost IoT pulse oximetry model that will be used to monitor oxygen saturation levels and heart rate to help with COVID-19 detection. This can be an affordable supplement to help improve COVID-19 detection and to contain the spread of the virus. It will also assist hospitals in remote monitoring of symptoms in patients who have contracted the virus.

1.2 Problem Statement

The rapid rise in number of infected COVID-19 patients globally has caused a heavy strain on healthcare facilities. Several resources needed to detect SARS-CoV-2 have been depleted with the spread of the infection. For instance, the need for point of care monitoring of blood oxygen level (SpO_2) for large numbers of patients has led to both a supply shortage and a cost increase of commercial sensors (Metcalfe et al., 2021). These have slowed down testing and treatment of individuals affected by the virus (Cabitza et al., 2020). As of this writing, there are no specific treatments for COVID-19. RT-PCR tests are the gold standard for COVID-19 diagnosis, but they have several shortcomings ranging from long turnaround times to high false-negative results. These tests require certified laboratories, expensive equipment, and trained personnel (Cabitza et al., 2020).

Further, increased testing has led to shortages of molecular testing reagents for COVID-19 and for other molecular diagnostics (WHO, 2020b). Developing countries with poor health and scientific infrastructure are struggling to diagnose and track the disease due insufficient laboratory facilities and trained manpower. The lack of a treatment strategy has led to governments imposing uninformed measures on their citizens (Chamola et al., 2020). Lastly diagnosis using CT and X-Ray images are sometimes difficult to be used due to high costs, few available devices and radiation doses (Brinati et al., 2020)(Ai et al., 2020).

1.3 Objectives

1.3.1 General Objective

The aim of the study is to ensure early detection of COVID-19 in individuals by developing an IoT pulse oximetry model to monitor oxygen saturation levels and heart rate.

1.3.2 Specific Objectives

i) To analyze the determinants of COVID-19 diagnosis

ii) To review existing techniques used to detect COVID-19 disease.

iii) To develop an IoT pulse oximetry model for early detection of COVID-19.

iv) To test the developed model.

1.4 Research Questions

i) What factors are used to determine diagnosis of COVID-19 disease?

ii) How are existing techniques being used to detect COVID-19 disease?

iii) How can we develop an IoT pulse oximetry model to detect COVID-19 symptoms?

iv) How can we test the model developed?

1.5 Justification

Several countries are shifting from quarantine to reopening businesses throughout the world and it is critical that governments and hospitals take initiatives to ensure effective detection and tracking of COVID-19 disease. Early detection of COVID-19 is necessary to slow down the spread of the disease and avoid transmission by ensuring COVID-19 positive patients are quarantined early. This also helps in tracing and isolating their close acquaintances.

Monitoring of various COVID-19 symptoms to detect COVID-19 disease will provide easier and affordable detection to ease governments' efforts toward curbing the spread of the virus. This can help supplement RT-PCR as a second-level diagnostic procedure to improve diagnostic capabilities of hospitals by reducing the number of false negatives. This will help prevent unsafe exposure of patients and healthcare professionals in hospitals. Further, the study intends to provide guidelines for the formulation and implementation of public health policies that will help manage the spread of COVID-19. Results from this research will give insight on transmission of COVID-19 that can aid decisions on back-to-work policies. Moreover, the study will be useful to health practitioners, researchers, and academics in building their knowledge for future theory and practical research work on COVID-19 diagnosis.

1.6 Scope and Limitations

This study is limited to detecting COVID-19 disease by monitoring oxygen saturation levels and heart rate in people. Other COVID-19 symptoms will aid in building a detection model to improve on COVID-19 detection. The study will focus on the use of IoT technology and KNN algorithm in developing the system.

Chapter 2: Literature Review

2.1 Introduction

This section reviews COVID-19 disease, transmission, and diagnosis techniques. It explores various technologies that are being used to enhance COVID-19 diagnosis, machine learning techniques that can be applied in COVID-19 diagnosis and related systems.

2.2 Theoretical Framework

2.2.1 Epidemiology of Covid-19

Epidemiology is the study of virus spread in a population, and modern models consider the heterogeneity of the transmission process and the selectivity of the virus (Renganathan & Uner, 2020). COVID-19 displays different epidemiological traits when compared to previous coronavirus outbreaks of MERS-CoV and SARS-CoV. Many COVID-19 transmissions have occurred through human-to-human contact through respiratory aspirates, droplets, contacts, and feces, and aerosols transmission is highly possible according to the National Health Commission of China (Yadav, 2020).

Based on published literatures, all three epidemics caused by these three coronaviruses are linked to wild animal markets(Singhal, 2020). All populations, regardless of age and gender are generally susceptible to COVID-19. Unlike other coronaviruses, those infected with COVID-19 have shown high viral loads even with no fever or mild symptoms (Zainol Rashid et al., n.d.). These were detected in upper respiratory specimens of patients with COVID-19, and viral shedding pattern of COVID-19 patients resembles that of patients with influenza. This suggests that SARS-CoV-2 may stay around for some time like influenza viruses. Extensive measures are therefore needed to reduce human to human transmission of COVID-19 thereby controlling the spread of the virus.

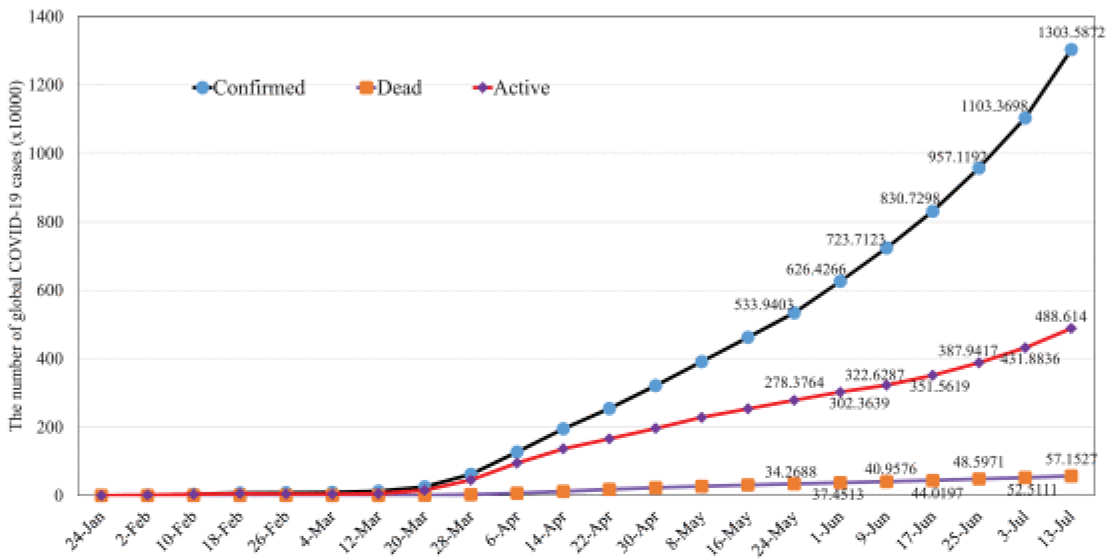

Figure 2. 1 Global COVID-19 trend (Amulya et al., 2021). *Data accurate as of July 13, 2020*

2.3 COVID-19 Disease

A wide variety of COVID-19 symptoms have been reported, including fever, body or muscle aches, headache, stuffy or congested nose, dry cough, fatigue, sore throat, and loss of taste or smell, 2-14 days after exposure to the virus. Severe symptoms like high fever, severe cough, and shortness of breath indicate the onset of pneumonia. Less common gastrointestinal symptoms like diarrhea, nausea, and vomiting have also been listed at the Centers for Disease Control and Prevention (Yadav, 2020). A study on COVID-19 indicates that an abrupt decline in Heart Rate Variability and a decrease in heart rate may signal the onset of COVID-19 before common symptoms (Buchhorn et al., 2020). COVID-19 manifests with clinical features ranging from the asymptomatic state to acute respiratory distress syndrome (ARDS) and multiple organ dysfunction syndrome (MODS). The results of a study conducted by the WHO in collaboration with China show that while most of the COVID-19 patients developed a mild to moderate disease, a few patients were diagnosed with a severe (13.8%) and a critical (6.1%) form of the same. Patients with a severe or a critical form of the disease often develop bluish lips/face and are prone to a variety of complications, including ARDS, acute heart injury, and secondary infection (Loey et al., 2020).

Most Common Symptoms	
Fever	87.9%
Dry Cough	67.7%
Fatigue	38.1%
Sputum Production	33.4%
Less Common Symptoms	
Shortness of Breath	18.6%
Myalgia / Arthralgia	14.8%
Sore Throat	13.9%
Headache	13.6%
Chills	11.4%
Rare Symptoms	
Nausea	5.0%
Nasal Congestion	4.8%
Diarrhea	3.7%
Hemoptysis (coughing up blood)	0.9%
Conjunctival Congestion	0.8%

Source: WHO

Figure 2. 2 COVID-19 Symptoms (Chamola et al., 2020)

According to the World Health Organization, testing is the key to controlling the spread of COVID-19 (Motley et al., 2020) It is essential for treatment, contact tracing and early isolation for those infected by the virus. Tests for Covid-19 are conducted in commercial, private, and state-owned health labs. Currently, COVID-19 tests are prioritized to vulnerable patients and health care workers especially in settings with limited resources (Motley et al., 2020). Focused testing in health care facilities ensures that correct prevention and control measures for the infection are implemented such that vulnerable patients who do not have COVID are protected from contracting the virus (Chamola et al., 2020). Moreover, testing among vulnerable populations is important for early treatment to minimize progression to severe diseases. Testing results can give a rough estimate of the size of the outbreak in of specific populations and be used to track trends in those areas. As of 1 June 2020, anyone with symptoms that could indicate the COVID-19 infection can be tested.

2.3.1 Oxygen Circulation and COVID-19

A key component of COVID-19 treatment has been the monitoring of a patient's Saturated Percentage of Oxygen (SpO_2), particularly with reference to respiratory conditions. The saturation of blood oxygen is the measurement of the percentage of oxygen available in the bloodstream (Ramirez Lopez et al., 2019). Studies have indicated that a great number of COVID-19 patients were found to be suffering from pneumonia from their chest x-rays taken in hospitals (Teo, 2020). These patients often have very low oxygen saturations ranging from 50% to 80% without showing any signs of breathing difficulties (Levitan, 2020). Initially, patients with COVID-19 pneumonia do not experience any noticeable breathing difficulties due to oxygen deprivation. This is usually difficult to detect causing a condition terms as “silent” hypoxia. By the time COVID-19 patients notice they are short of breath, their conditions have already significantly worsened into moderate-to-severe levels of pneumonia.

Additionally, it has been discovered that the virus attacks the lungs in an unusual way. It causes the air sacs of patients to collapse which reduce their oxygen levels leading levels. This is different from normal pneumonia infections where patient lungs fill with fluid or pus. However, COVID-19 patients are still able to maintains the lungs’ normal ability to expel carbon dioxide which is the reason why they do not feel shortness of breath in the initial stages of COVID-19 pneumonia (Teo, 2020).

The SpO_2 metric has been greatly used in predicting outcomes and triage of patients with suspected COVID-19 (Metcalfe et al., 2021). This approach is becoming necessary for clinical routine because it provides vital information on the cardiovascular function of patients. The current target blood oxygen level range as recommended by the National Institutes of Health for patients with COVID-19 is 92–96% (Shenoy et al., 2020)(Miron-Alexe, 2020)**.** The pulse oximetry is the non-invasive measurement technique of the oxygen saturation, which is an essential parameter used in medicine to determine the health status of a patient, in correlation with his lungs’ functionality and the capability of the red blood cells to carry oxygen into the body by the help of Hemoglobin proteins (Chen et al., 2020).

2.4 Covid-19 Diagnosis Techniques

2.4.1 RT-PCR Test

Reverse-Transcription Polymerase Chain Reaction (RT-PCR) is considered the gold standard for identification of SARS-CoV-2 virus (Linda J. et al., 2020). These COVID-19 tests mostly use samples obtained from the upper respiratory system using swabs. RT-PCR can be done as a one-step or a two-step procedure. One-step real-time RT-PCR uses a single tube containing the necessary primers to run the entire RT-PCR reaction. Two-step real-time RT-PCR involves more than one tube to run the separate reverse transcription and amplification reactions but offers greater flexibility and higher sensitivity than the one-step procedure(Linda J. et al., 2020). The preferred approach for detection of SARS-CoV-2 is the one-step procedure because it involves limited sample handling and is quick to set up, reducing possibility for pipetting errors and cross-contamination between the RT and real-time PCR steps (Chamola et al., 2020). These samples are then assessed in laboratories to detect the presence of viral RNA using a real-time Reverse-Transcription Polymerase Chain Reaction (rRT-PCR) test (Chen et al., 2020). A diagnosis of the COVID-19 is only confirmed if the test identifies the presence of two discriminatory targets for the SARS-CoV-2 genome, one of which is preferably explicit to the SARS-CoV-2 or the presence of beta coronavirus followed by the identification of SARS-CoV-2 using partial or complete sequencing of the virus genome (Chen et al., 2020).

Polymerase chain reaction (PCR) testing has the advantage that the primers needed for tests can be produced with relative speed as soon as the viral sequence is known. The first RT-PCR tests for detecting SARS-CoV2 were designed and distributed in January 2020 by the World Health Organization (WHO), soon after the virus was identified. However, the test protocol is complex and expensive and is mainly suited to large, centralized diagnostic laboratories. Another great challenge arises from logistic requirements for transporting clinical samples which lengthens the time patients can get test results. If the test result is positive, testing ought to repeated for verification. In patients with confirmed COVID-19 diagnosis, the laboratory evaluation should be repeated to evaluate for viral clearance prior to being released from observation. Their performance depends on many factors such as the sample types, different stage of infection in patients, the skill of sample collection, and the quality and consistency of the PCR assays being used.

There is an urgent need for an accurate and rapid test method to quickly identify many infected patients and asymptomatic carriers to prevent virus transmission and assure timely treatment of patients. However, RT-PCR faces the limitations of a complicated sample preparation, low detection efficiency, and high false-negative (Fong et al., 2020). Moreover, RT-PCR cannot be used to monitor the progress of the disease stages and cannot be applied to broad identification of past infection and immunity.

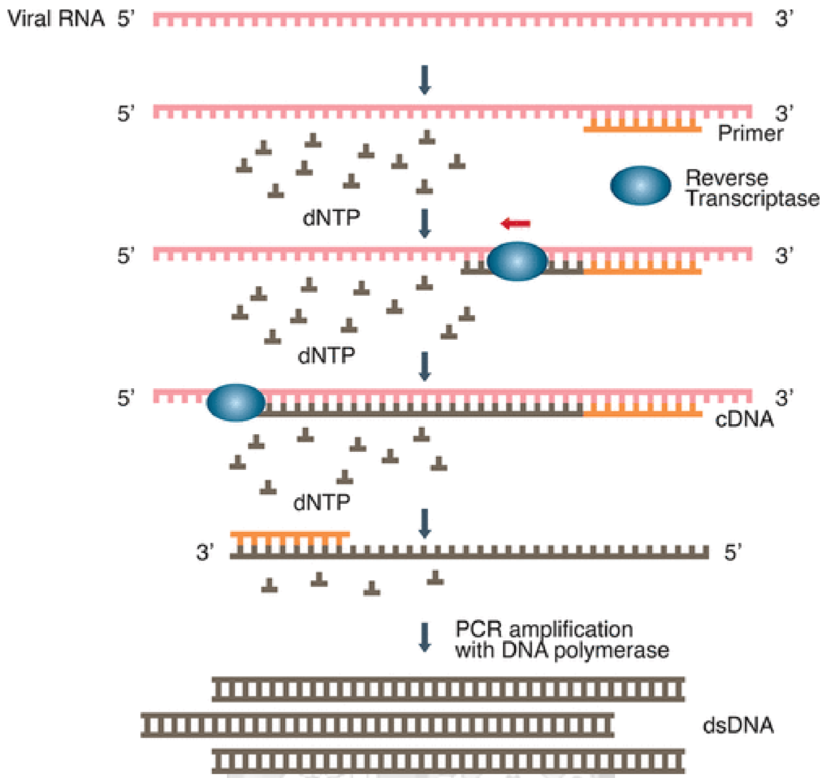

Figure 2. 3 Reverse transcription-polymerase chain reaction (RT-PCR).

2.4.2 Medical imaging inspection

Medical imaging inspection is a clinical technique used detection and diagnosis of COVID-19. It is more accurate compared to RT-PCR tests but requires expensive equipment operated by radiologists (Chen et al., 2020). COVID-19 medical image inspection mainly includes chest X-ray and lung Computerized Tomography (CT) imaging (Fong et al., 2020). CT scans of the lung reveals tissues of irregularity in patients infected with COVID-19. However, the anomalies are hard to be distinguished between that of viral pneumonia. The characteristics of COVID-19 as observed on CT scan include murky paving shades, air space contraction, traction bronchiectasis, peripheral ground glass opacities, and bronchovesicular thickening (Lerner et al., 2020).

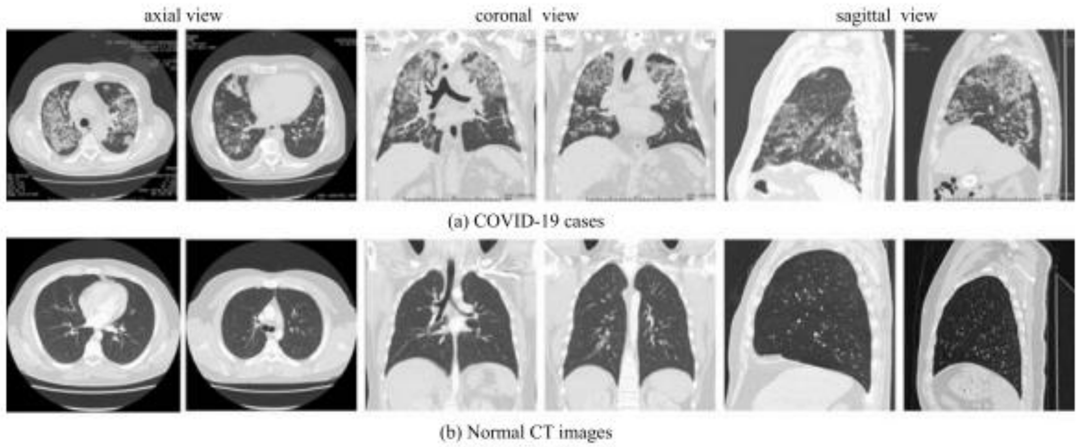

Figure 2. 4 Examples of lung CT images of normal and COVID-19 cases(Chen et al., 2020)

2.4.3 Serology Testing

Serology testing measure the antibody response in an individual (Chamola *et al.,* 2020). Antibodies are a type of protein produced as an immune response to invading pathogens (Graves et al., 2020). Unlike molecular tests, serology tests can be used to confirm suspected cases of individuals who have been exposed to the virus, developed antibodies, and have recovered. They could be used to estimate the timing of infection, dating when people were infected and distinguish people recently in close contact with the virus from people infected early during the outbreak (Amanat et al., 2020).

Serology tests can also help to identify asymptomatic carriers of the virus. They are crucial for patient contact tracing and for epidemiological studies. Serological tests are also needed for evaluation of the results of vaccine trials and development of therapeutic antibodies. Serological tests are essential to study the immune response to SARS-CoV-2 in a qualitative and quantitative manner (Linda J. et al., 2020). They help detect presence of immunoglobulin M (IgM) and immunoglobulin G (IgG) antibodies. IgM first becomes detectable in serum after a few days and lasts a couple of weeks upon infection and is followed by a switch to IgG(Linda J. et al., 2020). Thus, IgM can be an indicator of early-stage infection, and IgG can be an indicator of current or prior infection (Motley et al., 2020). IgG responses are also important for long term immunity and immunological memory (Erdem et al., 2021a). Several serological surveys in special populations are underway. Enzyme-linked immunosorbent assay (ELISA), Rapid diagnostic test

(RDT), Neutralization assay and Luminescent Immunoassay are some of the serology testing techniques used to detect antibodies.

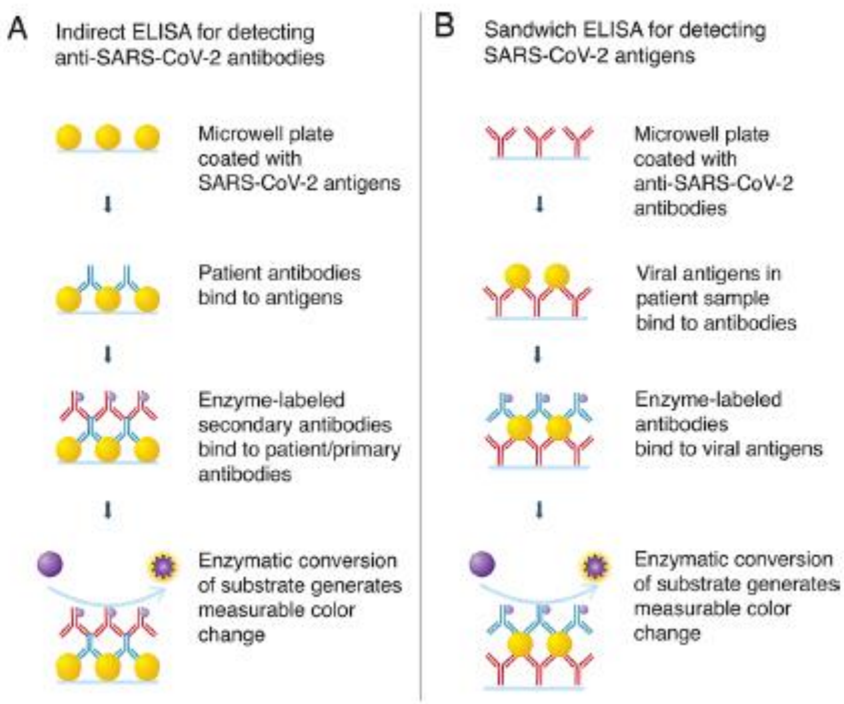

Figure 2. 5 ELISA test detecting antibodies(A) or antigens(B) (Linda J. et al., 2020)

2.5 Technologies Applied in COVID-19 Diagnosis

2.5.1 IoT

Wearable smart materials which are usually integrated with sensors collect huge amounts of data on human physiology every day (Mishra et al., 2020). Activity tracking and health monitoring via consumer wearable devices may be used for the large-scale, real-time detection of respiratory infections, often pre-symptomatically (Erdem et al., 2021b) A study by researchers from University College Cork has led to the development of a wearable device called COVID-19 Remote Early Warning System (CREW). This device detects the most common symptom in COVID-19 patients (Jeong et al., 2020). It consists of a wearable thermometer sensor that measures body temperature, a cloud system for the data collection and a mobile phone-based platform for its operation. When tested, the device identified individuals whose body temperature exceeded the critical value of the master data system and recommended necessary precautions to be taken. Lastly, researchers at Northwestern University achieved to record the severity, duration, and frequency of people's cough using a sensor to be placed in the suprasternal notch, making it possible for them to diagnose COVID-19 (Sohail, 2020).

2.5.2 Computer Vision

X-rays and CT scans which fall under medical imaging are commonly used for diagnostic and treatment purposes. They have played very important role in SARS-COV-2 virus detection during the COVID-19 pandemic (Chamola et al., 2020). For this reason, emerging technologies such as AI have been used to reinforce the imaging analysis by intelligent detection and classification of these X-rays and CT scans. For instance, AI-enabled COVID-19 disease screening can help toward automation of the screening process, to reduce human interaction among the doctors and the patients. Thus, it helps to reduce the spread of the virus and protect medical imaging professionals. Furthermore, the computer-assisted AI-enabled automated classification and recommender systems support the radiologists to make better clinical judgments with better accuracy, and efficiency.

2.5.3 Voice Detection

COVID-19 disease has led to the development of several voice-detection apps for COVID-19 screening. These applications analyze voice samples of their users' voice and decide whether a person has symptoms relating to COVID-19. One such AI application has been designed by a team of researchers from Carnegie Mellon University to detect the presence of the COVID-19 in an individual based on his/her voice (Chamola et al., 2020). When using the application, a user keys in his height and weight, followed by a request to cough three times. Afterwards, they are asked to recite any alphabet and a vowel loudly. This assists in measuring the lung capacity of the user by comparing it with thousands of other users' data, including those who are infected. The users then obtain results by getting a score out of 10. A higher score indicates that a user's features are highly like the features exhibited by COVID-19 patients. The researchers, however, have added a word of caution stating that this is not a diagnostic process and can never be substituted for tests conducted in hospitals and laboratories (Chamola et al., 2020).

2.6 Machine Learning Algorithms for Disease Prediction

2.6.1 K-nearest Neighbors (kNN)

The K-nearest neighbor (KNN) algorithm is one of the earliest simplest and classification algorithms. Unlike the Naive Bayes technique, the KNN algorithm does not require to consider probability values. The '*K*' is the KNN algorithm is the number of nearest neighbors considered to take 'vote' from. The selection of different values for '*K*' can generate different classification results for the same sample object (Uddin et al., 2019). The *k* nearest neighbors are determined based on some distance metric measure. Two mostly used distance metrics include Manhattan and Euclidean distance measure. Figure 2.6 illustrates a simplified K-nearest neighbour algorithm. When K = 3, the sample object ('star') is classified as 'black' and when K = 5 the same sample object is classified as 'red' since it now gets more 'vote' from the 'red' class

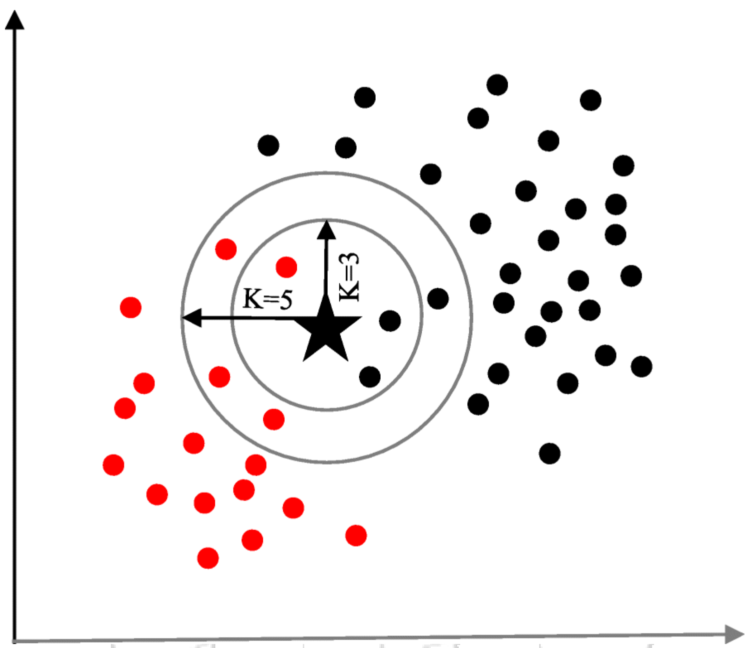

Figure 2. 6 K-nearest neighbor

KNN has been implemented in several COVID-19 detection studies. It has been used to develop a COVID-19 detection method using the complete genome sequences of human coronaviruses (Arslan & Arslan, 2021), to classify lung CT scan images into COVID-19 and normal classes (Kadry et al., n.d.), among other implementations.

2.6.2 Naive Bayes

Naïve Bayes is a simple, yet effective and commonly used, machine learning classifier. It is a classification technique based on the Bayes' theorem which describe the probability of an event based on the prior knowledge of conditions related to that event. This classifier assumes that a particular feature in a class is not directly related to any other feature although features for that class could have interdependence among themselves (Uddin et al., 2019). Naïve Bayes has been applied several complex real-world applications such as medical diagnosis, real-time prediction, spam filtering, and weather forecasting despite its oversimplified assumptions and its Naïve design. It has been used for detection of COVID-19 based on patient laboratory test collected in hospitals, in a study conducted by (Mansour et al., 2021). Naïve bayes has also been used to develop a framework for COVID-19 pneumonia to help with rapid diagnosis of COVID-19 on computed tomography scans (Akram et al., 2021).

2.6.3 Decision Trees

Decision tree (DT) is one of the earliest and prominent machine learning algorithms for the representation of classification using categorical data (Uddin et al., 2019). A Decision tree is a directed tree that obtains its structure by recursively separating the set of observations. It consists of a root with no incoming edges, internal or test nodes with exactly one outgoing edge for each, and leaves which represent the decision node and have no outgoing edges (Maimon and Rokach, 2010).

The decision tree development algorithm is a greedy algorithm which is a top-down recursive divide-and-conquer in nature. The first and the node at the highest level of the tree is called the root node where the data is partitioned first using some attribute selection measure. Then the nodes at the lower levels are also partitioned recursively 11 using a test on an attribute. The last level of the tree denotes the output class labels (Han & Kamber, 2010). DTs have been found easy to interpret and quick to learn and are a common component to many medical diagnostic protocols (Uddin et al., 2019). When traversing the tree for the classification of a sample, the outcomes of all tests at each node along the path will provide sufficient information to conjecture about its class. Figure 2.7 Shows a decision tree with its elements and rules.

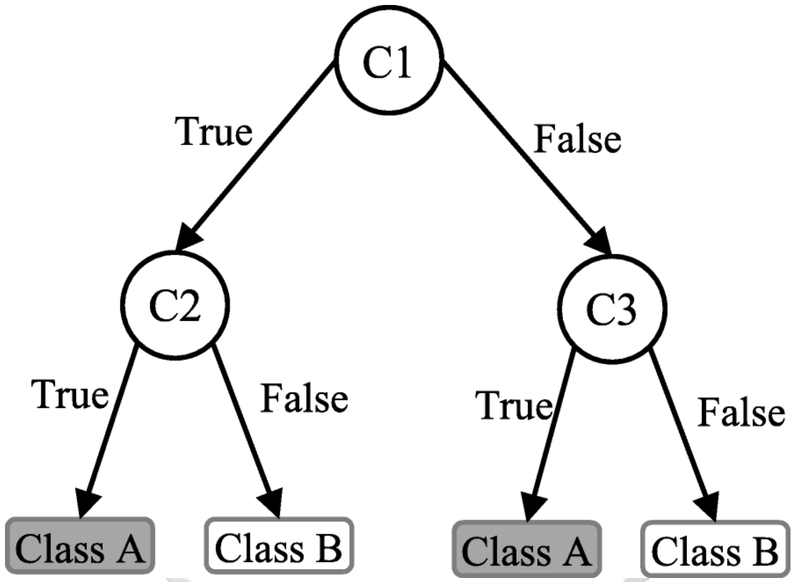

Figure 2. 7 Decision Trees

Each variable (C1, C2, and C3) is represented by a circle and the decision outcomes (Class A and Class B) are shown by rectangles. Decision trees have been used in COVID-19 diagnosis in classifying COVID-19 based on chest C-ray imaging (Yoo et al., 2020) and diagnosis of COVID-19 from routine blood tests (Alves et al., 2021).

2.6.4 Strengths and Weaknesses

Table 2. 1 Strengths and weaknesses of machine learning algorithms (Uddin et al., 2019)

ML Algorithm	Strengths	Weaknesses
Naïve Bayes	- Simple and very useful for large datasets. - Can be used for both binary and multi-class classification problems. - It requires less amount of training data. - It can make probabilistic predictions and can handle both continuous and discrete data.	- Classes must be mutually exclusive. - Presence of dependency between attributes negatively affects the classification performance. - It assumes the normal distribution of numeric attributes.
Decision Trees	- Resultant classification tree is easier to interpret. - Data preparation is easier. - Multiple data types such as numeric,	- Require classes to be mutually exclusive. - Algorithm cannot branch if any attribute or variable value for a non-leaf node is

	nominal, categorical are supported. - Can generate robust classifiers and can be validated using statistical tests.	missing. - Algorithm depends on the order of the attributes or variables.
k-NN	- Simple algorithm and can classify instances quickly. - Can handle noisy instances or instances with missing attribute values. - Can be used for classification and regression.	- Computationally expensive as the number of attributes increases. - Attributes are given equal importance, which can lead to poor classification performance. - Provide no information on which attributes are most effective in making a good classification.

2.7 Related Works

2.7.1 A Cost-Effective Pulse Oximeter Designed in Response to the COVID-19 Pandemic

(Metcalfe et al., 2021) designed as system to measure both heart rate and saturated oxygen content using dual wavelength photoplethysmography. The optical sensors were from the MAXIM range of integrated circuits such as MAX30100 and include all the required drive and sense electronics. Data in the system was transferred from the optical sensor to an Arduino Nano, which computes a time averaged heart rate and SpO_2 for display on an OLED screen.

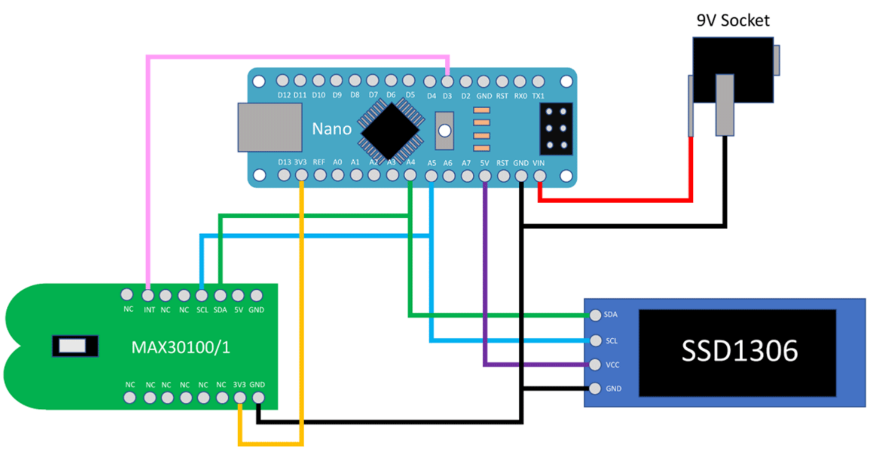

Figure 2. 8 Architectural plan of the system (Metcalfe et al., 2021)

2.7.2 Designing IoT-Based Independent Pulse Oximetry Kit as an Early Detection Tool for Covid-19 Symptoms

(Hidayat et al., 2020) focused on the development of pulse oximetry kits by utilizing IoT tools to perform remote monitoring of covid-19 patients through smartphones regarding physical and social distancing protocols. The design and development of portable pulse oximetry kit products that are equipped with GPS and its integration with IoT technology that is mass production is the focus on this research. The system works by relying on ESPDUINO-32 as a receiver or Bluetooth Low energy receiver connected with Pulse Oximetry BLE. ESPDUINO-32 and NEO-6M GPS relate to Node MCU serially to Server Databases to be monitored in real-time. Fingerprints are used as identity of the patient using this tool.

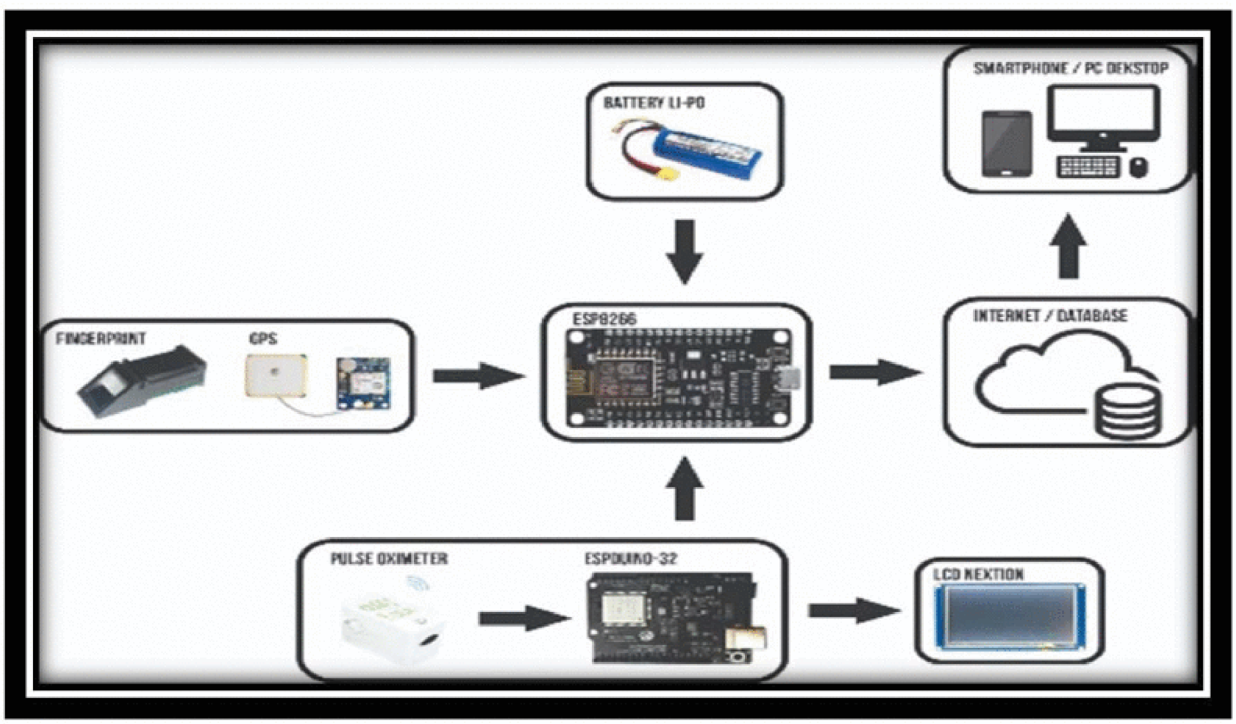

Figure 2. 9 System design (Hidayat et al., 2020)

2.7.3 IoT Pulse Oximetry Status Monitoring for Home Quarantined COVID-19 Patients

(Andreswari et al., 2020) developed an affordable, wireless, and online monitoring pulse oximetry device, for a Covid-19 infected patient that is quarantined at home. By using an Arduino based IoT embedded platform and a free online API (Application Programming

Interface) live data stream server, a family physician can remotely monitor the patient's pulse and blood oxygen level, by using a mobile phone, a tablet or a computer, without any contamination risk or contact, with the infected patient.

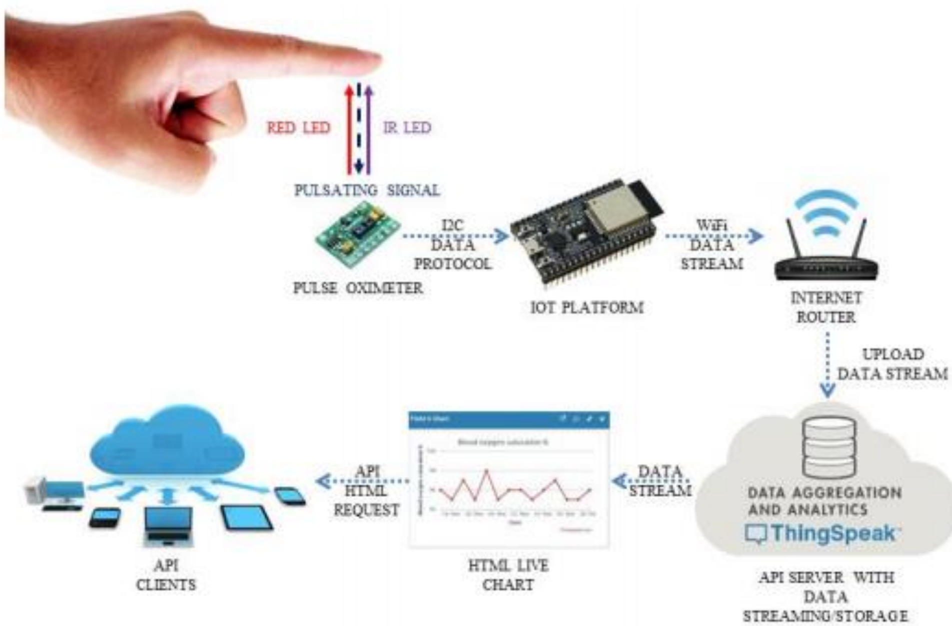

Figure 2. 10 System Architecture (Miron-Alexe, 2020)

2.7.4 Gaps in related systems

The reviewed systems above have several gaps that this study aims to fill. First, they have been developed to only monitor COVID-19 vital symptoms such as oxygen saturation level and heart rate without any added functionality. Secondly, other key COVID-19 symptoms such as fever, fatigue, dry cough, headache, shortness of breath, sore throat, not employed in these systems need also to be monitored. Lastly, external datasets to develop machine learning models which enhances prediction and monitoring of COVID-19 from these vital symptoms have not been used. The research focuses on monitoring various symptoms for early detection of COVID-19 which the reviewed systems have not employed.

2.8 Conceptual Framework

Figure 2.11 shows the conceptual framework of the proposed system. The proposed system is a low cost IoT system which will use a pulse oximeter sensor to measure oxygen saturation levels and heart rate in patients. These are tracked by the system for monitoring and early detection of COVID-19 symptoms. The system will employ KNN algorithm to build a prediction model based on various COVID-19 parameters which include readings from the pulse oximeter sensor. Data from the oximeter sensor will go through the model to predict whether users are likely to be COVID-19 positive or not.

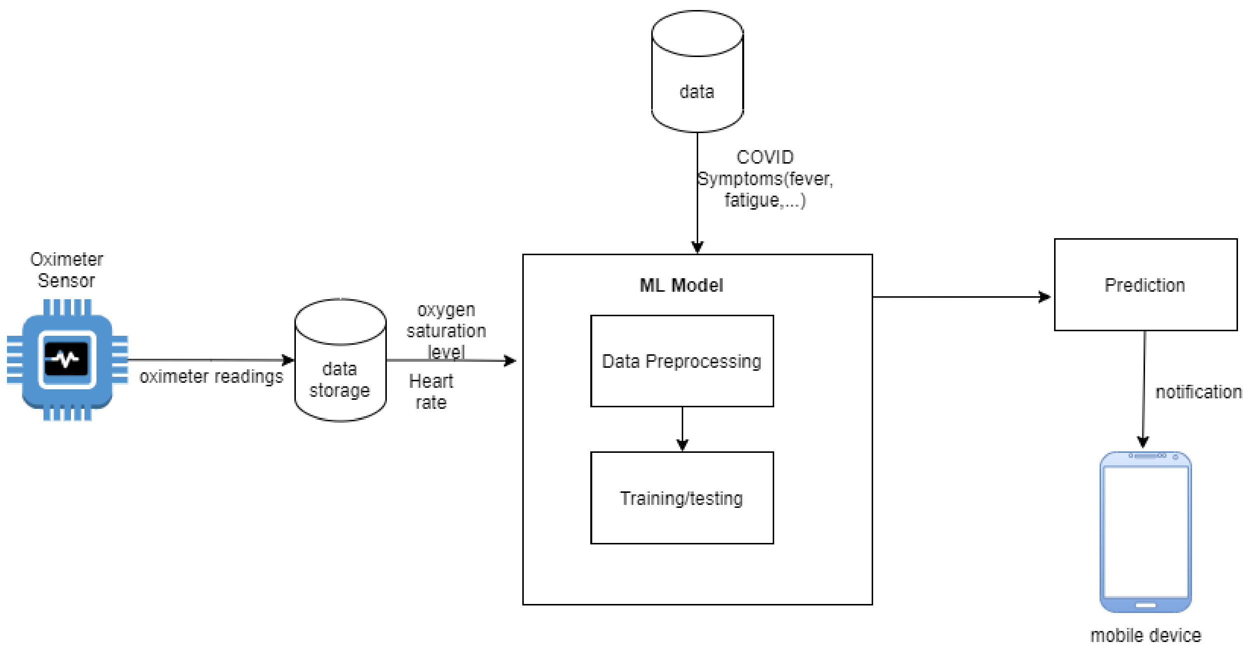

Figure 2. 11 Conceptual Framework of Proposed system

Chapter 3: Methodology

3.1 Introduction

Research methodology allows one to systematically solve a research problem. It is necessary for to know not only the research methods but also the methodology (Kothari, 2004). This chapter elaborates the development methodology that was used to develop the model, the research strategy and approaches to data collection and data analysis.

3.2 System Development Methodology

The proposed solution was developed using Rapid Application development methodology (RAD). RAD is a condensed software development process that produces a high-quality system with low investment costs aimed at reducing the amount of construction needed to build a product (Korkishko, 2017). RAD methodology was suitable for this research because it incorporates special techniques and computer tools to speed up the analysis, design, and implementation phases to get some portion of the system developed quickly (Metcalfe et al., 2021) and into the hands of the users for evaluation and feedback. Figure 3.1 shows the stages of RAD that was used in the implementation of the solution.

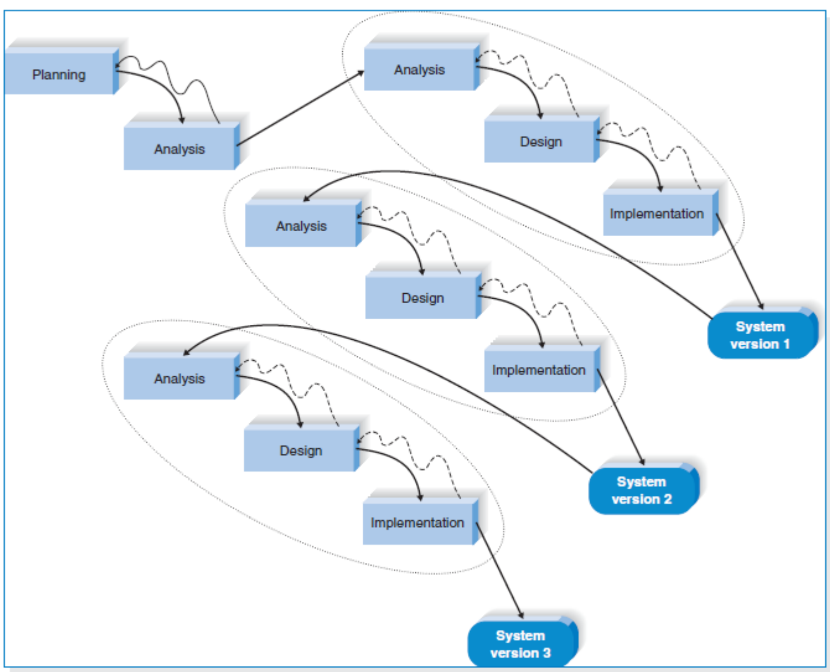

Figure 3 1 RAD Stages

Requirement’s planning was the first phase in which the requirements for the system were defined. Moreover, the systems scope and subject areas for data, which was key for the development of the system, were also defined. This involved gathering of system requirements where tools and techniques were used to collect and define both the functional and nonfunctional requirements of the application.

System Design was the second stage in the study. Here, the system’s data and processes were modelled and architecture to build a working prototype of critical system components designed. It was guided by both the function and non-functional requirements of the application. Moreover, inputs, processes and outputs were identified hence contributing to logical and physical design of the application. Unified Modelling Language (UML) diagrams were designed to depict various components and aspects of the system including use case diagrams, data flow diagrams and sequence diagrams.

The third stage was the construction stage which completed the construction of the physical design of the application system. After completion of the detailed design of the proposed system, the application was built.

The final stage was the implementation stage. This stage included final user testing and training, data conversion, and the implementation of the application system. According to IEEE standards, system testing is the investigation conducted to evaluate whether a complete and integrated software system complies with its specified requirements. The system was put in to testing to establish whether the functionality and usability of the application were met. This stage also assisted in identifying any possible errors in the application.

3.3 Research Design

Research design is the conceptual structure within which research is conducted and guides in the collection, measurement, and analysis of data (Lewis, 2015). There are two fundamental research approaches which are qualitative and quantitative approaches. This study employed both quantitative and qualitative research design to improve research rigor to develop the system. It involved the identification of research objectives, building of the prediction model and validation of the system using several tests to ensure the best performance. Qualitative approach was used to evaluate various parameters from the data used.

3.4 Data Collection

This research employed publicly available data from the Khorshid COVID Cohort (KCC) study containing 50 different COVID-19 epidemiological parameters (Sami et al., 2020). The study contains data of patients admitted for COVID-19 from February 2020 until September 2020 in the Khorshid Hospital in Isfahan. This was used to help analyze epidemiological outbreak information of COVID-19 infected patients. It explored signs and symptoms findings in patients with COVID-19 pneumonia for temporal changes and establish the incidence of psychological disorders and related prevalent symptoms after discharge from the hospital.

These signs and symptoms comprised of, runny nose, fever, fatigue, dry cough, headache, shortness of breath, diarrhea, nausea, sore throat, vomiting, abdominal pain, vital symptoms including pulse rate, respiratory rate, SpO2, systolic blood pressure among others. The data was pre-processed to remove missing data and key features were identified. The dataset was then divided into training and testing data set after which the training dataset was labelled. Training data set consisted of 80% data to train the algorithm effectively and for prediction accuracy and the testing data consisted of 20% data to test the prediction accuracy of the algorithm.

	Gender\n1: male\n2: female	Age\nyears	Occupation\n1: employed\n2: unemployed	Marital status\n1: single\n2: married	Number of family \nwho infected	Symptom duration\ndays	Smoking status\n1: Smoker\n2: Non-smoker	Travel history to \nthe high-risk regions\n1: yes\n2: no	Quarantine before admission \n1: yes\n2: no	Hypertension\n1: yes\n2: no	...	AST\nU/L	ALT\nU/L
0	1	57	1	2	2	10	2	2	2	2	...	38	15
1	2	73	1	2	0	4	2	2	1	1	...	19	14
2	2	75	1	2	0	20	2	2	2	1	...	40	21
3	1	62	2	2	0	4	2	2	1	2	...	69	39
4	1	64	2	2	0	3	2	2	1	1	...	40	26

5 rows × 50 columns

Figure 3 2 Sample data in the dataset (Sami et al., 2020)

Pulse Oximetry data

The pulse oximeter is a device for non-invasive, continuous measurement of oxygen saturation within arterial blood. Pulse oximeters use light-emitting diodes (LEDs) to emit two different wavelengths of light, usually to compare the absorption spectra of oxyhaemoglobin (HbO2) and deoxyhaemoglobin (HHb)(Shalannanda et al., 2020). The MAX30100 Pulse Oximeter is an integrated heart-rate monitor and pulse oximetry sensor solution includes internal optical elements, such as light-emitting diodes, photodetectors (Shalannanda et al., 2020). It combines two LED's, a photodetector, optimized optics, and low-noise analog signal processing to detect pulse and heart-rate signals. The MAX30100 sensor was used to collect patient blood oxygen level and heart rate.

3.5 Data Analysis

Building the model comprised of four phases. First, the preprocessing phase where the data was cleaned, and missing data were removed. Secondly, feature extraction where the informative features were extracted for training. Thirdly, the training phase where the supervised learning algorithm, K-nearest neighbor, were used to train the model and finally, the testing phase where the trained model was tested, and its performance evaluated. These processes have been summarized in Figure 3.3.

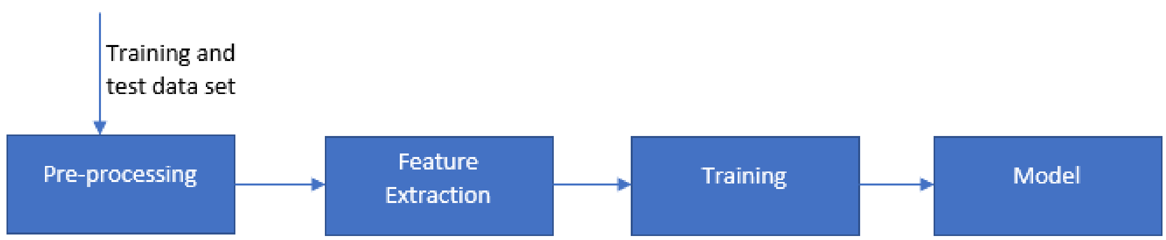

Figure 3 3 Data Analysis for the prediction model

3.6 Research Quality

Research quality is a measure to which the results of a study exhibit correctness and the extent which the researcher ensure that research quality is upheld. The research model was validated to ensure objectivity, quality, and reliability of results. Validity of the data used in this research was ensured by acquiring only data from credible sources to build the model. Reliability is the extent to which a research instrument consistently has the same results if it is used in the same situation on repeated occasions (Heale & Twycross, 2015). To ensure reliability of the model, adequate data was used for training and testing the model. This data was preprocessed to ensure that they are in the same format. The same methodology was used for the preprocessed data to increase accuracy of the model. To ensure objectivity, the study was based on the data to be collected and analyzed and not based on personal biases. In addition, the tools used in data collection, were developed with the objectives in mind.

3.7 Ethical Considerations

Systems must be built from the ground up to be ethical (Burton *et al.*, 2017). This helps to address risk and responsibilities of designing and deploying them. The study considered and addressed ethical issues regarding privacy and security. The system to be developed ensured anonymity of individuals from the data set used which had no personal data of patients. Moreover, to address security concerns, the system did not provide any contact tracing information.

Chapter 4: System Design and Architecture

4.1 Introduction

This section presents the system architecture, key functional and non-functional requirements of the systems. The system architecture and design illustrate the interconnection between the system components and is guided by the functional requirements specified. Unified Modeling Language diagrams are also presented in this section.

4.2 System Analysis

System analysis aids the research in decomposing the pulse oximetry IoT system used in detection of COVID-19 into its various components. It will also help in identifying problems during the design of the system.

4.2.1 Requirements Gathering

Gathering requirements involves collection of the requirements of the system that will be used for prediction. Any existing project requirements must be thought of thoroughly to be able to capture all the needs of key stakeholders of the system. These requirements comprise of functional and non-functional requirements.

4.2.2 Functional Requirements

Functional requirements are the desired functionality of the system to be built and describe the interaction between the system and its environment. Functional requirements generally describe what the system must do. These were obtained through document analysis by identifying in existing literature determinants of COVID-19 diagnosis, key symptoms indicative of a COVID-19 infection and existing techniques developed for diagnosis of the disease. The functional requirements of the system include:

1. The system should make accurate prediction of whether a patient is likely COVID-19 positive or not based on patient symptoms data fed into the system.
2. The oximeter sensor should capture and record blood saturation level and heart rate from patients.
3. The system should allow patients to monitor their blood saturation levels and heart rate.
4. The system should allow patients to login to their account and enter their health information.

5. The system should allow patients to view theirs results.
6. The system should allow administrators to add users to the system.
7. The system should allow administrators to monitor system health.

4.2.3 Non-functional Requirements

Non-functional requirements are nonfunctional characteristics that address operational and technical requirements of the system. The non-functional requirement of the system includes:

1. **Responsiveness** - the proposed system should have a maximum response time from the time of the prediction by the model to the time the results are produced. The process for capturing and analysis of blood saturation level should be reasonably fast for timely detection.
2. **Accuracy** - the prediction results of whether a patient is COVID-19 positive or not should meet certain levels of accuracy. With accuracy of results in the system, there will be objective decision making about hospital staff recommending further diagnostic test to the patience for timely treatment and quarantine.
3. **Flexibility** - the system should be flexible to allow for modifications. It should enable easier finding and fixing of errors to enable the system to perform its functions as required.
4. **Usability** - the system should allow users to easily operate and learn its functions. The users should easily be able provide the system with input and interpret the output given by the system.
5. **Reliability** - the system should perform its specified functions consistently and without failure at any point in time.

4.3 System Architecture

System architecture provides an understanding of the proposed system's design, structure and user requirements that must be supported or achieved by the system. The components in the proposed system architecture include users, a database, a mobile app frontend, and backend that has the prediction model. The system will accept blood parameters oxygen level from patient using the pulse oximeter sensor. Afterwards, these will be analyzed by the prediction model to give a prediction on whether a patient is COVID-19 positive or negative. Patients can view their results from the system after every test. Figure 4.1 shows the visual representation of the system, how its components will interact and tasks that can be carried out in the system.

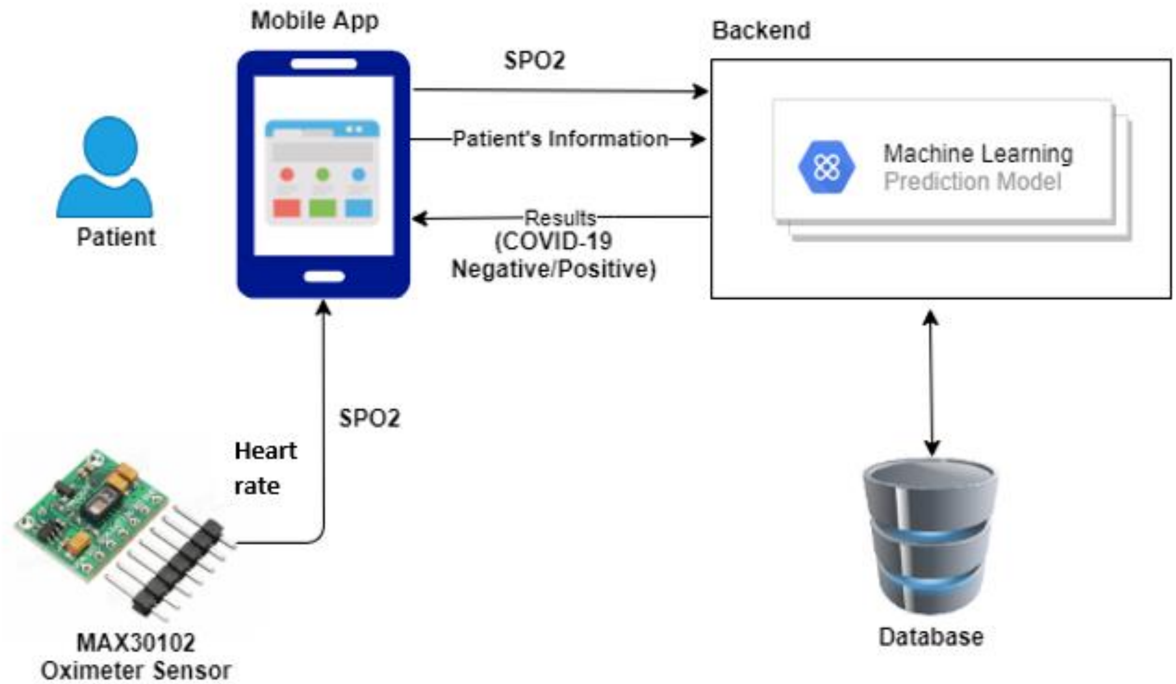

Figure 4. 1 Architecture of proposed system

4.4 Use case diagram

Use Case Diagrams are abstract representations of the system prepared in the requirement analysis phase while finalizing the functional scope of the system (Sahoo and Mohanty, 2017). They are used to summarize the details of the systems actors through Unified modeling Language. Figure 4.2 illustrates the use case diagram that is made up of various actors that interact with the system. A patient is the primary actor of the system as they operate the system to meet its goals. They register themselves to the system, use the sensor to get oximeter readings, key in their health information for analysis by the prediction model and view the results. The oximeter sensor gets readings which include heart rate and blood oxygen levels from the patients.

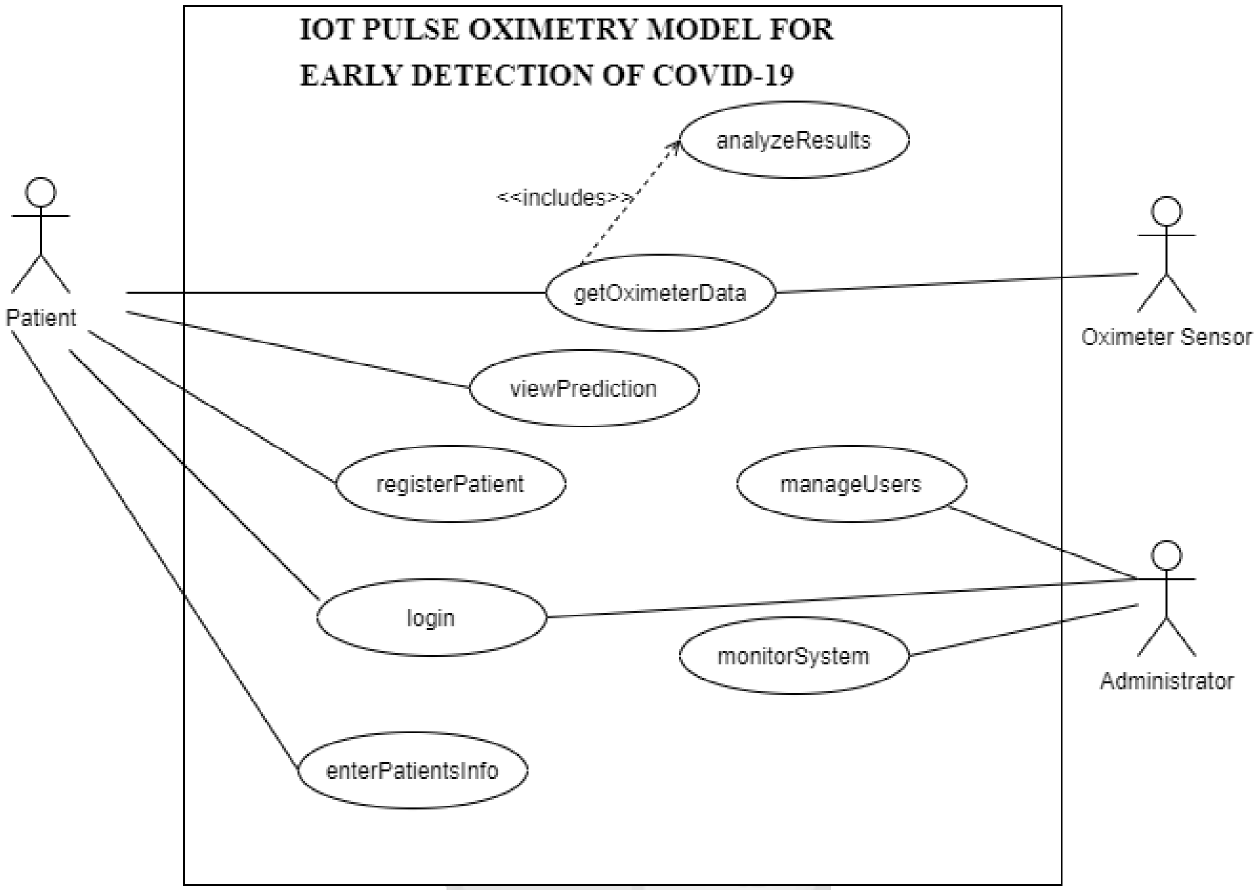

Figure 4. 2 Use Case Diagram of the system

Use Case Descriptions:

This section covers major use cases in the system summarized as system contracts in tabular form.

Use Case Name	Get oximeter data
Description	The pulse oximeter sensor captures SPO2 and heart rate from the patient.
Primary Actor	Oximeter sensor
Trigger	Patient finger placed on the sensor
Pre-condition	The Sensors is connected to a powered Node microcontroller unit.

	Code is uploaded to the Node MCU. IoT module connected to the database
Post-condition	Data recording in the database

Table 4. 1 Get oximeter data use case description.

Use Case Name	Analyze results
Description	Patient data is analyzed by the prediction model of the system
Primary Actor	Patient
Trigger	Patient clicks on analyze on the mobile application
Pre-condition	All patient data keyed in Sensor data obtained from database
Post-condition	Results from the analysis show on the mobile application

Table 4. 2 Analyze results use case description.

Use Case Name	Enter patient information
Description	The patient uses the mobile application to record their symptoms
Primary Actor	Patient
Trigger	Need for analysis of patient symptoms.
Pre-condition	Patient must register and login to the system
Post-condition	Patient data stored in the system database

Table 4. 3 Enter patient information use case description.

Use Case Name	View prediction
Description	The patient views prediction from analysis
Primary Actor	Patient
Trigger	Analysis on patient symptoms performed
Pre-condition	Patient data recorded are analyzed

Post-condition	Patient results from analysis by the model is show on the screen

Table 4. 4 View prediction use case description.

4.5 Sequence diagram

Sequence diagrams show the progression of events and related processes from one system unit another. The patient first logs in to the system by keying in their login details and upon successful login they are able to access the system. Keying in incorrect login information prevents them from accessing the system. The patient then places their finger on the oximeter sensor to get their blood oxygen level and heart rate. The patient can key in other details of the symptoms they are experiencing in the app. This patient data is analyzed by the prediction model. Finally, the system returns a detailed result with a prediction of whether a patient is COVID-19 positive or not.

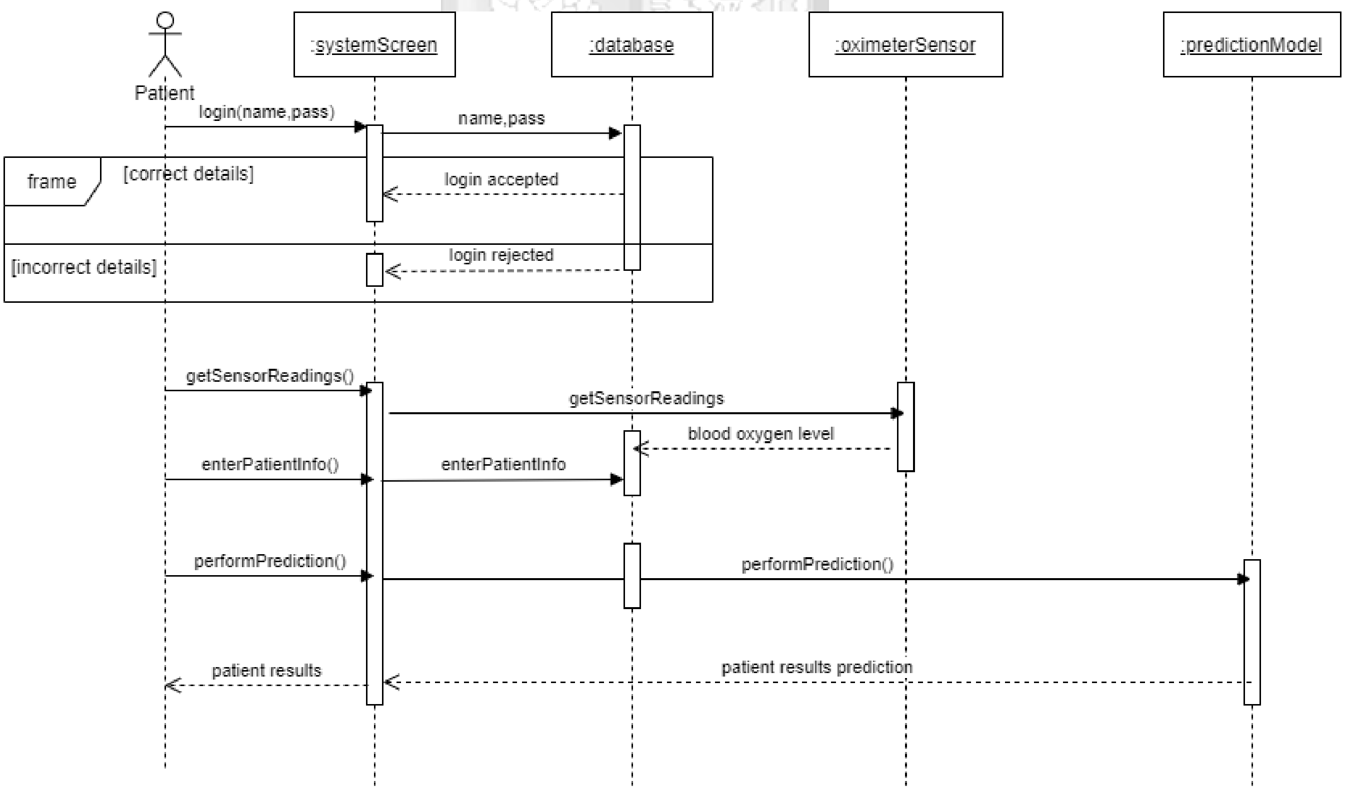

Figure 4. 3 Sequence diagram for proposed system

4.6 ERD

4.6.1 Database Schema

Figure 4.4 shows the Entity Relationship Diagram (ERD) for the system. The database for the system will consist of four tables, oximeter sensor table to record sensor readings, patient table to record patient details, health information table, and result table to record prediction results for a patient.

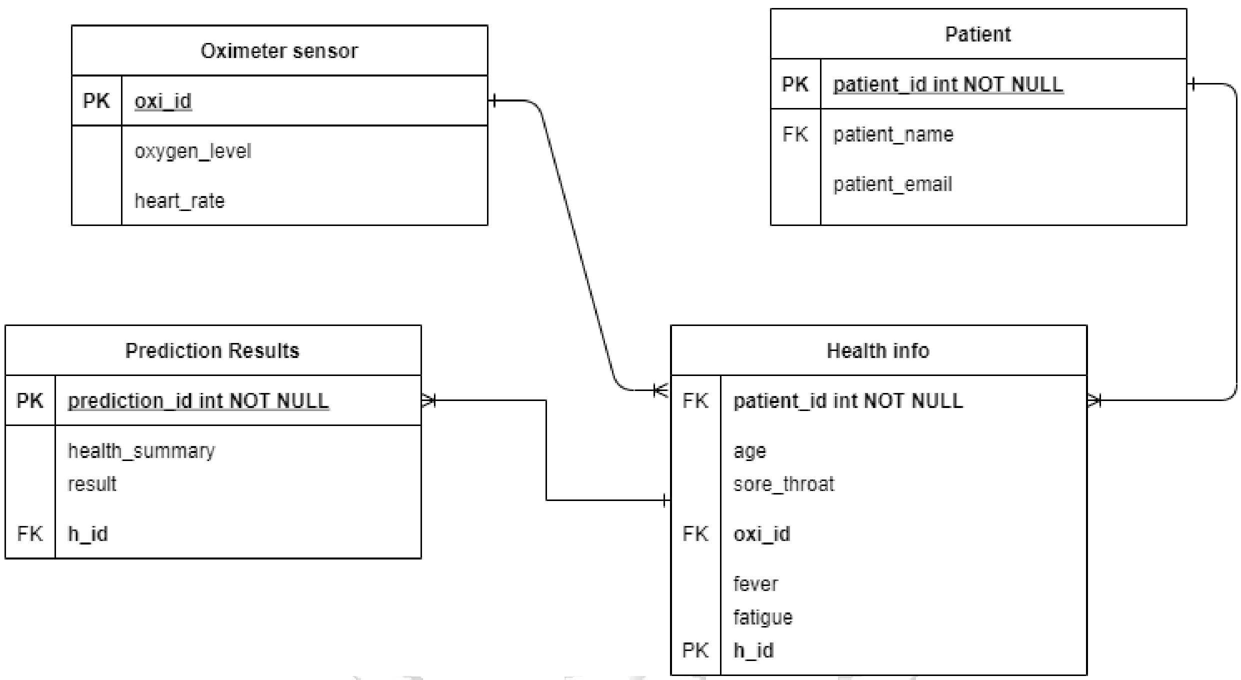

Figure 4. 4 Entity Relationship Diagram for proposed system

4.7 Class Diagram

Classes are the logical building units of Object-Oriented Systems. Class Diagrams are network of identified classes and represent relationships between them(Sahoo & Mohanty, 2017). They describe the structure of the system by showing its classes, attributes, and the relations among the classes. The class diagram illustrated in Figure 4.5 below that shows the blueprints of the COVID-19 detection system. Apart from modelling all the objects that make up the system, the class diagram also displays the interactions between these objects and their functionalities. During system design, a class diagram helps bring out a better understanding of the requirements of the problem domain and in the identification of system's components.

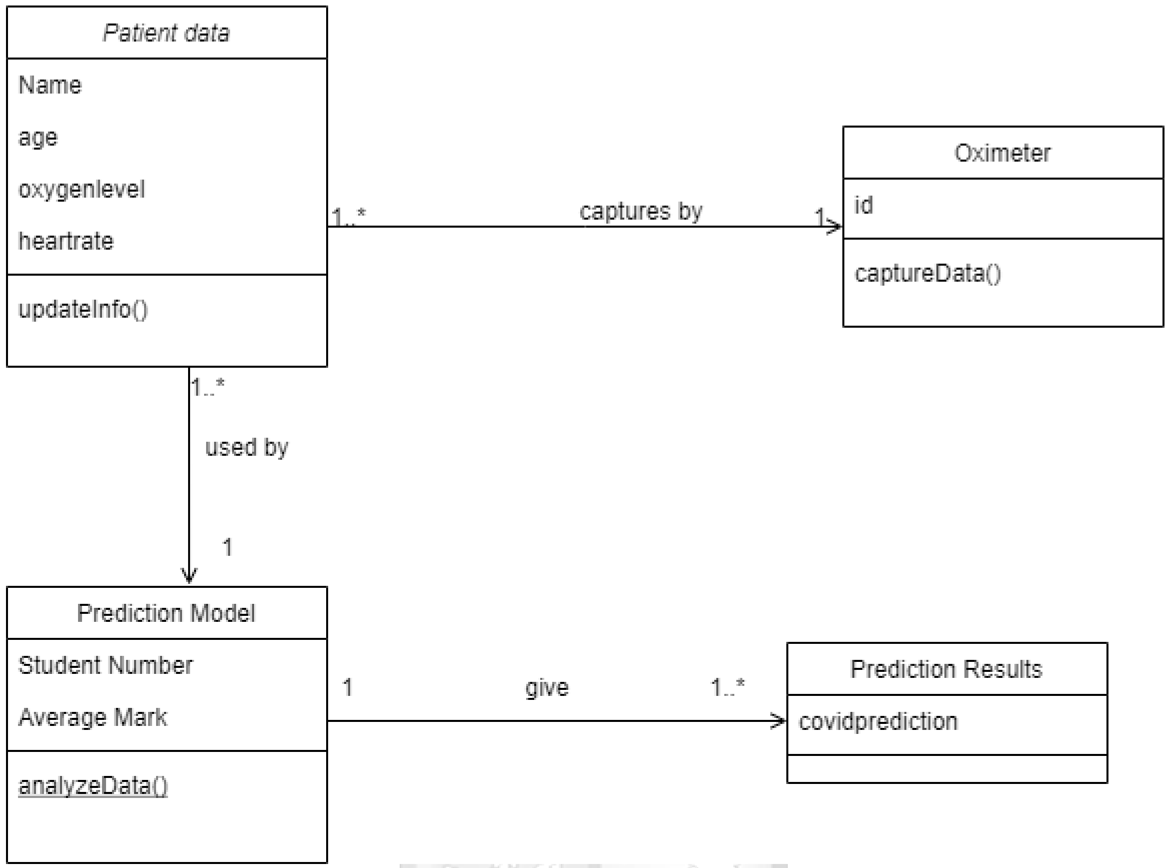

Figure 4. 5 Class Diagram

4.8 Wireframes of the system

The proposed system is a web application. Patients will be required to create accounts and login to their accounts thereafter. The application will enable them to enter their health information, run the prediction model to detect COVID-19 and monitor their heartrate and blood oxygen level. The wireframe of the system is illustrated Figure 4.6 below.

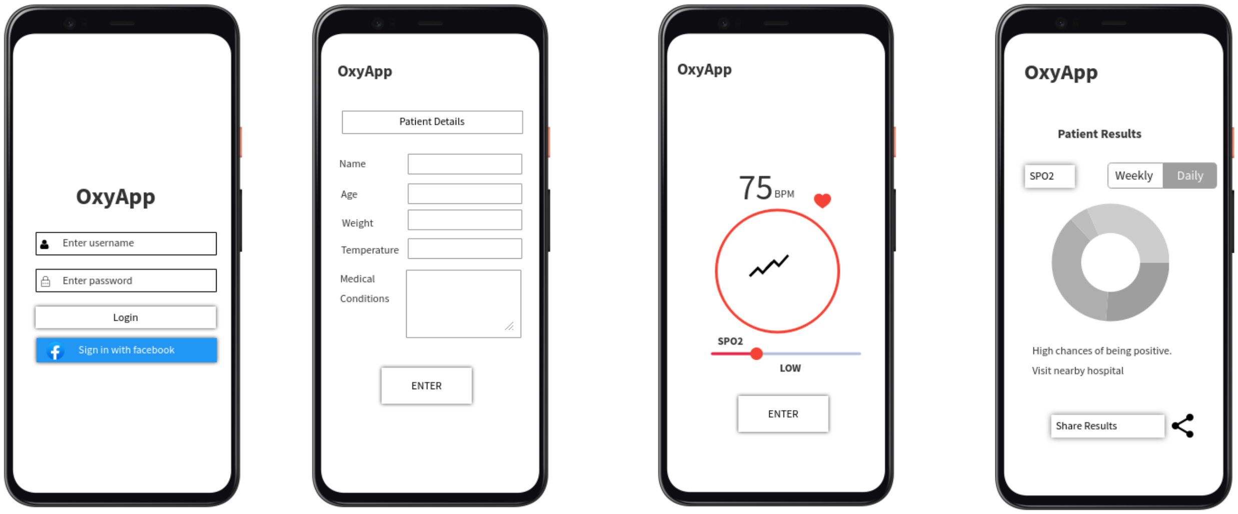

Figure 4. 6 Wireframes of the system

Chapter 5: System Implementation and Testing

5.1 Introduction

This chapter presents the implementation and testing of the system that will be used for early COVID-19 symptoms detection. System implementation and testing incorporate a series of steps that determine the nature of the delivered system.

5.2 Components of the System

A MAX30100 pulse oximeter sensor connected to an ESP8266 Node Microcontroller Unit (MCU) is used to get blood oxygen level and heart rate readings from the patient's finger placed on the sensor. At every instance, blood oxygen level and heart rate measurements are transmitted to the database through the Node MCU module. The KNN algorithm based on COVID-19 symptoms interprets and analyzes the data to determine the likelihood of COVID-19 infection. The prediction results of a particular patient are then automatically viewed on the mobile application by the patient after analysis.

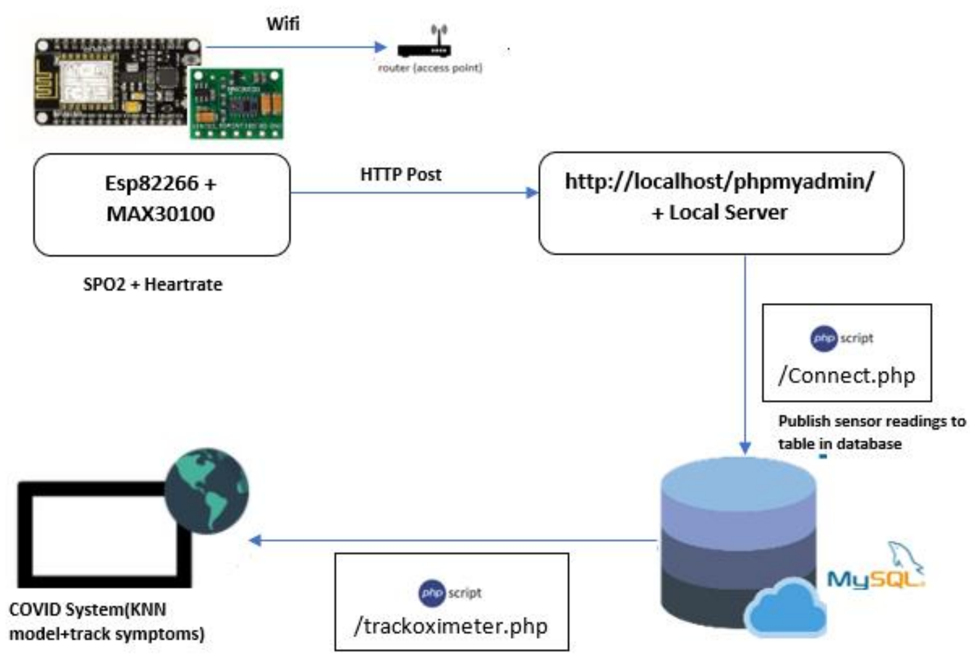

Figure 5. 1 System Components

5.2.1 IoT System Components

The MAX30100 Pulse Oximeter Sensor is an integrated heart-rate monitor and pulse oximetry biosensor solution. It combines two LEDs, a photo detector, optimized optics, and low-noise analog signal processing to detect pulse oximetry and heart-rate signals (Hema et al., n.d.). The MAX30100 operates from 1.8V and 3.3V power supplies. The MAX30100 provides a complete system solution to ease the design-in process for mobile and wearable devices. Communication is through a standard I2C-compatible interface. The module can be shut down through software with zero standby current, allowing the power rails to always remain powered.

Figure 5. 2 MAX30100 Sensor (Hema et al., n.d.)

ESP8266 Node MCU module has integrated WiFi which provides connection to the WiFi, gathering the signals from the sensors (Škraba *et al.*, 2017). The module is best suited for Internet of Things (IoT) at its low cost, low power consumption capability as it requires 3.3V power, built in WiFi module, integrated TCP/IP protocol stack, easy to flash and erase firmware and is usb powered (Škraba et al., 2017).

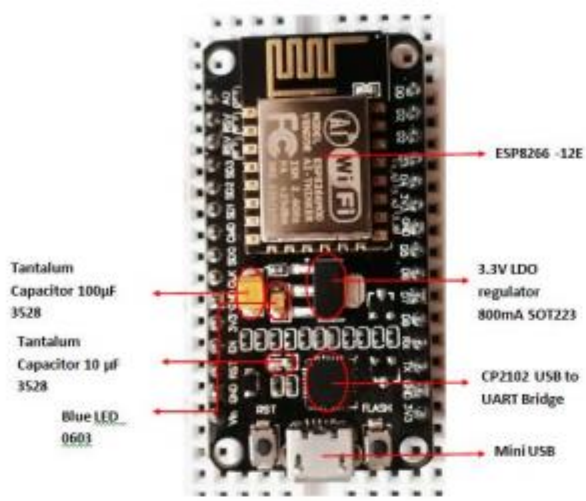

Figure 5. 3 ESP8266 Development Board (Škraba et al., 2017)

The microcontroller used for interfacing the sensor is the Node MCU consisting of an ESP8266 Wi-Fi module (Hema et al., n.d.). This is the controller used for communicating the results to online web server. An OLED Display, Breadboard and Jumper wires(male-female) are used to interconnect different components that make up the system.

5.2.2 Software Requirements

The table describes the software components of the soiled linen detection prototype.

Software	**Description**
MYSQL	Open-source relational database management system.
Arduino IDE	Arduino IDE Version 1.8.13
Imported libraries (Arduino)	MAX30100_PulseOximeter.h ESP8266WiFi.h FirebaseArduino.h Adafruit_GFX.h
Jupyter Notebook	Interactive environment to build the model
Python Libraries	Numpy, Matplotlib, sklearn, pandas, pickle, flask
Visual Studio Code	Development environment for the web application

Table 5. 1 Software Requirements

5.2.3 Machine Learning Model

The dataset consists of the features listed in appendix A and on that dataset various machine learning algorithms were applied to predict whether one has COVID-19 or not. The analysis of decision trees, naïve bayes and KNN performance on the dataset is shown in Table 5.2. The KNN model produced the best results as compared to the other algorithms.

Table 5. 2 ML Algorithms on the dataset.

ML Algorithm	Accuracy score
Naïve Bayes	58.33%
KNN	66.67%
Decision Tree	66.67%

Scikit-learn library, used for implementation of the KNN model is a python library that focuses on bringing machine learning to non-specialist using general-purpose high-level language. To train the model, the excel data file containing the preprocessed and labelled COVID-19 symptoms was read into Pandas Data Frame. The dataset was split on 8:2 ratios for training and testing respectively using the train_test_split method of the scikit-learn library.

```
In [12]: from sklearn.model_selection import train_test_split
         X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.20, random_state = 0)
```

For KNN, Minkowski distance was used with 5 nearest neighbors set. The model was then saved using the pickle library as a .sav file to be used for analysis by the mobile application.

```
In [14]: from sklearn.neighbors import KNeighborsClassifier
         classifier = KNeighborsClassifier(n_neighbors = 5, metric = 'minkowski', p = 2)
         classifier.fit(X_train, y_train)

Out[14]: KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
                              metric_params=None, n_jobs=None, n_neighbors=5, p=2,
                              weights='uniform')
```

The Confusion Matrix is a means to visualize the per class prediction performance of the model. The confusion matrix is of the form (true class, predicted class) where the rows represent true class and columns represented predicted class. We can therefore infer that correctly classified points are grouped corresponding to the classes in the diagonal entries of the confusion matrix.

```
In [36]: #KNN
         from sklearn.neighbors import KNeighborsClassifier
         from sklearn.metrics import confusion_matrix,accuracy_score
         classifier = KNeighborsClassifier(n_neighbors = 5, metric = 'minkowski', p = 2)
         classifier.fit(X_train, y_train)
         y_predknn = classifier.predict(X_test)

         cm = confusion_matrix(y_test, y_predknn)
         ac = accuracy_score(y_test,y_predknn)
         ac

Out[36]: 0.6666666666666666
```

5.3 Implementation of the System

The web application for the system was created using flask, php, HTML and CSS. It is connected to a MYSQL database which stores patient data and sensor readings. The sensor readings from the MAX30100 sensor are viewed on the app as show in Figure 5.4.

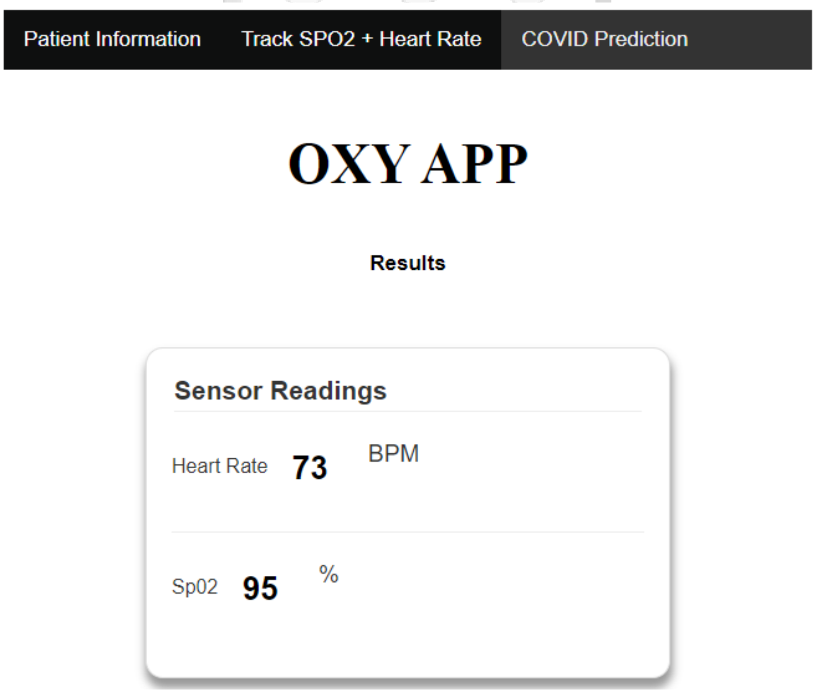

Figure 5. 4 Web application

To begin integration of the system, the development environment was setup. The Arduino IDE (Integrated Development Environment) development tool was used to write the code to be uploaded to the ESP8266 node MCU. The IoT components were connected as show in Figure 5.5 below

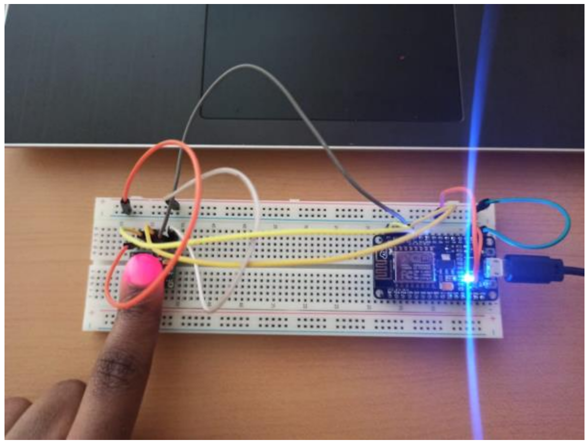

Figure 5. 5 Circuit connection of the system

As in the circuit diagram, ESP8266 Node MCU is USB powered. It picks data from the MAX30100 pulse oximeter. The readings captured are as shown in Figure 5.6 below on the serial monitor of Arduino ide.

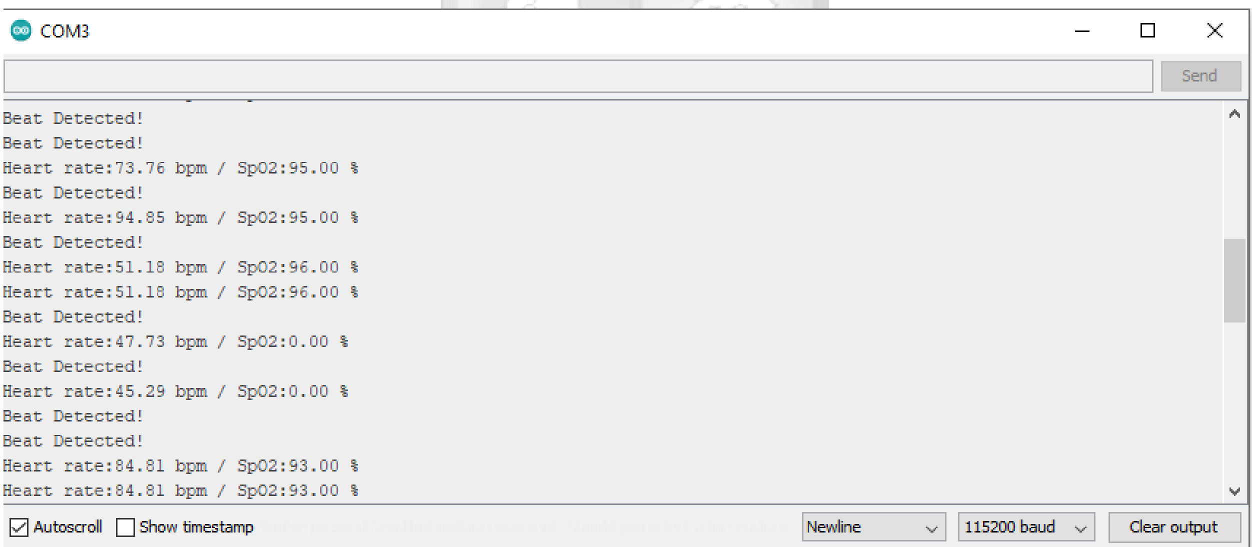

Figure 5. 6 MAX30100 sensor readings

The data received from the ESP8266 Node MCU is sent to the MYSQL database using the WIFI connection and obtained by the mobile application. The information is then displayed on the web application. The app analyzes other COVID-19 symptoms together with the sensor readings to predict likelihood of infection using the prediction model.

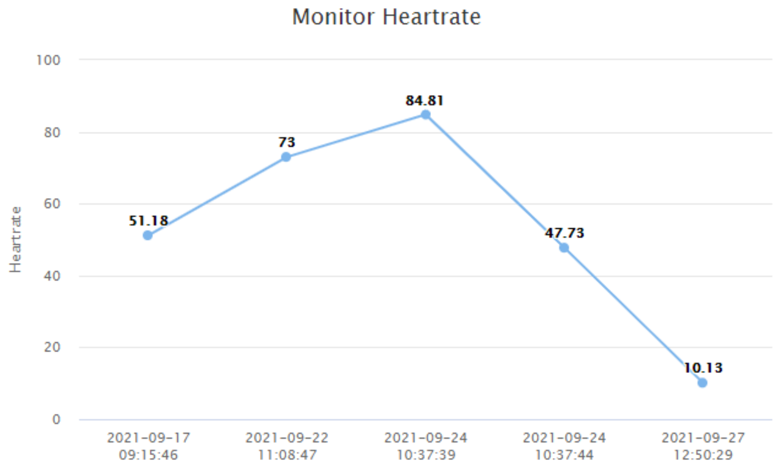

Figure 5. 7 Heartrate Readings

5.4 Testing

After integrating all system components, the system was tested to verify it is working as required. Functionality testing was carried iteratively during the development process against. This was to ensure that requirements were met, and that the system executed perfectly. The testing process also ensured that all bugs were fixed. The test cases conducted for testing are show in the table below.

Test case	Pre-condition	Test Data	Expected Results	Pass/Fail
Check response on entering valid email and password	Mobile app must be installed	Email: lesleybonyo@gmail.com Password: 12345678	Login must be successful	Pass
Confirm ESP2866 sends sensor readings to database	IoT module must be connected to a computer using a USB cable	Heart rate SPO2	Readings uploaded on the firebase database	Pass
Confirm that	MAX30100	Heart rate	Readings	Pass

the mobile application displays the sensor data in real time	uploaded to the database	SPO2	displayed on the mobile app	
Confirm analysis of patient’s symptoms by the prediction model	Patient selects symptoms. Heart rate and SPO2 captured	Fever, heart rate, SPO2, headache, dry cough	Prediction results by the model	Pass

Table 5. 3 Test Cases

Chapter Six: Discussion

6.1 Introduction

This chapter reviews the objectives of the research specified in chapter one based on the results obtained from the system developed. The aim of the study was to ensure early detection of COVID-19 in individuals by developing an IoT pulse oximetry system to monitor oxygen saturation levels and heart rate. A model was developed using KNN algorithm to perform predictions based on patient symptoms.

6.2 Determinants of COVID-19 diagnosis

The first objective of this research was to analyze the determinants of COVID-19 diagnosis. The research study established that COVID-19 diagnosis is mainly based on nasal swabs tested using RT-PCR. Other diagnosis techniques are based on antibodies, X-RAYs and CT scans reviewed in chapter 2 of the research. Additionally, it was determined that COVID-19 symptoms monitoring is as essential in the diagnosis process because a significant number of COVID-19 infected patients show no symptoms.

6.3 Techniques used to detect COVID-19 disease

The second objective was to analyze technologies that are used for COVID-19 detection. Several techniques such as voice detection, using computer vision techniques for analysis of X-Rays and CT-Scans and smart devices to monitor COVID-19 symptoms have been developed for the detection of COVID-19. These approaches have significantly provided cost effective and convenient means for testing. However, these innovations have not been approved by the WHO and only molecular and serology tests are being used in hospitals for detection of COVID-19 disease. It is noted therefore that they should only complement the existing diagnosis techniques to counter challenges mentioned in chapter 2 of the study.

6.4 IoT pulse oximetry model for detecting COVID-19 symptoms

Based on the determinants for diagnosis of COVID-19 disease established from the research, the inputs used for the prediction model were fever, fatigue, dry cough, headache, shortness of breath, sore throat, pulse rate and SPO2. The system is made up of an oximeter sensor which captures heart rate and blood oxygen levels. The sensor data is displayed on a mobile app and are analyzed with other COVID-19 symptoms using the KNN model. The mobile application offers a platform for monitoring of these symptoms which can help in early detection of COVID-19 disease.

6.5 System Testing

In this study, the model was developed based on COVID-19 symptoms using KNN. This model was tested using various metrics to ascertain the performance. The results were presented in chapter 5.

Chapter 7: Conclusion and Recommendation

7.1 Conclusion

Several studies have indicated how a significant number of those infected by COVID-19 disease show little to no symptoms. From the study we observe that without early diagnosis, individuals face risks of serious deterioration. Parameters such as heart rate and oxygen saturation can not only assist in early COVID-19 detection but also help to identify patients at risk from pulmonary complications. A great number of COVID-19 patients have been found to be suffering from pneumonia from analysis of their chest x-rays. This is coupled with low oxygen levels ranging from 50% to 80% that is hard to detect because patients do not show any noticeable breathing difficulties. When COVID-19 patients discover they are short of breath, their conditions have already worsened from moderate-to-severe levels of pneumonia. Therefore, patients with COVID-19 who do not need immediate hospital attention but are at high risk of developing serious symptoms ought to be monitored to reduce risk. The research was able to employ a KNN model based which was able predict COVID-19 infection based on these symptoms with an accuracy of 66.67 percent. A key challenge encountered was inadequate data to improve accuracy of the model. The system was able to classify individuals based on the SPO2 and heart rate readings from the oximeter sensor. With the low cost IoT base oximeter system, patients can be able to monitor these crucial COVID-19 symptoms.

7.2 Recommendation

From the study outcome, to guarantee proper use of the system users measure their SPO2 and heart rate several times a day. This enables constant monitoring of the patient's health status. Moreover, rather than accepting the first number that appears on the screen, the individual should observe the readings for a few seconds to identify the most measured value and should only accept values associated with a strong pulse signal.

7.3 Future Works

The oximeter based COVID-19 detection system developed employs the use of a medical sensor to remotely capture patient data. This can be integrated into a smart device that is portable such as a smart watch to conveniently monitor COVID-19 symptoms. Moreover, MAX30100 pulse oximeter sensor captures only two metrics, the oxygen saturation level and heart rate, for identifying COVID-19. Other sensors to capture symptoms such as body temperature can be incorporated for more accurate detection of COVID-19 disease.

References

Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., Tao, Q., Sun, Z., & Xia, L. (2020). Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology*, *296*(2), E32–E40. https://doi.org/10.1148/radiol.2020200642

Akram, T., Attique, M., Gul, S., Shahzad, A., Altaf, M., Naqvi, S. S. R., Damaševičius, R., & Maskeliūnas, R. (2021). A novel framework for rapid diagnosis of COVID-19 on computed tomography scans. *Pattern Analysis and Applications*, *1*, 3. https://doi.org/10.1007/s10044-020-00950-0

Alves, M. A., Castro, G. Z., Oliveira, B. A. S., Ferreira, L. A., Ramírez, J. A., Silva, R., & Guimarães, F. G. (2021). Explaining machine learning based diagnosis of COVID-19 from routine blood tests with decision trees and criteria graphs. *Computers in Biology and Medicine*, *132*, 104335. https://doi.org/10.1016/j.compbiomed.2021.104335

Amanat, F., Stadlbauer, D., Strohmeier, S., Nguyen, T. H. O., Chromikova, V., McMahon, M., Jiang, K., Arunkumar, G. A., Jurczyszak, D., Polanco, J., Bermudez-Gonzalez, M., Kleiner, G., Aydillo, T., Miorin, L., Fierer, D. S., Lugo, L. A., Kojic, E. M., Stoever, J., Liu, S. T. H., … Krammer, F. (2020). A serological assay to detect SARS-CoV-2 seroconversion in humans. *Nature Medicine*, 1–4. https://doi.org/10.1038/s41591-020-0913-5

Amulya, M. P., Niranjanamurthy, M., Yogish, H. K., & Ravikumar, G. K. (2021). *Tracking and Analysis of Corona Disease Using Intelligent Data Analysis*. 347–364. https://doi.org/10.1007/978-981-16-1574-0_16

Andreswari, R., Ambarsari, N., Syahrina, A., Puspitasari, W., Novianti, A., & Darmawan, I. (2020). *INTERNATIONAL JOURNAL OF INNOVATION IN ENTERPRISE SYSTEM Design of e-Marketplace for Village-owned Small, Micro and Medium Enterprise using Rapid Application Development*. https://ijies.sie.telkomuniversity.ac.id/index.php/IJIES/index

Arpaci, I., Huang, S., Al-Emran, M., Al-Kabi, M. N., & Peng, M. (2021). Predicting the COVID-19 infection with fourteen clinical features using machine learning classification algorithms. *Multimedia Tools and Applications*, 1–15. https://doi.org/10.1007/s11042-020-10340-7

Arslan, H., & Arslan, H. (2021). A new COVID-19 detection method from human genome sequences using CpG island features and KNN classifier. *Engineering Science and Technology, an International Journal*. https://doi.org/10.1016/j.jestch.2020.12.026

Brinati, D., Campagner, A., Ferrari, D., Locatelli, M., Banfi, G., & Cabitza, F. (2020). Detection of COVID-19 Infection from Routine Blood Exams with Machine Learning: A Feasibility Study. *Journal of Medical Systems*, *44*(8), 1–12. https://doi.org/10.1007/s10916-020-01597-4

Buchhorn, R., Baumann, C., & Willaschek, C. (2020). Heart Rate Variability in a Patient with Coronavirus Disease 2019. https://doi.org/10.17987/icfj.v20i0.685

Cabitza, F., Campagner, A., Ferrari, D., di Resta, C., CerIoTti, D., Sabetta, E., Colombini, A., de Vecchi, E., Banfi, G., Locatelli, M., & Carobene, A. (2020). Development, evaluation, and validation of machine learning models for COVID-19 detection based on routine blood tests. *Clinical Chemistry and Laboratory Medicine*, *59*(2), 421–431. https://doi.org/10.1515/cclm-2020-1294

Chamola, V., Hassija, V., Gupta, V., & Guizani, M. (2020). A Comprehensive Review of the COVID-19 Pandemic and the Role of IoT, Drones, AI, Blockchain, and 5G in Managing its Impact. *IEEE Access*, *8*, 90225–90265. https://doi.org/10.1109/ACCESS.2020.2992341

Chen, J., Li, K., Zhang, Z., Li, K., & Yu, P. S. (2020). *A Survey on Applications of Artificial Intelligence in Fighting Against COVID-19*.

Erdem, Ö., Derin, E., Sagdic, K., Yilmaz, E. G., & Inci, F. (2021a). Smart materials-integrated sensor technologies for COVID-19 diagnosis. *Emergent Materials*, 1–17. https://doi.org/10.1007/s42247-020-00150-w

Erdem, Ö., Derin, E., Sagdic, K., Yilmaz, E. G., & Inci, F. (2021b). Smart materials-integrated sensor technologies for COVID-19 diagnosis. *Emergent Materials*, 1–17. https://doi.org/10.1007/s42247-020-00150-w

Fong, S. J., Dey, N., & Chaki, J. (2020). *AI-Enabled Technologies that Fight the Coronavirus Outbreak* (pp. 23–45). Springer, Singapore. https://doi.org/10.1007/978-981-15-5936-5_2

Graves, J., Byerly, J., Priego, E., Makkapati, N., Parish, S., Medellin, B., & Berrondo, M. (2020). A Review of Deep Learning Methods for Antibodies. Antibodies, 9(2), 12. https://doi.org/10.3390/antib9020012

Hema, L. K., Mohana Priya, R., Indumathi, R., & Ii, G. (n.d.). DESIGN AND DEVELOPMENT OF IOT BASED PULSE OXIMETER. Retrieved May 3, 2021, from http://www.acadpubl.eu/hub/

Hidayat, A., Subono, Wardhany, V. A., Nugroho, A. S., Hakim, S., Jhoswanda, M., Syamsiana, I. N., & Agustina, N. A. (2020). Designing IoT-Based Independent Pulse Oximetry Kit as an Early Detection Tool for Covid-19 Symptoms. *2020 3rd International Conference on Computer and Informatics Engineering, IC2IE 2020*, 443–448. https://doi.org/10.1109/IC2IE50715.2020.9274663

Jeong, H., Rogers, J. A., & Xu, S. (2020). Continuous on-body sensing for the COVID-19 pandemic: Gaps and opportunities. In *Science Advances* (Vol. 6, Issue 36, p. eabd4794). American Association for the Advancement of Science. https://doi.org/10.1126/sciadv.abd4794

Kadry, S., Rajinikanth, V., Rho, S., Sri, N., Raja, M., Rao, V. S., & Thanaraj, K. P. (n.d.). Development of a Machine-Learning System to Classify Lung CT Scan Images into Normal/COVID-19 Class.

Kang, H., Xia, L., Yan, F., , Zhibin Wan, Feng Shi, Huan Yuan, Huiting Jiang, D. W., & He Sui, C. Z. and D. S. (2020). *Diagnosis of Coronavirus Disease 2019 (COVID-19) with Structured Latent Multi-View Representation Learning*. https://ieeexplore.ieee.org.ezproxy.library.strathmore.edu/stamp/stamp.jsp?tp=&arnumber=9086482

Kumar, K., Kumar, N., & Shah, R. (2020). Role of IoT to avoid spreading of COVID-19. *International Journal of Intelligent Networks*, *1*, 32–35. https://doi.org/10.1016/j.ijin.2020.05.002

Lerner, A. M., Eisinger, R. W., Lowy, D. R., Petersen, L. R., Humes, R., Hepburn, M., & Cassetti, M. C. (2020). The COVID-19 Serology Studies Workshop: Recommendations and Challenges. *Immunity*. https://doi.org/10.1016/j.immuni.2020.06.012

Levitan, R. M. (2020). Pulse Oximetry as a Biomarker for Early Identification and Hospitalization of COVID-19 Pneumonia. In Academic Emergency Medicine (Vol. 27, Issue 8, pp. 785–786). Blackwell Publishing Inc. https://doi.org/10.1111/acem.14052

Linda J., C., Linda V., G., Jeffrey W., S., Yingzhu, L., Qiongqiong, Z., Catherine J., S., Janet M., S., Anne C., G., Divya J., S., Tiffany R., B., And, S. R. J., & Liu, C. (2020). *Assay Techniques and Test Development for COVID-19 Diagnosis*. https://pubs.acs.org/doi/pdf/10.1021/acscentsci.0c00501

Loey, M., Manogaran, G., Hamed, M., Taha, N., Eldeen, N., & Khalifa, M. (2020). *Sustainable Cities and Society xxx (xxxx) xxx Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection*. https://doi.org/10.1016/j.scs.2020.102600

Mansour, N. A., Saleh, A. I., Badawy, M., & Ali, H. A. (2021). Accurate detection of Covid-19 patients based on Feature Correlated Naïve Bayes (FCNB) classification strategy. Journal of Ambient Intelligence and Humanized Computing, 1, 3. https://doi.org/10.1007/s12652-020-02883-2

Metcalfe, B., Iravani, P., Graham-Harper-Cater, J., Bowman, R., Stirling, J., & Wilson, P. (2021). A Cost-Effective Pulse Oximeter Designed in Response to the COVID-19 Pandemic. *Journal of Open Hardware*, *5*(1). https://doi.org/10.5334/joh.26

Miron-Alexe, V. (2020). IoT PULSE OXIMETRY STATUS MONITORING FOR HOME QUARANTINED COVID-19 PATIENTS. *Journal of Science and Arts Year*, *20*(4), 1029–1036. https://doi.org/10.46939/J.Sci.Arts-20.4-c04

Mishra, T., Wang, M., Metwally, A. A., Bogu, G. K., Brooks, A. W., Bahmani, A., Alavi, A., Celli, A., Higgs, E., Dagan-Rosenfeld, O., Fay, B., Kirkpatrick, S., Kellogg, R., Gibson, M., Wang, T., Hunting, E. M., Mamic, P., Ganz, A. B., Rolnik, B., … Snyder, M. P. (2020). Pre-symptomatic detection of COVID-19 from smartwatch data. *Nature Biomedical Engineering*, *4*(12), 1208–1220. https://doi.org/10.1038/s41551-020-00640-6

Motley, M. P., Bennett-Guerrero, E., Fries, B. C., & Spitzer, E. D. (2020). Review of Viral Testing (Polymerase Chain Reaction) and Antibody/Serology Testing for Severe Acute Respiratory Syndrome-Coronavirus-2 for the Intensivist. *Critical Care Explorations*, *2*(6), e0154. https://doi.org/10.1097/CCE.0000000000000154

Narzullaev, A., Muminov, Z., & Narzullaev, M. (2020). Contact Tracing of Infectious Diseases Using Wi-Fi Signals and Machine Learning Classification. *2020 IEEE 2nd International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, 1–5. https://doi.org/10.1109/IICAIET49801.2020.9257812

Ramirez Lopez, L. J., Aponte, G. P., & Garcia, A. R. (2019). Internet of things applied in healthcare based on open hardware with low-energy consumption. *Healthcare Informatics Research*, *25*(3), 230–235. https://doi.org/10.4258/hir.2019.25.3.230

Renganathan, D. Y. E. Y. V., & Uner, K. A. R. U. ¨ O. ¨. (2020). *A Vision-based Social Distancing and Critical Density Detection System for COVID-19*. https://www.researchgate.net/profile/Ekim_Yurtsever/publication/342763051_A_Vision-based_Social_Distancing_and_Critical_Density_Detection_System_for_COVID-19/links/5f08fbdc92851c52d628c8eb/A-Vision-based-Social-Distancing-and-Critical-Density-Detection-Sy

Sahoo, P., & Mohanty, J. R. (2017). Early Test Effort Prediction using UML Diagrams. Indonesian Journal of Electrical Engineering and Computer Science, 5(1), 220–228. https://doi.org/10.11591/ijeecs.v5.i1.pp

Sami, R., Soltaninejad, F., Amra, B., Naderi, Z., Haghjooy Javanmard, S., Iraj, B., Haji Ahmadi, S., Shayganfar, A., Dehghan, M., Khademi, N., Sadat Hosseini, N., Mortazavi, M., Mansourian, M., Mañanas, M. A., Marateb, H. R., & Adibi, P. (2020).

A one-year hospital-based prospective COVID-19 open-cohort in the Eastern Mediterranean region: The Khorshid COVID Cohort (KCC) study. *PLOS ONE*, *15*(11), e0241537. https://doi.org/10.1371/journal.pone.0241537

Shalannanda, W., Zakia, I., Fahmi, F., & Sutanto, E. (2020, November 4). Implementation of the hardware module of IoT-based infant incubator monitoring system. *Proceeding of 14th International Conference on Telecommunication Systems, Services, and Applications, TSSA 2020*. https://doi.org/10.1109/TSSA51342.2020.9310901

Shenoy, N., Luchtel, R., & Gulani, P. (2020). Considerations for target oxygen saturation in COVID-19 patients: Are we under-shooting? *BMC Medicine*, *18*(1), 260. https://doi.org/10.1186/s12916-020-01735-2

Shorten, C., Khoshgoftaar, T. M., & Furht, B. (2021). Deep Learning applications for COVID-19. *Journal of Big Data*, *8*(1), 1–54. https://doi.org/10.1186/s40537-020-00392-9

Singhal, T. (2020). A Review of Coronavirus Disease-2019 (COVID-19). In *Indian Journal of Pediatrics* (Vol. 87, Issue 4, pp. 281–286). Springer. https://doi.org/10.1007/s12098-020-03263-6

Škraba, A., Koložvari, A., Kofjač, D., Stojanović, R., Stanovov, V., & Semenkin, E. (2017, July 12). Prototype of group heart rate monitoring with NODEMCU ESP8266. 2017 6th Mediterranean Conference on Embedded Computing, MECO 2017 - Including ECYPS 2017, Proceedings. https://doi.org/10.1109/MECO.2017.7977151

Sohail, S. (2020). Rational and practical use of imaging in COVID-19 pneumonia. *Pakistan Journal of Medical Sciences*, *36*(COVID19-S4), S130–S133. https://doi.org/10.12669/pjms.36.COVID19-S4.2760

Teo, J. (2020). Early Detection of Silent Hypoxia in Covid-19 Pneumonia Using Smartphone Pulse Oximetry. In Journal of Medical Systems (Vol. 44, Issue 8, pp. 1–2). Springer. https://doi.org/10.1007/s10916-020-01587-6

Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. BMC Medical

Informatics and Decision Making, 19(1), 281. https://doi.org/10.1186/s12911-019-1004-8

WHO. (2020a). *Coronavirus disease (COVID-19)*. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200626-covid-19-sitrep-158.pdf?sfvrsn=1d1aae8a_2

WHO. (2020b). *Situation Report-51 SITUATION IN NUMBERS total and new cases in last 24 hours*. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf?sfvrsn=1ba62e57_10

World Health Organization. (2020). COVID-19 Weekly Epidemiological Update 22. *World Health Organization*, *December*, 1–3. https://www.who.int/docs/default-source/coronaviruse/situation-reports/weekly_epidemiological_update_22.pdf

Yadav, S. (2020). Deep Learning based Safe Social Distancing and Face Mask Detection in Public Areas for COVID-19 Safety Guidelines Adherence. *International Journal for Research in Applied Science and Engineering Technology*, *8*(7), 1368–1375. https://doi.org/10.22214/ijraset.2020.30560

Yoo, S. H., Geng, H., Chiu, T. L., Yu, S. K., Cho, D. C., Heo, J., Choi, M. S., Choi, I. H., Cung Van, C., Nhung, N. V., Min, B. J., & Lee, H. (2020). Deep Learning-Based Decision-Tree Classifier for COVID-19 Diagnosis From Chest X-ray Imaging. Frontiers in Medicine, 7, 427. https://doi.org/10.3389/fmed.2020.00427

Zainol Rashid, Z., Norlia OTHMAN, S., Najihan ABDUL SAMAT, M., Kalsom ALI, U., & Kon Ken, W. (n.d.). *Diagnostic performance of COVID-19 serology assays*.

Appendix

Appendix A: Dataset Features

```
In [3]: dataset.columns

Out[3]: Index(['Gender\n1: male\n2: female', 'Age\nyears',
               'Occupation\n1: employed\n2: unemployed',
               'Marital  status\n1: single\n2: married',
               'Number of family \nwho infected', 'Symptom duration\ndays',
               'Smoking status\n1: Smoker\n2: Non-smoker ',
               'Travel history to \nthe high-risk regions\n1: yes\n2: no ',
               'Quarantine before admission \n1: yes\n2: no',
               'Hypertension\n1: yes\n2: no', 'Ischemic heart disease\n1: yes\n2: no',
               'Diabetes\n1: yes\n2: no', 'Immunological problems\n1: yes\n2: no',
               'Acute kidney disease\n1: yes\n2: no ', 'COPD\n1: yes\n2: no',
               'Cancer\n1: yes\n2: no', 'Sneeze\n1: yes\n2: no',
               'Runny nose\n1: yes\n2: no', 'Fever\n1: yes\n2: no',
               'Fatigue\n1: yes\n2: no', 'Dry cough\n1: yes\n2: no',
               'Headache\n1: yes\n2: no', 'Shortness of Breath\n1: yes\n2: no',
               'Diarrhea\n1: yes\n2: no', 'Nausea\n1: yes\n2: no',
               'Sore throat\n1: yes\n2: no', 'Vomiting\n1: yes\n2: no',
               'Abdominal pain\n1: yes\n2: no', 'Pulse Rate\n /min',
               'Respiratory Rate\n /min', ' SpO2\n %',
               'Systolic blood pressure\nmm Hg', 'Diastolic blood pressure\nmmHg',
               'PCR results\n1: positive\n2: negative',
               'White blood cell count\n x109 /L', 'Neutrophil count\n x109 /L',
               'Lymphocyte count\n x109 /L', 'Platelet count\n x109 /L ',
               'Hemoglobin\n g/dL ', 'LDH\nU/L ', 'AST\nU/L', 'ALT\nU/L', 'ALP\nU/L ',
               'Na\nmeq/L', 'Ca\nmg/dL', 'P\nmg/dL', 'Mg\nmg/dL', 'BUN\nmg/dL',
               'Cr\nmg/dL ', 'Primary composite endpoints\n1: positive\n0: negative'],
              dtype='object')
```

Appendix B: Sample Database Table

+ Options

←T→		id	oxygenlevel	heartrate	time_stamp
☐	Edit Copy Delete	1	96	51.18	2021-09-17 09:15:46
☐	Edit Copy Delete	2	95	73	2021-09-22 11:08:47
☐	Edit Copy Delete	3	93	84.81	2021-09-24 10:37:39
☐	Edit Copy Delete	4	0	47.73	2021-09-24 10:37:44
☐	Edit Copy Delete	18	244	10.13	2021-09-27 12:50:29

Appendix C: Naïve Bayes Model Accuracy

```
In [35]: #Naive Bayes
         from sklearn.naive_bayes import GaussianNB
         from sklearn.metrics import confusion_matrix,accuracy_score
         from sklearn.preprocessing import StandardScaler
         #sc = StandardScaler()

         #X_train = sc.fit_transform(X_train)
         #X_test = sc.transform(X_test)
         classifier = GaussianNB()
         classifier.fit(X_train, y_train)
         y_prednaive  =  classifier.predict(X_test)
         print('y_pred', y_prednaive)
         print('y_test', y_test)

         cm = confusion_matrix(y_test, y_prednaive)
         ac = accuracy_score(y_test,y_prednaive)
         ac

         y_pred [1 1 1 1 1 2 2 1 1 2 2 1]
         y_test [1 2 1 1 1 1 2 2 1 1 1 1]

Out[35]: 0.5833333333333334
```

Appendix D: Decision Tree Model Accuracy

```
In [37]: #Decision Tree
         from sklearn.tree import DecisionTreeClassifier
         from sklearn.tree import DecisionTreeClassifier
         from sklearn.metrics import confusion_matrix,accuracy_score, classification_report

         dtree=DecisionTreeClassifier()
         dtree.fit(X_train,y_train)
         print('Decision Tree Classifier Created')
         dtree=DecisionTreeClassifier()
         dtree.fit(X_train,y_train)
         print('Decision Tree Classifier Created')
         from sklearn.metrics import classification_report, confusion_matrix#for visualizing tree
         y_preddecision = dtree.predict(X_test)
         #print("Classification report - \n", classification_report(y_test,y_preddecision))

         cm = confusion_matrix(y_test, y_preddecision)
         ac = accuracy_score(y_test,y_preddecision)
         ac

         Decision Tree Classifier Created
         Decision Tree Classifier Created

Out[37]: 0.6666666666666666
```

Appendix E: Ethics Review Certificate

Final Decision

This document certifies that the study:

\\\"Automating COVID-19 Screening Protocols in Buildings\\\"

Principal Investigator: Ms. Bonyo, Lesley Auma
Reference number: SU-IERC0960/20

Was reviewed and received the following status:
\\\"approved\\\"
Additional Comments:

No Comments from the Reviewers.

Appendix F: Similarity Report

Curiginal

Document Information

Analyzed document	IOT PULSE OXIMETRY MODEL FOR EARLY DETECTION OF COVID-19.pdf (D109940783)
Submitted	2021-06-30 11:21:00
Submitted by	
Submitter email	Lesley.Bonyo@strathmore.edu
Similarity	13%
Analysis address	library.strath@analysis.urkund.com