COVID-19 & Obesity: Artificial Intelligence (AI) and software-based analysis using Machine learning Models (ML).

By

Sadafa Mehreen.

Student Id-2385711.

Supervisor Dr Uday Reddy.

A thesis submitted to the University of Birmingham.

For the degree of MSc in Artificial Intelligence and Computer Science.

School of Computer Science.

University of Birmingham.

Dubai, United Arab Emirates.

September 2023.

Contents

Acknowledgement.

I would like to thank Dr Leonardo Stella and Dr Uday Reddy for their continuous support and coordination throughout my project. Their guidance and advice from the start has been truly benefiting and inspiring for me and motivated me to approach the project in a positive manner. I would also like to thank my parents and my colleagues from the University of Birmingham, Dubai for their constant support and believing the best in me while I was working on this project.

Abstract.

The obesity epidemic and COVID-19 pandemic have led to a complex interplay between global health challenges. Changes in diet and lifestyle during pandemic has led to a global surge in people being obese and losing track of their dietary patterns .During the pandemic many studies were conducted to study the linkage between these two global challenges to predict the mortality rates such as the Healthy Artificial nutrition Analysis Model [1] but few have considered the impact of COVID-19 and obesity on changes in diet of individuals when transiting from COVID-19 to post COVID-19 era. The WHO (World Health Organization) had declared the end of pandemic in May 2023, but the surge of COVID-19 new variants continues to rise. Hence it becomes necessary in the present world to tackle these challenges from a broader spectrum, keeping in mind the emergence of COVID-19 that leads to a drastic change in all the activities worldwide. A critical problem is to determine how strong the linkage between these two global concerns is and how much it has affected the health of individuals in a broader way. We focus on this aspect by using machine learning models and predicting the two global concerns relationship along with developing an evidence-based web application called ‘OBC HEALTH ‘which acts an early warning system for healthcare professionals by providing them with relevant information on obesity and COVID-19 and helps them in handling the potential future risks of the two global challenges. The web application acts as a self-assessment tool for individuals to assess their potential long-term risks of COVID-19 and BMI levels before approaching a healthcare expert for a medical opinion. We also extend our findings in the form of a survey which will be used to assess the change in dietary and lifestyle habits of individuals during and post COVID-19 period and help us give valuable insights for developing web application and for knowing the transitionary changes of food and lifestyles habits of individuals.

Keywords*: COVID-19, obesity, web application, machine learning models, linkage, early warning system, .*

CHAPTER I

1.1 Introduction.

Since the COVID-19 outbreak, there has been a continuous surge in obesity among individuals globally. All countries actively report obesity rates exceeding 20%. Moreover, there has not been any instance of the lowering of obesity rates in any country till this date, making COVID-19 and obesity a common phenomenon in both developing and developed countries worldwide [2]. This global crisis demands a universal focus and requires critical evaluation to develop an understanding of the possible link between obesity and COVID-19 severity even in the post-COVID times [3]. Policy measures taken to settle the consequences of COVID-19 are paving the way for demanding economic challenges. The pandemic had bound all nations to enforce movement restrictions. These adaptations of measures have led to disruptions in the food system, changing food habits and physical activity routines to a larger extent. The already existing home working styles may further intensify the existing trends in obesity rates. These changes signify lasting effects that exist beyond addressing the immediate concerns of COVID-19 and can potentially compromise public health. Subsequently, a critical evaluation of the already existing studies relating to the correlation between obesity and COVID-19 is needed along with in-depth explorations of the results arising from this relationship between COVID-19 and obesity to solve these global issues in depth and overcome future health challenges [3].

1.2 Literature Review.

Obesity, a known risk factor for respiratory infection, has been increasingly recognized as a predisposing factor in the development of COVID-19 [4]. This has important implications on global health as excess weight, usually represented by a raised body mass index (BMI), affects vast numbers of people worldwide. The World Health Organization (WHO) estimates that nearly out of 2 billion individuals globally, 650 million experience obesity. It is important to clarify whether these individuals should be considered at increased risk of developing severe long COVID-19.

A Healthy Artificial Nutrition Analysis (HANA) model was proposed in[1]. The HANA model was used to generate a food recommendation system and track individual habits during the COVID-19 pandemic to ensure healthy foods are recommended used ML (Machine Learning) bases analysis to study the nutrition styles of individuals during the COVID-19 period. The main aim of this proposed model was to design a fast-learning model using reliefF and stochastic gradient as a healthy food recommendation system that incorporated WHO advice for COVID-19 and nutrition studies. However, this study did not consider the body mass index (BMI) of individuals rather it was more focused on the mortality rates due to COVID which gives an edge to our study as we focus on the BMI as well as the mortality rate of an individual due to COVID-19. This makes our study unique as we not only recommend diets but also provide a self-assessment tool for individuals to predict their potential risk for long-term COVID-19 effects in the long run.

Research proposed by N Rangelov, and M Schluneger [5] suggested a Web application called the 'The Altea Network 'which educated individuals on long-term COVID-19 effects. This network was launched by a nonprofit organization in Zurich which specializes in respiratory health and recognizes the need to address post-COVID-19 conditions in an evidence-based and accessible way. It educates individuals through stories of COVID-19-affected patients and engages individuals in forums to discuss their complications on a larger scale. This is very close to our study but doesn't consider the suggestions or recommended diets for these COVID-19-affected people as we do on our web application. Hence our study is unique in its way as it not only educates individuals on COVID-19 long-term effects but also provides solutions in the form of recommended diets and charts for a healthy lifestyle. Another study analyzed the various mechanisms of obesity affecting the severity of COVID-19, it summarized both individual-level and hospital-level preventive and management measures for COVID-19 and discussed the impact of isolation on people with obesity. This gives an edge to our paper as we evaluate the mechanisms of obesity in general and see its correlation with COVID-19 with the help of machine learning models [6].

A study done by the European Association for the Study of Obesity (EASO) reveals that COVID-19 and obesity have two realms. One is that obesity is the most significant risk factor for COVID-19 and the other one is that health emergency caused by COVID-19 diverts the attention from the prevention and care of non-communicable chronic diseases to communicable diseases like COVID-19 resulting in an imbalance of focus to be given to reduce the spread of these diseases in the long run. This aspect provides a broad spectrum for our study as we not only shed light on the effects of COVID-19 overall but see how likely an individual is prone to getting long-term effects of COVID-19 along with predicting the BMI levels of an individual to analyze their obesity levels with the help of the BMI calculator present in our web application OBC HEALTH. [7].

Many of the studies have addressed obesity and COVID-19 with mortality rates but fewer studies are showing how the pandemic has affected the obese people post-pandemic and how diet affects the overall health of these individuals collectively. Nevertheless, while the linkage between obesity and COVID-19 is well established there is still very little evidence regarding how the diet intake of these individuals matters and how it has changed post-pandemic. Thus, the main goal of this paper is to assess the dietary patterns of obese people affected by COVID-19 during COVID-19 and post-COVID-19 and recommend effective diets to tailor their overall health accordingly in the long run.

1.3 Paper Overview.

The remainder of the paper is as follows.

- Chapter II will talk about the background, highlight the problem statement, and explain the significance along with relevant works and gaps present in the previous works.

- Chapter III introduces the methodology starting with data collection, defining the machine learning models used and the evaluation metrics.

- Chapter IV confirms the models by highlighting the evaluation metrics and outlines the performance of the models in detail alongside some visual analysis.

- Chapter V highlights the web application development which will include the user interface of the web application ‘OBC HEALTH along with the survey to know the changes in food and lifestyle habit during and post-pandemic with the survey analysis and critical evaluation of it as well as the testing for the web application which includes the feedback form for user testing and validation.

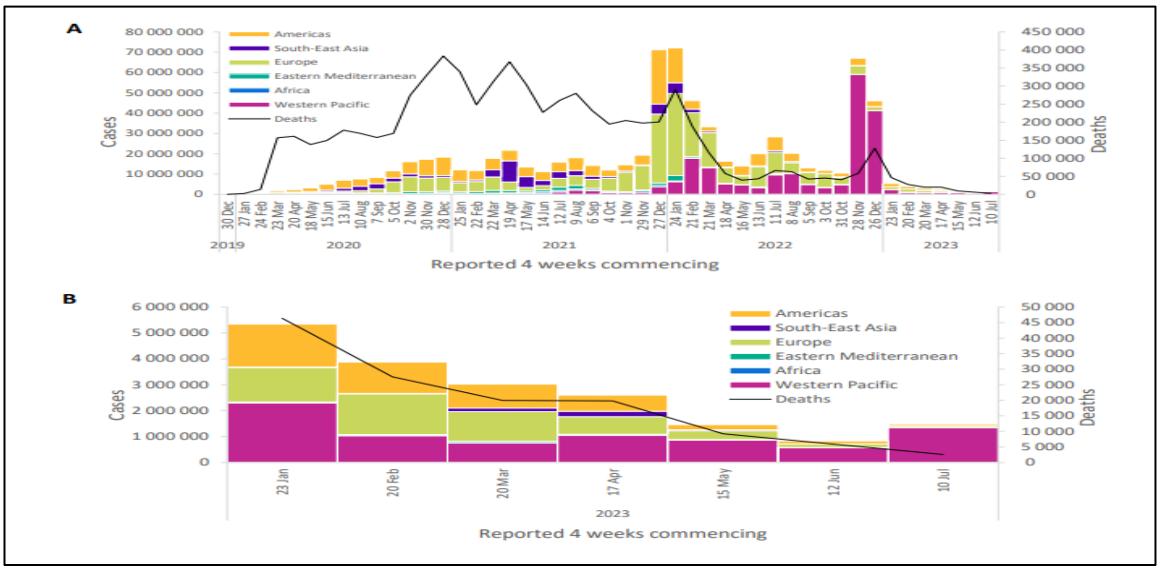
- And lastly Chapter VI will talk about the future works ,limitations and conclusion for the paper.

CHAPTER II

2.1 Background.

COVID-19.

The emergence of COVID-19 originated from a novel coronavirus (SARS-COV-2) that developed in China's Wuhan province. The earliest COVID-19 case was recorded in December 2019. Subsequently, the disease spread rapidly with even more variants in recent years. The most recent one being called the EG.5 strain also known as the Eris variant appeared in the world on 9th August 2023 according to the WHO, it's known to be a descendant of the previously deadly variant Omicron. Despite implementing subsequent interventions, COVID-19 continues to stride globally resulting in global morbidity and mortality rates[8]. The emergence of COVID-19 has paved the way for the intersection between other non-communicable diseases. Various metabolic diseases like diabetes, obesity, and hypertension have been linked with the prolonging and severity of COVID-19, making it one of the hot topics for research and development. As of 6 August 2023, over 769 million confirmed cases and over 6.9 million deaths have been reported globally. Although the public health emergency of international concern for COVID-19 was declared as 'over' on 5 May 2023, COVID-19 remains a major threat as a highly mutated new COVID-19 variant, tagged as EG.5, has been found in a handful of countries, including Denmark, USA, UK. The World Health Organization (WHO) has kept the variant under monitoring and is urging the public to take preventive measures against the variants. [9].



In fig (1) shows the COVID-19 cases region wise as per The World Health Organization reported on August 6th, 2023[10].

Appendix H

Obesity.

As per The World Health Organization(WHO),obesity is defined to be accumulation of excess fat in the body which can potentially affect health conditions. Using the body mass index(BMI) as a measure an individual can determine obesity levels accurately. The BMI scale is classified in 5 categories.ie

Table 1.

BMI	Weight Status
<18.5	Underweight
18.5-24.9	Healthy
25-29.9	Overweight
30-39.9	Obese
40+	Extremely obese

In the past, overweight and obesity were initially perceived as challenges limited to affluent nations, however these two causes are now witnessing a remarkable surge in low- and middle-income countries, particularly within the urban areas. Most of the individuals dealing with obesity mostly exist in developing nations, where the increase rate of COVID-19 has surpassed the rate of obesity in developed nations by 30%.The evolution of diets in the recent years has brought a shift towards individuals consuming energy-dense foods abundant in fats and sugars. This dietary transformation is accompanied by a decline in physical activity, giving rise to sedentary lifestyle changes and the growing phenomena of urbanization. Mitigating the risk of obesity involves strategies such as reducing calorie consumption from fats and sugars, increasing intake of fruits vegetables,legumes,whole grains,fruits,whole grains, and nuts, and embracing regular physical activity(with recommended duration of 150minutes per week for adults [9]. The estimation of obesity global levels also referred to as high obesity suggest that over 4 billion people may be affected by 2023,compared with over 2.6 billion in 2020.This shows an increase from 38% of the world's population in 2020 to over 50% by 2035.The prevalence of obesity (BMI $\geq$30kg/m^2) alone is anticipated to rise from 14% to 24% of the population over the same period, affecting nearly 2 billion adults, children and adolescents by 2035[11]

The interlink between obesity and COVID-19 .

Multiple studies reveal a strong association between COVID-19 and obesity. COVID-19 patients with obesity have an enhanced hospitalization rate, more severe progression, and worse clinical outcomes. Studies proposed by López de la Torre [12] also suggest that there exists a strong linkage between obesity and restricted mobility on health. During the lockdown, the most prevalent explanation for the weight gain that impacted nearly half of the world's population was a mix of increased food intake and sedentary lifestyle. The purchasing of high-calorie products such as alcoholic beverages, candies, and snacks surged by more than 50% after the first few weeks. Furthermore, the lockdown restricted access to sports facilities and made it difficult to engage in physical activity outside, which, combined with a lack of regular exercise at home, made it difficult to maintain an active lifestyle [12].This possibility of interlinking demands a comprehensive approach to be tackled and eliminated either through analyzing the linkage or verifying it with the help of machine learning models.

2.2 COVID-19 mortality rates of individuals with and without obesity.

A study done on 482 patients in Italy figured out obesity to be a strong independent risk factor for severe diseases and mortality caused due to COVID-19. While patients with a BMI>30kg/m2 had a high risk for severe illnesses and chronic diseases, a BMI of 35kg/m2 or more subsequently increased the likeliness of death to a greater extent. Cài et al. investigated the relationship of obesity with the severity of COVID-19 in designated hospitals in Shenzhen, China and concluded that the obese patients had increased chances of progressing towards the long-term effects of COVID-19 in the long run [13].

Differences in the rate of death involving coronavirus (COVID-19) between people with and without obesity may be partly explained by people with obesity having a higher prevalence of several health conditions. For instance, the prevalence of cardiovascular diseases, diabetes, and asthma before the COVID-19 pandemic was higher among people with obesity than those without obesity [12]

People without obesity or any underlying medical conditions may be subject to ‘immune system gaps’ a norm where some people who seem perfectly healthy and are not termed as immune-deficient may be not prone to having inherited immune system features that make them more at risk than average to certain viral infections. This may be due to random genetic versatilities or a lack of earlier exposure to viral infections could be more likely to develop COVID-19. Studies and research linking immune system gaps to COVID-19 susceptibility are in their beginning stage and more up-to-date information might be needed to address this issue [13]

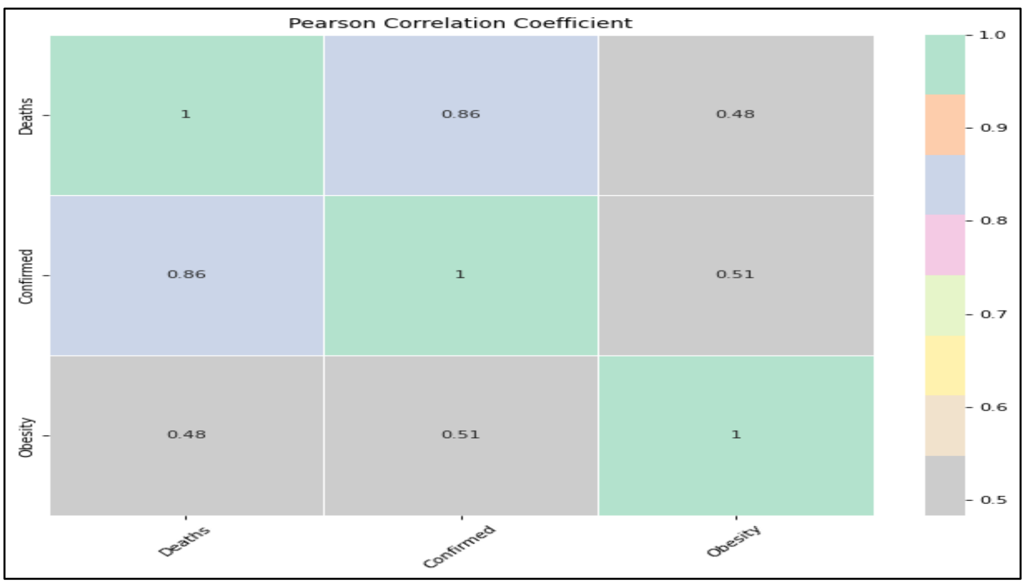
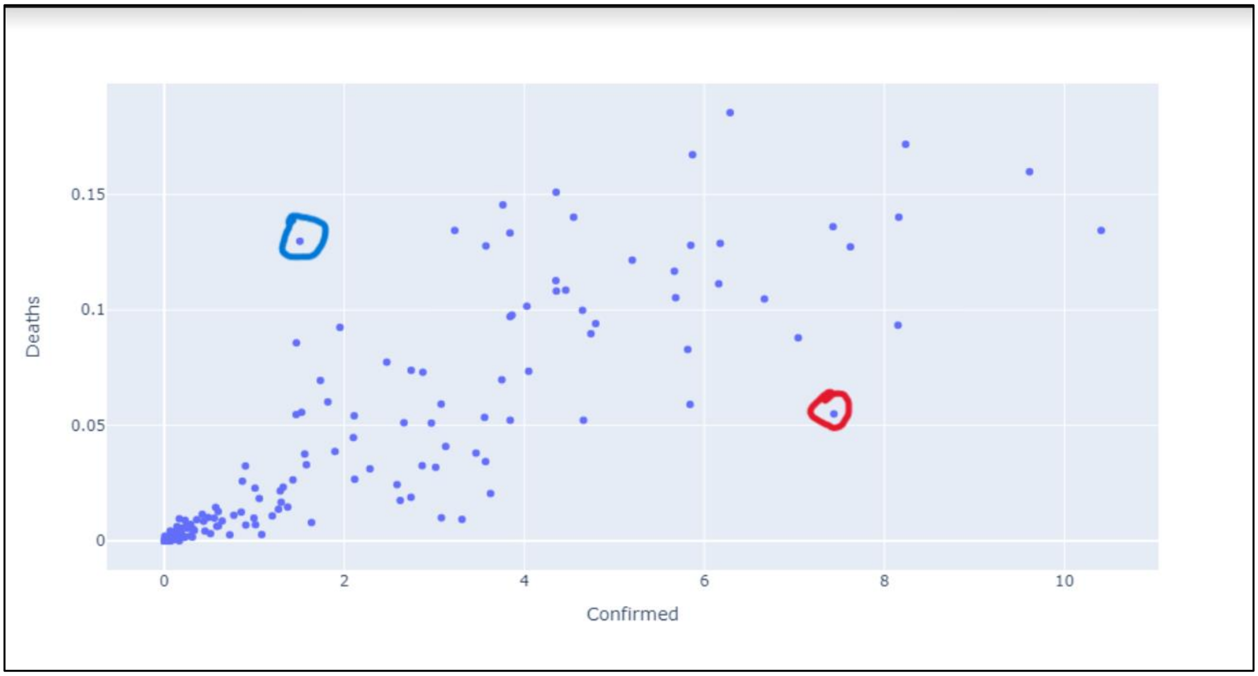


Fig (2) shows the Pearson correlation coefficient between obesity, confirmed cases and deaths due to COVID-19

It measures the direction and strengths of the relationships between variables ranging from -1 to 1.The relationship between deaths and obesity is 0.48 indicating a medium correlation between the two variables relationship between confirmed cases and deaths show 0.86 indicating a very strong correlation between them and the relationship between obesity and confirmed cases of COVID-19 show a medium correlation. This shows that the confirmed variable has the highest correlation with deaths indicating that the risk of mortality increases when an individual gets inflicted by COVID-19.

2.3 COVID-19 mortality rates and COVID-19 Confirmed cases.



Fig(3) shows that Mexico (Blue circle) is having highest death rate in comparison to the confirmed rates. And Israel (Red circle) is having a higher rate of confirmed rates compared to death rates in the country. More discussion on this is given below.

1.Mexico.

Mexico ranks among the top five countries with the highest excess mortality with COVID-19 being the leading cause. The largest increases in cause-specific mortality occurred in diabetes, heart diseases, and hypertension, all linked to obesity. The increase in other causes which led to death was mainly due to health service utilization changes or misclassification. Mexico has extensive experience with infectious diseases, it was the epicenter of the 2009 swine flu pandemic. But unlike many nations, it has not rushed to scale up testing. Mexico did not report its first case of COVID-19 until Feb. 27, 2020, a month after the virus was detected in the United States. This shows the reason why the number of COVID-19 cases wasn't reported, and the statistics show that Mexico has the lowest reported confirmed cases of COVID-19 when compared to other countries in the world [14].

2.Israel.

The mortality rate in Israel was kept low despite high rate of confirmed COVID-19 cases as the outbreak was handled very well by the decision makers of the country. The government shut down many activities early and protected high-risk populations. When you look at the number of cases in Israel, it’s relatively low per capita, but still not as low as the mortality rate. Israel emerged as one of the leading countries in the world in testing people accurately for COVID-19 at fast pace in limited amount of time. This helped the country in early detection of the disease and early prevention. The other reason also being that most of the population affected in the country constituted young people who had an exceptionally low mortality rate as compared to the elderly [15].

If you look at the distribution of sick people in Israel, fewer than 5% are over the age of 80. And in Israel, the population over 70 and 80 was quite well protected. Resulting in low death rates among elderly individuals [16].

2.4 Problem Statement.

There exists a connection between obesity and COVID-19.Changes in diet and lifestyle activities during COVID-19 have contributed to this cause. It is important to address the correlation between them as each one increases the vulnerability of one another. Effective strategies should be identified to combat these global challenges. The pandemic according to The World Health Organization has officially ended as per May 5th, 2023.But the long-term effects of the disease continue to linger in people especially those who already have existing conditions one such is the obesity condition.

Hence to reduce and prevent these two challenges we analyze ways to suggest people with effective remedies such as personalized diets with the help of a web application. The web application educates people about the long-term effects of the two global causes and helps them estimate their body mass index accurately with the help of a BMI calculator and help them assess how prone they are too long-term effects on COVID-19 with the help of COVID-19 assessment tool. These steps will make individuals aware about their health conditions and make them more responsive to changing their lifestyle and diet patterns in a more positive manner.

2.5 Significance of the study.

This paper aims at using machine learning models to predict the interlink between COVID-19 and obesity. The paper is driven by the following significant features.

1.The obese population is identified as the most risk if infected by COVID-19.This paper underlines the need for centralized long-term strategy and leadership in fighting global obesity as the largest long lasting global health issue and the recent pandemic that has developed at an alarming rate worldwide.Analysing the link between COVID-19 and obesity by bridging the gap between them can guide medical treatment decisions and help healthcare providers tailor the treatment plans uniquely while addressing the individual challenges and complications of both the causes.

2.The link between COVID-19 and obesity could inspire further research into the underlying mechanisms that make obese individuals more susceptible to severe illness. This research could lead to innovative treatments and interventions that can mitigate the impact of obesity on COVID-19 susceptibility and outcomes more effectively.

2.6 Related works.

A possible protective effect of obesity (obesity paradox) in relation to infectious diseases has previously been proposed in [17], however, several studies evaluating risk factors in relation to COVID-19 severity have pointed out obesity to potentially augment the severity of the disease, including increased risk of hospitalization and an increased need for invasive mechanical ventilation (IMV) [3]. There are various other phenomenon's related to obesity such as[4] present the habits of individuals in eating. The study of 170 countries was performed to discover the correlation between these habits and death rates caused by COVID-19 based on machine learning approaches taking the distribution of energy, fat, and protein through twenty-three different sorts of diets into consideration. The results indicate that 95% predicted correctly using a regression model based on Principal Component Analysis (PCA). but this study did suggest ways to improve the overall health of individuals. [18] Shams et al. [19] proposed a regression model Based on Support Vector Machine (SVM) and Deep Learning (DL) approaches given a dataset containing both confirmed deaths and recovered cases. The results achieved indicate that the RMSE using SVM's with the Radial Basis Function (RBF) kernel is 0.27, while the SVM with linear Kernel achieves 0.18 RMSE, and the deep regression model achieves 0.29 RMSE [20].

The researchers in [21] formulated an epidemic model called SAIR (susceptible-asymptomatic-infected-removed that discriminates between asymptomatic and symptomatic COVID-19 infections. This work develops a predictive model which highlights the impact of asymptomatic infections in spreading the disease through two different measures of their interactions. The measures include the presence of large number of asymptomatic cases of COVID-19 and the intensity of the asymptomatic infections on individuals. This study was one of the first attempts to use the Ista seroprevalence study to model the evolution of COVID-19 at a national level and by making use of complex networks to model the inter-regional spread of the virus for the situation in Italy [21].

Another study included the use of Virtual reality technology (VR) to visual the impact of social distancing in an effort to mitigate the spread of COVID-19.The approach involved was multi-disciplinary as it included the digital twin application named soDAIVR (Social Distancing Algorithm in Virtual Reality).This research as unique in its kind as it used a renowned game machine 'Unity' in its application for visualizing the effect of social distancing due to COVID-19in a university based environment.[22]

CHAPTER III

3.1 Methodology

This paper is divided into two parts.

Problem classification.

- The first part examines the relationship between COVID-19 and obesity with the help of machine learning models such as linear regression, XGB regressor for COVID-19 Healthy Diet Dataset and XGB classifier for the BMI dataset. There are two machine learning problems discussed in this paper which we will shed light upon.
- The first problem is the regression problem which is to examine the relationship between COVID-19 and obesity with the help of linear regression and the relationship between COVID-19 and other variables of the dataset. The relationship between COVID-19 confirmed cases and obesity is predicted with the help of XGB regressor as this model gives best results in comparison to the linear regression model due to its advantageous features such as capacity to handle large outliers and imbalances in the dataset more accurately. The justification of which is given in the critical evaluation of the models. However, the nature of this prediction or analysis may involve various statistical and machine learning techniques beyond regression if the type of variables involved in the prediction get complex. Critical evaluation at the end of the paper given in Chapter (IV) concludes the results.
- The second problem we try to solve is a classification problem for which we will be using the BMI data set that shows the gender, height, weight and obesity levels or classes of individuals ranging from 0-5 indexes The XGB classifier is used to analyze this dataset which will help us predict the BMI levels of individuals and help us develop a BMI calculator for our web application which will be discussed in the next section.

Web application development.

- The second part of the paper will talk about the Web application 'OBC HEALTH' curated for assessing the individual's obesity levels with the help of a BMI calculator and their potential risk of long-term COVID-19 effects with the help of COVID assessment tool. An inbuilt survey on the Web application will be used to gain insights on how the food habits and lifestyle has changed of people post-pandemic and what changes they need to make in their lifestyles to prevent obesity and reduce the long-term COVID-19 effects to a large extend. Suitable diet plans are listed, and informative interventions are displayed to educate the individuals on how to live a healthy lifestyle and bridge the gap between the COVID-19 pandemic and the obesity epidemic gradually. The Web application serves as an early warning system to the healthcare providers giving them the flexibility to intervene the two causes proactively with the help of primary data generated by the inbuilt survey and allocate the resources efficiently to combat these two challenges effectively over time.

3.2 Data collection

The datasets used in this paper are taken from Kaggle, a repository of numerous data science and artificial intelligence datasets. Two types of datasets are used for this research the links of which are given below:

1.COVID-19 Healthy Diet Dataset | Kaggle

2.BMI Dataset | Kaggle

Data sources and information.

1. The COVID-19 healthy diet dataset.

The first dataset used is the COVID-19 healthy diet dataset that was prepared during the COVID-19 pandemic when there many death cases due to the disease and various changes in eating habits and lifestyle of people. It consists of 31 columns and 171 rows .The columns show different types of food consumed by each country of the world in percentages along with the total population of each country in numbers. The information provided in this dataset is crucial for our study as it contains information about COVID-19 such as death rates, recovered and active cases along with obesity and malnourished figures that will help us determine the relationship between COVID-19 and obesity accurately. The other information in this dataset includes data of different types of food consumed by people which is stored in five comma-separated values format files. The data in these datasets except the food description dataset is in percentages. It mainly shows the different types of food consumed by people in various countries. The categories of food include alcoholic beverages, animal products, animal fats, aquatic products, cereals excluding beer, oil crops,eggs,seafood,sugar and sweetners,fruits,meat,miscellaneous,milk excluding butter,offal,spices,starchy roots,pulses,stimulants,meat,tree nuts,vegetable oils,vegetal products, and vegetables (all calculated as percentage of total intake amount). The dataset is divided into five comma-separated files containing the following :

- The Fats are represented in (%) in each category of food in the dataset.
- Food consumed(kg) in (%) of 170 countries.
- The (%) of energy consumption (calories) from different categories of food.
- The percentage of protein is present in different categories of food.
- Detailed description of the subcategories of food present in each category of the dataset.

This information would help us in determining the diet of people during COVID-19 and which all food categories were consumed the highest which can help us analyze the possibility of countries that have the greatest number of obese people and determine the most COVID-19 affected country by analyzing the confirmed ,active and death rates given in percentages. By collecting this information, we can draw our conclusions from the survey highlighted on the Web application that will provide us insights on how food habits of people have changed post COVID and recommend suitable diets to people for adopting a healthy lifestyle in the long run.

2.The BMI Dataset

The second dataset used was the BMI dataset taken from Kaggle. This dataset helps us determine body mass index (BMI) of individuals based on gender, height, and weight of individuals. This data set is useful for providing information on the obesity levels of individuals in contrast to the COVID-19 Healthy Diet dataset which only gave us information regarding countries obesity on a whole. This dataset is crucial for our Web application as its the data present in it is used in the building of the BMI calculator that takes input as gender, height and weight of an individuals and gives output as the BMI category or the BMI score in accordance with the 5 classes of the obesity mentioned in the dataset. The formula used to calculate the BMI with the inputs taken from this dataset is:

$$BMI = \frac{Weight}{Height^2}$$

The BMI dataset consisted of 500 columns and 4 rows. The rows contained information of gender, height ,weight and obesity levels of individuals ranging from index 0-5.

Index 0-Extremely weak.

Index 1-Weak.

Index 2-Normal weight.

Index 3-Overweight.

Index 4-Obese.

Index 5-Over obese.

3.Data Pre-processing

For the COVID-19 Healthy Diet Dataset, each of the 4 datasets consisted of 170 rows and 31 columns. These 4 datasets were merged for better analysis and cleaned accurately after which the total entries came up to 680 rows, these rows were then extracted for data cleaning. During the data preprocessing a total of more than 2 null values were found in the dataset in the obesity,active,undernourished column, hence those columns were removed respectively. The 'fillna' method was used to fill information in certain columns where the information could be filled with values, particularly in the undernourished column there was some values that were objects hence those values were converted into integers for better calculations. We have added three extra columns in the dataset while doing the data preprocessing i.e,Total obese to know the number of obese people in each country, the NON-COVID column, to know the number of people who were not inflicted by COVID-19 in each of the country and non-obese column to know the number of people neither affected by COVID-19 or obesity.

For preprocessing the data in the BMI dataset, the gender column consisted of 2 categories i.e,male and female, these two categories were converted into numeric form using one-hot-encoding technique which is a method to quantify the categorical data, since these were object values. After the data preprocessing is complete the machine learning models are trained which will be discussed in section 3.4 of this paper.

3.3 Challenges of the dataset.

Some of the challenges i faced while collecting the information from the datasets include:

- Data quality-The data presented did not contain uniform reporting standards.ie.The data was presented in percentages which made it difficult for analyzing the actual values of the data indicated in the dataset.

- Data integration-Since the data was only available on a country-wise level it became essential to look upon a secondary dataset to get insights on the individuals body mass index (BMI) to analyze obesity. This became challenging as it was time consuming and finding datasets that integrated well with my already existing dataset was difficult.

3.4 Machine learning models .

Our approach with machine learning models is that we will be doing our model training and evaluation with Supervised learning models. These models are given a set of labelled training data. Each record in the data has a set of features and is assigned to a final label (or class). Supervised learning is used for both classification and regression problems: in classification problems, the labels are in categorical and discrete forms, such as “healthy” or “Obese”, while regression problems have continuous numbers, such as “height” or “weight” [23]. Several supervised machine learning models exist, such as linear regression, XGB regressor, naïve Bayes, Support vector machine(SVM),XGB classifiers some of which we will be using in our paper for training and testing of the dataset [24]

The three different models which we will use in our paper are:

1.Linear regression.

2.XGB regressor.

3.XGB classifier.

1.Linear regression.

Linear regression is a supervised machine learning model which predicts the value of a dependent variable (y) against independent value (x) in the form of $y = mx + c$. Linear regression uses the mathematical equation as the line of best fit for the relationship between y (dependent variable) and x (independent variable), m is the slope of the line. And C is the y-intercept. Linear regression is best suited for our research goals as it will help us in prediction of how different types of food present in the Healthy COVID-19 dataset are related to the COVID-19 variables and to help us find the strength of the obesity and COVID-19 linkage to a certain extend [25].

$$y = mx + c$$

Regression can be of two types, linear and nonlinear.

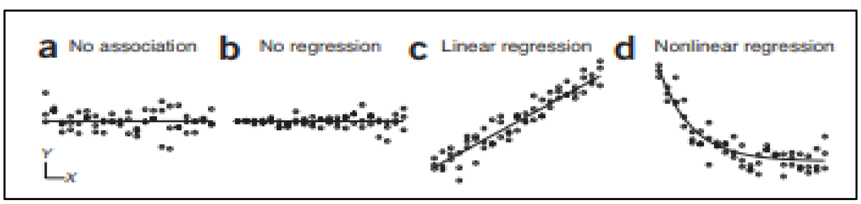


Fig (4) explains different kinds of regressions and their properties in detail. Appendix H.

In Fig (4) [25] (a) If properties of Y do not change with X there is no association.(b) Association is possible without regression. Here f(Y|X) is constant the variance of Y increases with X.(c) Linear regression where there is positive relation between X and Y.(d) Nonlinear regression where there is no relationship between X and Y due to the nature of the variables.

2.XGB Regressor.

XGBoost Regressor Boosting is an ensembled technique that contains decision trees which accurately predict the errors made by previous models with the help of a single decision tree at a time. As per the gradient boosting technique the models are fitted using gradient descent optimization set of rules which also include arbitrary differentiable loss function which helps in reducing the loss gradient accurately. Compared to other machine learning models, XGB Regressor emerges a winner in terms of handling sparse data and instances with varying weights which make it a suitable model for various range of regression tasks. According to Rumahorbo et.al,2023 XGBoost has been used for analysis if significant statistical features relating to facial behavior in human depression [26].

The XGBoost regressor must be developed without overfitting as this may be one of the major issues that conflict in most of the machine learning models. Hyperparametric tuning is one way to reduce overfitting of the model and improve the overall behavior and accuracy. It’s a method to reduce the loss function of a model [27].

3.XGB Classifier.

The XGBoost classifier is a machine learning method that specializes in binary and multiclass classification tasks. It is an ensemble learning approach that combines the predictions of numerous weak learners, such as decision trees, to produce a more powerful model. XGBoost classifier is well-known for delivering outstanding results in machine learning contests and real-world applications. It is performance and efficiency optimized, with techniques like as parallel processing, tree pruning, and regularization used to reduce overfitting and enhance training speed [28]

XGBoost classifier has a feature important ranking, which allows users to see which characteristics contribute the most to the model's predictions. It can manage unbalanced datasets by employing various objective functions and weighting procedures, making it effective in cases with dramatically varying sample sizes. It is capable of handling both numeric and categorical information, as well as optimizing bespoke loss functions for specific problem domain [28].Despite its widespread usage, XGBoost requires rigorous preprocessing and correct hyperparameter adjustment to produce excellent results for specific situations. The XGB classifier is evaluated with the help of classification report which is discussed in the due course of this paper [28].

3.5 Evaluation metrics of the machine learning models.

1.Mean squared error (MSE)

MSE is the most used metric for regression tasks in machine learning. It is calculated as the average of the squared difference between the predicted value and the Actual value. Due to its differentiable property, it is easier to optimize and has the capability to penalize large errors [21].One major disadvantage of MSE is that it is not robust to outliers. In case the sample has an error that is much larger than the other samples in the dataset the square of error will even be larger. Hence MSE calculated the average of errors and is prone to outliers. It's given by the following formula.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_{i-y_i}\right)^2$$

In the equation the $\frac{1}{n}\sum_{i=1}^{n}$ represents the squares of the errors of $\left(\hat{y}_{i-y_i}\right)^2$ which represent the difference between the actual and predicted values.

2.Root means squared error (RMSE).

The Root Mean Squared Error (RMSE) is a relatively common measure for regression models .It is calculated as a difference between the estimated and actual values. It is the square root of Mean square error (MSE). Unlike MSE,the RMSE gives an error measure in the same unit as target variable, it's a very useful metric for continuous numeric variables. However, some of its limitations include that it gives weight to all errors regardless of whether the values are positive or negative. It has the same property for outliers as the MSE and is very sensitive to large outliers. It is shown by the following formula [29].

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{\left(\hat{y}_{i-y_i}\right)^2}{n}}$$

This equation shows the square root of MSE formula. $\left(\hat{y}_{i-y_i}\right)^2$ is the difference between the actual and predicted values which is divided by the total values to get the root mean square error.
[29].

3.Mean absolute error (MAE).

This is just the average of the absolute difference between the target and anticipated values. Shouldn't be used when outliers in the model are prominent. Also, RMSE, MSE and MAE do not penalize large errors mainly due to the absence of squaring in its calculations. Below is the following formula for MAE [30]

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|^2$$

4. R squared (R^2)or coefficient of determination.

This metric represents a part of the variance, or the dependent variable explained by the independent variables of the model .This metric measures the strength of the relationship between them There are 3 scenarios for R^2 score to be satisfied:

1. Perfectly depicts the dependent variable's variance If R^2 is high (assume=1) then the model depicts the variance of the dependent variable perfectly.

2.If the R^2 value is very low this shows that it does not represent the variance of the dependent variable and it would be better for the regression model to take the mean values for the model evaluation

3.A negative R^2 shows that that score of R's^2 is worse than the mean value and that the predictors do not explain the dependent variables to the required accuracy.

Hence it is very important to see the difference in the R^2 and predict the value of R^2 precisely as there may be scenarios where if the R^2value is 1 then all predicted values are the same as actual values and this essentially means that all values fall on the regression line. This doesn't give us a proper analysis. Because sometimes a good model can have low R^2 value and a biased model can have a high R^2 value as well. That is the reason we should make use of appropriate visualizations such as residual plots for estimating better accuracy of the models [24].The formula for R^2 is as follows:

Total sum of squares:

$$Sum = \sum_{i=1}^{n} (y_i - \overline{y_i})^2$$

The sum of residual squares is also referred to as the residual sum of squares:

$$Sum = \sum_{i=1}^{n} (y_i - f_i)^2$$

The most common expression of the coefficient of determination is given by the following equation:

$$R^2 = 1 - \frac{RSS}{TSS}$$

Where RSS is the sum of squared individuals and TSS is the total sum of squares.

Lower MSE, RMSE and MAE indicate the best result and high value for the determinant of coefficient with higher accuracy [31].

5. Classification report.

Performance evaluation after training the classification model is important to evaluate and measure classifier performance. For classification reports a confusion matrix is used for evaluation. The confusion matrix shown in fig (5) [32] counts the number of instances that were truly classified as positive or negative, which are referred to as true positive (TP) and true negative (TN). It also counts the number of cases that were falsely classified as positive or negative, referred to as false positive (FP) and false negative.

		Predicted Class	
	Total Population n = a number	False (0)	True (1)
Actual Class	False (0)	TN True Negative	FP False Positive
	True (1)	FN False Negative	TP True Positive

Fig (5) shows the confusion matrix for the classification report. Appendix H

- Precision.

The precision is the ratio of the predicted values against all other values computed by the XGB classification model.Mathematicsally precision is the number of true positive values divided by the number of true positive values plus the number of false positive values. The true positive values are data points that are classified as positive values by the model (either are correct) and false negatives are values that the model identifies as negative but still are positive (either are incorrect) [32].

$$precision = \frac{tp}{tp + fp}$$

- Recall.

The recall in the classification report is the ability of the model to find all relevant cases within the dataset. It is the ratio of the actual values that were predicted to be belonging to the actual values but now they don't belong to empathetically it is defined as the number of true positives divided by true positives plus false negatives. It's like the precision formula [32]

$$Recall = \frac{tp}{tp + fn}$$

- Fi score

Precision and recall usually have inverse relations, and the f1-score is a metric that measures both precision and recall together. The Fi-score shows a combined figure of both precision and recall. It is an optimal blend of the two metrics also known as the harmonic mean [32]

Harmonic means instead of the simple average is used because it punished the extreme values. For evaluating the model to the best of accuracy the FI score must be maximized with optimal balance of precision and recall. F1 Score might be a better measure to use if we need to seek a balance between Precision and Recall AND there is an uneven class distribution (large number of Actual Negatives) [32]

$$F1 = 2 \cdot \frac{(precision \cdot recall)}{(precision + recall)}$$

- Accuracy.

Accuracy is one metric for evaluating classification models. Informally, accuracy is the fraction of predictions the model gets right. Formally, accuracy has the following definition:

$$Accuracy = \frac{(Number\ of\ correct\ predictions)}{Total\ predictions}$$

OR

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

Where *TP* = True Positives, *TN* = True Negatives, *FP* = False Positives, and *FN* = False Negatives.

Accuracy measures the model's performance across all classes and is beneficial when balanced but can be misleading when dealing with imbalanced datasets.

- Support

And lastly the support is the number of actual instances of the class in the provided dataset is referred to as support. Imbalanced support in the training data may suggest fundamental flaws in the classifier's reported scores, indicating the necessity for stratified sampling or rebalancing. Support doesn't vary between models rather it diagnoses the assessment process. Higher support scores indicate that the precison,recall and F1-scores are likely to be more reliable as higher support score means higher instances for the class which provides a strong base for evaluating the model's performance and accuracy effectively [32].

CHAPTER IV

4.1 Model Evaluation.

For the linear regression model, we take the confirmed cases, it serves as baseline model as it provides a starting point for comparison with other complex machine learning models for example in our case, we used to XGB regressor to train the ‘Obesity’ and ‘Confirmed’ variables instead of using linear regression. In our paper we will be looking at two cases to predict the linear regression between them and the third case would be for the classification problem using XGB classifier.

We have 3 cases to evaluate our models used in both datasets.

1. Case 1 for linear regression.
2. Case 2 for XGB Regressor.
3. Case 3 for XGB Classifier.

- Case 1 (Linear Regression)

Determining the relationship between COVID-19 confirmed cases and other independent variables in the dataset using linear regression.

Predicting the linear regression between COVID-19 confirmed rates and different types of foods mentioned in the healthy COVID-19 dataset will help us to assess which types of food have a positive as well as negative correlation with confirmed COVID-19 cases. We shall be using a bar plot to determine the positive and negative correlation of each of the food categories with the COVID-19 confirmed variable after we determine the relationship using linear regression model.

STEP 1

```
x = df.drop(['Country', 'Confirmed'], axis = 1)
y = df['Confirmed']

# splitting the data

x_train , x_test, y_train , y_test = train_test_split(x,y, test_size=0.2, random_state=20)

# model training

lr = LinearRegression()
lr.fit(x_train , y_train)

y_pred_test = lr.predict(x_test)
y_pred_train = lr.predict(x_train)
```

Fig (6) Training the model with all variables of the COVID-19 Healthy Diet Dataset and confirmed cases of COVID-19.

As shown in fig (6) We choose our independent variable (x).All variables are selected except the ‘Country’ variables. It’s a categorical variable and cannot be quantified for analyses using linear regression.

We choose our dependent variable(y) as ‘Confirmed’ as we want to see how the different categories of food are correlated with the ‘Confirmed’ variable in the dataset.

Other columns such as ‘Obesity’,’Active’,’Recovered’,’Population’ aren’t being analyzed for this case. We then split the data into training and testing sets and perform the model training. Which includes training on the training data (x_train,y_train) and predicting the target variable on the testing data.

STEP 2

```
def rmsse(y_test, y_pred_test):
    # Calculate RMSE
    rmse = np.sqrt(mean_squared_error(y_test, y_pred_test))

    # Calculate mean of the observed data
    mean_observed = np.mean(y_test)

    # Calculate RMSSE
    rmsse = rmse / mean_observed

    return np.around(rmsse, 2)

# Evaluate on training data
mse_train = mean_squared_error(y_train, y_pred_train)
rmse_train = np.around(np.sqrt(mse_train),3)
r2_train = np.around(r2_score(y_train, y_pred_train)*100,2)
score_train = rmsse(y_train, y_pred_train)

# Evaluate on testing data
mse_test = mean_squared_error(y_test, y_pred_test)
rmse_test = np.around(np.sqrt(mse_test),3)
r2_test = np.around(r2_score(y_test, y_pred_test)*100,2)
score_test = rmsse(y_test, y_pred_test)


# Print evaluation results
print("Training Data:")
print("RMSE:", rmse_train)
print("R-squared:", r2_train)
print("RMSSE Score:", score_train)

print("\nTesting Data:")
print("RMSE:", rmse_test)
print("R-squared:", r2_test)
print("RMSSE Score:", score_test)
```

fig(7) shows the performance of the linear regression model.

We evaluate the model on both training and testing data and find the accuracy of the model along with the evaluation metrics of linear regression which include: RMSE,R^2 and RMSE score. Fig (7) explains the process in detail.

First, we start with calculating the root mean squared error (RMSE) to assess the model's performance in terms of accuracy and the variability in the target varible.Then we evaluate the model using the required metrics, that is R^2 and RMSSE score. The RMSSE score is computed as 0.0 indicating that the predictions made perfectly match the observed data in the model indicating no prediction error. And the R^2 which represents a part of the variance of the dependent variable explained by the independent variables of the model shows that a perfect 100% score is achieved indicating that the model is overfitting and requires regularization methods to be applied to reduce the overfitting to some extend such as lasso regression.

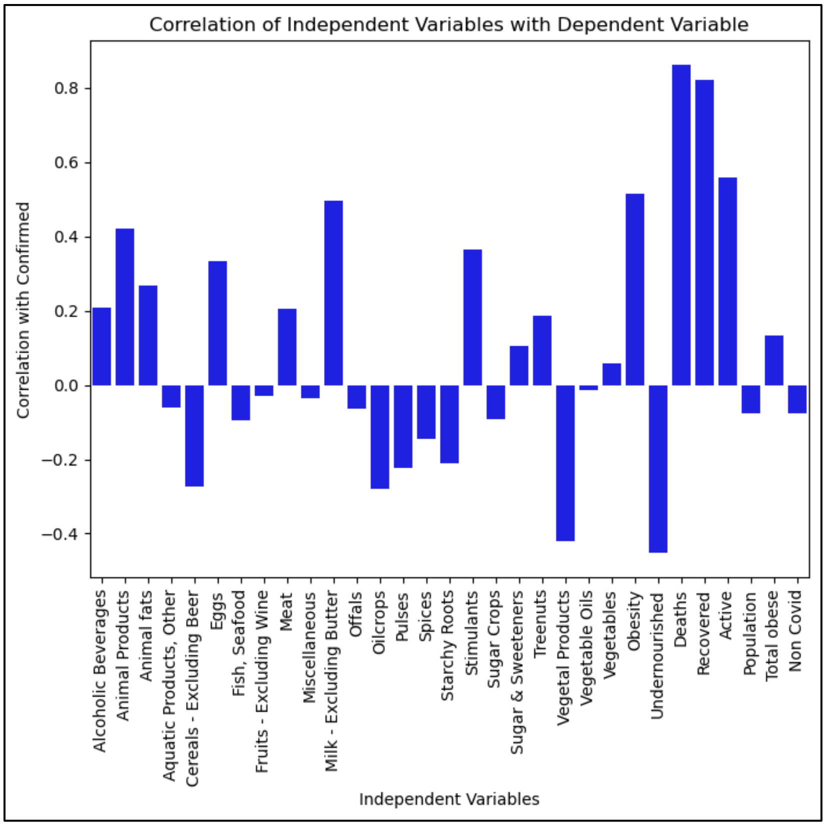
Step 3

```
# Create the Lasso regression object with the regularization parameter alpha
lasso_model = Lasso(alpha=0.25)

# Fit the model to the training data
lasso_model.fit(x_train, y_train)

# Make predictions on the test data
y_pred_test = lasso_model.predict(x_test)

# Make predictions on the training data (to check for overfitting)
y_pred_train = lasso_model.predict(x_train)

# Evaluate the model on the test data
mae_test = mean_absolute_error(y_test, y_pred_test)
mse_test = mean_squared_error(y_test, y_pred_test)
rmse_test = mean_squared_error(y_test, y_pred_test, squared=False)
r2_test = r2_score(y_test, y_pred_test)

# Evaluate the model on the training data
mae_train = mean_absolute_error(y_train, y_pred_train)
mse_train = mean_squared_error(y_train, y_pred_train)
rmse_train = mean_squared_error(y_train, y_pred_train, squared=False)
r2_train = r2_score(y_train, y_pred_train)

# Calculate RMSSE for the test data
residuals_test = y_test - y_pred_test
rmse_consecutive_test = np.sqrt(mean_squared_error(y_test[1:], y_test[:-1], squared=False))
rmsse_test = rmse_test / rmse_consecutive_test

# Calculate RMSSE for the training data
residuals_train = y_train - y_pred_train
rmse_consecutive_train = np.sqrt(mean_squared_error(y_train[1:], y_train[:-1], squared=False))
rmsse_train = rmse_train / rmse_consecutive_train

# Print the evaluation metrics for the test data
print("Test Data:")
print("Mean Absolute Error (MAE):", np.around(mae_test, 3))
print("Mean Squared Error (MSE):", np.around(mse_test, 3))
print("Root Mean Squared Error (RMSE):", np.around(rmse_test, 3))
print("R-squared (R2):", np.around(r2_test, 3))
print("Root Mean Squared Scaled Error (RMSSE):", np.around(rmsse_test), 3)

# Print the evaluation metrics for the training data
print("\nTraining Data:")
print("Mean Absolute Error (MAE):", np.around(mae_train),3)
print("Mean Squared Error (MSE):", np.around(mse_train, 3))
print("Root Mean Squared Error (RMSE):", np.around(rmse_train, 3))
print("R-squared (R2):", np.around(r2_train, 3))
print("Root Mean Squared Scaled Error (RMSSE):", np.around(rmsse_train), 3)
```

Fig (8) shows the method to reduce the overfitting of the model.

To reduce the overfitting of the model Lasso regression (Least Absolute Shrinkage and Selection Operator) is applied as shown in fig (8). It is a method that minimizes prediction error by identifying variables and their corresponding regression coefficients. It achieves this by shrinking the regression coefficients towards zero, forcing the absolute value to be less than a fixed value (λ). This constraint reduces the model's complexity, and variables with zero regression coefficients are excluded. The value of alpha is determined by predicting a suitable value that minimizes prediction error on the validation data. Hence lasso regression was applied to the model with regularization parameter alpha set to 0.25. The model was based on the training data and the predictions were made on the same to check if the overfitting has reduced. After evaluating the model both on training and testing data the evaluation metric scores improved to a certain extent. The R^2 score was reduced to 0.983 for training data and 0.987 for testing data and the RMSSE score and the RMSE improvised respectively and were similar for both training as well as testing data indicating that the model has fitted properly with the values. The scores were low indicating that the models' predictions are very close to the actual observed values and there is no room for errors [33].This suggests that linear regression was the best choice made for analyzing the variables taken in case 1.

Step 4



A bar plot is shown in fig (9) which shows the variables in the dataset that have a positive and a negative correlation with the target variables ('Confirmed') in the dataset.

From the above bar plot, we can see that there are some variables positively correlated and some negatively. The variables which are above zero are positively correlated and below zero are negatively correlated with COVID Confirmed variable.

Food categories that are positively correlated variables are : 'Alcoholic Beverages', 'Animal Products', 'Animal fats', 'Eggs', 'Meat', 'Miscellaneous', 'Milk - Excluding Butter', 'Stimulants', 'Sugar & Sweeteners', 'Trees, 'Vegetables', 'Vegetable Oils', 'Obesity.'

Food categories that are negatively correlated variables are : 'Aquatic Products, Other', 'Cereals - Excluding Beer', 'Fish, Seafood', 'Fruits - Excluding Wine', 'Offal's', 'Oil crops', 'Pulses', 'Spices', 'Starchy Roots', 'Sugar Crops', 'Vegetal Products'.

The observations show that foods high in saturated fats, sugars, and refined carbohydrates show positive correlation with the target variable. Higher consumption of these foods contributes to obesity and type 2 diabetes, increasing the risk of severe COVID-19. Therefore, promoting wider access to healthy foods and maintaining healthy eating habits among individuals of the countries is necessary to reduce susceptibility and long-term complications from COVID-19.it is critical to consider the impact of lifestyle habits, such as consumption of unhealthy diets, on the susceptibility to COVID-19 and consume high amounts of fiber, whole grains fruits and vegetables that are present as the negatively correlate variables in daily lives to boost the immune system and live a healthy lifestyle in the long run [34].

- Case 2. (XGB Regressor).

Determining the relationship between COVID-19 confirmed cases and obesity using linear regression.

Predicting the linear regression between COVID-19 confirmed cases and obesity is the main goal of our paper and analyzing this relationship will help us determine whether or if there is a linear regression or a linkage between the two variables of not. There are other studies that have tried to work with COVID-19 and linear regression but the relationship they tried to show with were with different variables such as the HANA model [1] which tried to perform linear regression to show the relationship between COVID-19 deaths and different type of food consumed by 170 countries taken from the COVID-19 Healthy Diet Dataset. But there are few studies that analyze the relationship between COVID-19 and obesity particularly on long-term effects of COVID-19 on individuals. Our study aims to find that relationship and see whether the linear regression model is the best fit or other regression models are required for better performance.

Step 1

```
# LINEAR REGRESSION FOR ObESITY AND COVID 19.

x_1 = df[['Obesity']]
y_1 = df['Confirmed']


# Split the data into training and testing sets
X_train1, X_test1, y_train1, y_test1 = train_test_split(x_1, y_1, test_size=0.2, random_state=42)

# Initialize and fit the linear regression model
lr = LinearRegression()
lr.fit(X_train1, y_train1)

# Make predictions on the test set
y_pred1 = lr.predict(X_test1)
```

Fig (10) Splitting and training the model based on obesity and confirmed variables.

As shown in fig (10) ,We choose our independent variable (x_1) as ‘Obesity and dependent or target variable (y_1) as ‘Confirmed’. As we want to predict the relationship of obesity with COVID-19 and determine the likeliness of obese individuals getting affected by COVID-19.

We then split the data into training and testing sets and perform the model training. Which includes training on the training data (x_train1,y_train1) and predicting the target variable on the testing data.

Step 2

```
# Calculate evaluation metrics (Root Mean Squared Error and R-squared)
rmse = np.sqrt(mean_squared_error(y_test1, y_pred1))
r2 = r2_score(y_test1, y_pred1)

# Print evaluation metrics
print("Root Mean Squared Error:", rmse)
print("R-squared:", r2)

# Visualize the linear regression line
plt.figure(figsize=(8, 6))
plt.scatter(X_test1, y_test1, color='blue', label='Actual')
plt.plot(X_test1, y_pred1, color='red', linewidth=2, label='Predicted')
plt.xlabel('Obesity')
plt.ylabel('Confirmed')
plt.title('Linear Regression Prediction')
plt.legend()
plt.show()
```

```
Root Mean Squared Error: 2.0357902953139275
R-squared: 0.23011405862611056
```

Fig (11) evaluates the linear model for obesity and confirmed variables on training and testing sets.

In fig (11) We then evaluate the model on both training and testing data and find the accuracy of the model along with the evaluation metrics of linear regression which include: RMSE and R^2 .Fig (11)explains the process in detail. First, we start by calculating the root mean squared error to assess the model performance in terms of accuracy and the variability in the target varible.Then we evaluate the model using the required metrics that is R^2 and RMSSE.The RMSSE is computed to be 2.03 indicating a bad fit for the model. The RMSE should be between 0 to 1.Thelower the RMSE the better for the model indicating that the model is able to fir the dataset properly. And the R^2 shows that the model is giving only 23% accuracy which indicates that there is a large variation between the actual and predicted values, and All RMSEvalues ≥ 0.5 reflects the poor ability of the model to accurately predict the data. Hence, we need a better model to capture the dataset with the desired variables accurately.

<u>Step 3</u>

```
from xgboost import XGBRegressor

xg = XGBRegressor()
xg.fit( x_train ,y_train )
y_pred_test_xg = xg.predict(x_test)
y_pred_train_xg = xg.predict(x_train)

# Evaluate on training data
mse_train = mean_squared_error(y_train, y_pred_train_xg)
rmse_train = np.around(np.sqrt(mse_train),3)
r2_train = np.around(r2_score(y_train, y_pred_train_xg)*100,2)
score_train = rmsse(y_train, y_pred_train)

# Evaluate on testing data
mse_test = mean_squared_error(y_test, y_pred_test_xg)
rmse_test = np.around(np.sqrt(mse_test),3)
r2_test = np.around(r2_score(y_test, y_pred_test_xg)*100,2)
score_test = rmsse(y_test, y_pred_test)


# Print evaluation results
print("Training Data:")
print("RMSE:", rmse_train)
print("R-squared:", r2_train)
print("RMSSE Score:", score_train)

print("\nTesting Data:")
print("RMSE:", rmse_test)
print("R-squared:", r2_test)
print("RMSSE Score:", score_test)
```

```
Training Data:
RMSE: 0.001
R-squared: 100.0
RMSSE Score: 0.13

Testing Data:
RMSE: 0.029
R-squared: 99.98
RMSSE Score: 0.15
```

Fig (12)shows the XGB regressor model.

To solve this problem, we tried to use XBG regressor fig (12) as our next best model to predict the relation between COVID-19 confirmed variable and obesity. The model showed very high accuracy compared to linear regression. The evaluation metrics R^2 and RMSE score show prominent results. The R^2 score for training data is 100 and for the testing data is 99 and the RMSE score for the training data is 0.13 and for testing data is 0.15brespectively indicating that the model is fitted perfectly with only very little variations of about 1% in the evaluation scores of trainings and testing sets of the XGB regressor model. This suggests that XGB regressor is best suited for predicting the relationship between confirmed and obesity rather than linear regression Fig (12) shows the working of XGB regressor in detail along with the evaluation metrics.

- <u>CASE 3 (XGB Classifier).</u>

The results for the classification report for the BMI dataset are as follows:

The input for the model was the height and weight of either male or female and the output given by the model is the body mass index of an individual which will help us determine the BMI levels of individuals when we prune this model into the Web application. This will help us filter individuals based on their health metrics and recommended diets for the same.

Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1
1	0.67	1.00	0.80	4
2	1.00	0.83	0.91	18
3	0.80	1.00	0.89	8
4	0.93	0.83	0.88	30
5	0.90	0.95	0.92	39
accuracy			0.90	100
macro avg	0.88	0.94	0.90	100
weighted avg	0.91	0.90	0.90	100

Fig (13) shows the classification report.

The XGB classifier is trained on 5 classes of the BMI as mentioned in the BMI dataset and in Chapter II of this paper. The overall accuracy of the model is 90%, which is satisfactory. The precision scores of all the classes of the BMI are between 0.5-1.0 indicating that the model's predictions for different BMI classes are reasonably accurate with values closer to 1.0 indicating a high level of precision. The recall also falls in the same range.Therecall scores for all the BMI classes fall between 0.8-1.0 indicating that the model has a high ability to correctly identify and capture true positive cases within each class. The recall scores indicate that the model is very effective in minimizing false negatives and ensuring most of the actual cases of the BMI are correctly identified.

The high recall scores suggest that the model is well suited for identifying and addressing obesity related data making it valuable for public health experts and medical applications where accurate detection is crucial. The high scores of F1shows a positive aspect of the model particularly for specific classes like in our case for class 0 and class 2 and class 5 indicating that the model is effectively identifying true positives (high recall) while minimizing false positives (high precision).This suggests there is a good balance between the precision and recall score highlighting the robustness of the model and lastly the model's macro average F1-score of 0.90 and weighted average F1-score of 0.90 indicate that the model performs well across classes.

However, it is critical to examine all the requirements of the dataset. for example, If certain classes are more crucial than others, we may need to prioritize precision or recall. Also, assessing the support for each class is vital, especially for classes with low support, since smaller sample sizes may impact performance measurements. To conclude the model captures all the features of the BMI dataset very well making it the best model for evaluation.

4.2 Critical evaluation of models.

For the linear regression we talk about why XGB regressor is better suited for obesity and confirmed COVID-19 variables than using linear regression. A comparison evaluation metrics table is given below:

Table 2.

Evaluation metrics	Linear regression	XGB Regressor
RMSE	2.03	0.029
R^2	0.23	99.98

1.Model accuracy-The model XGB regressor is advantageous in our case as the relationship between the obesity and the COVID-19 confirmed variable is not linear as shown in the previous model of linear regression while trained ,the XGB regressor captures the nonlinear relationships very well and interacts between the features while reducing noisy data to a larger extend. The linear regression models only work when the relationship between the continuous variables is linear and is noisy free. The R^2 score of XGB regressor is highest almost having a difference of 99.75 % when compared to the score of R^2 in linear regression.

Better predictability of the dependent variables-The RMSE score shows us how closely the model's predictions align with the actual dataset. In the table we can see that linear regression is having a RMSE of 2.03 which is quite high and suggests there is large variation between actual and predicted values and XGB regressor is having a RMSE of 0.029 that is in XGB regressor the difference between the actual and predicted values is only 0.029 which is quite low and good for the model and suggests that XGB regressor is a better model for the COVID-19 confirmed and the obesity variable.

The classification report provides us with a clear picture of how well the model distinguishes between different BMI classes. The report is well balanced in terms of the scores as most of the classes have high FI score and recall scores indicating the model correctly identifies most of the relevant instances while making very few false positive errors. The macro and weighted average scores show up to 90% accuracy which indicates there is very little chances fir class imbalances to be present and lastly with the accurate support values we get an estimation of the distribution of the number of samples in each class of BMI.Overall the report gives best results for the model and gives us insights on how well the model distinguishes between different BMI classes of the BMI dataset.

4.3 Visual Analysis

- Country wise Obesity and COVID-19.

<u>1.Top 5 high obesity countries from the COVID-19 healthy diet dataset.</u>

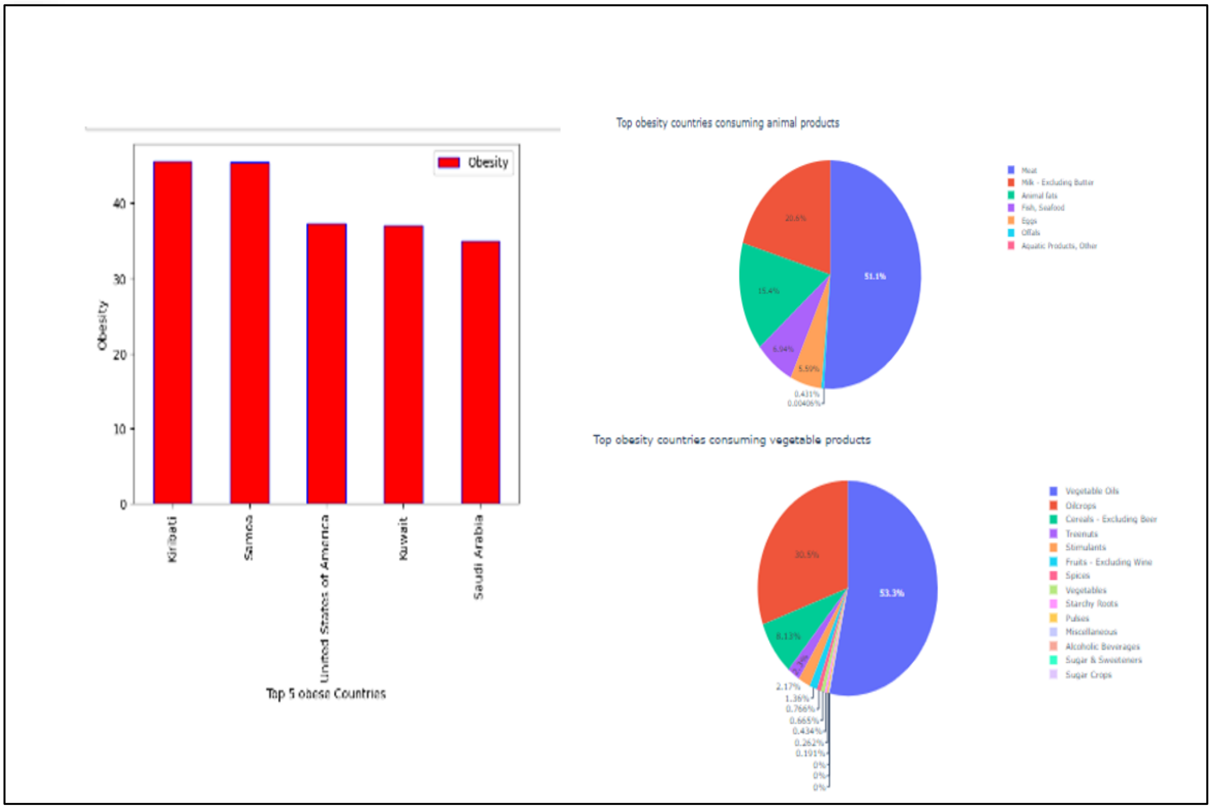


Fig (14) shows the Obesity in top 5 countries from the COVID-19 Healthy Diet dataset and shows the consumption of animal and vegetable products in those countries.

Fig (14)Urbanization in Kiribati, Samoa has led to a shift in the diet of Pacific Islanders, who rely on traditional, local plantations and traditional nutritious food. However, urbanization introduces Western diets, including processed energy-dense food, sweets, refined grains, and fat. Cultural perceptions and addictive tastes of processed food contribute to unhealthy eating habits. The consumption of sodium-rich, oily, and sweetened foods is common among Pacific Islanders, while fruits and vegetables are rarely included in meal menus. This has led to an increase in overweight and obesity rates in Gulf countries[35]. The World Health Organization reports that Gulf countries have the highest rate of obesity, with Kuwait, Saudi Arabia being among the top ten. Kuwait is the worst affected, with a 42.8 percent obesity rate. Countries like Saudi Arabia also have high obesity rates. Increased public awareness about healthy eating habits and promotion of physical activity and exercise is needed. A diet rich in fruits and vegetables can help reverse this trend. [36].

<u>2.Top 5 low obesity countries from the COVID-19 healthy diet dataset.</u>

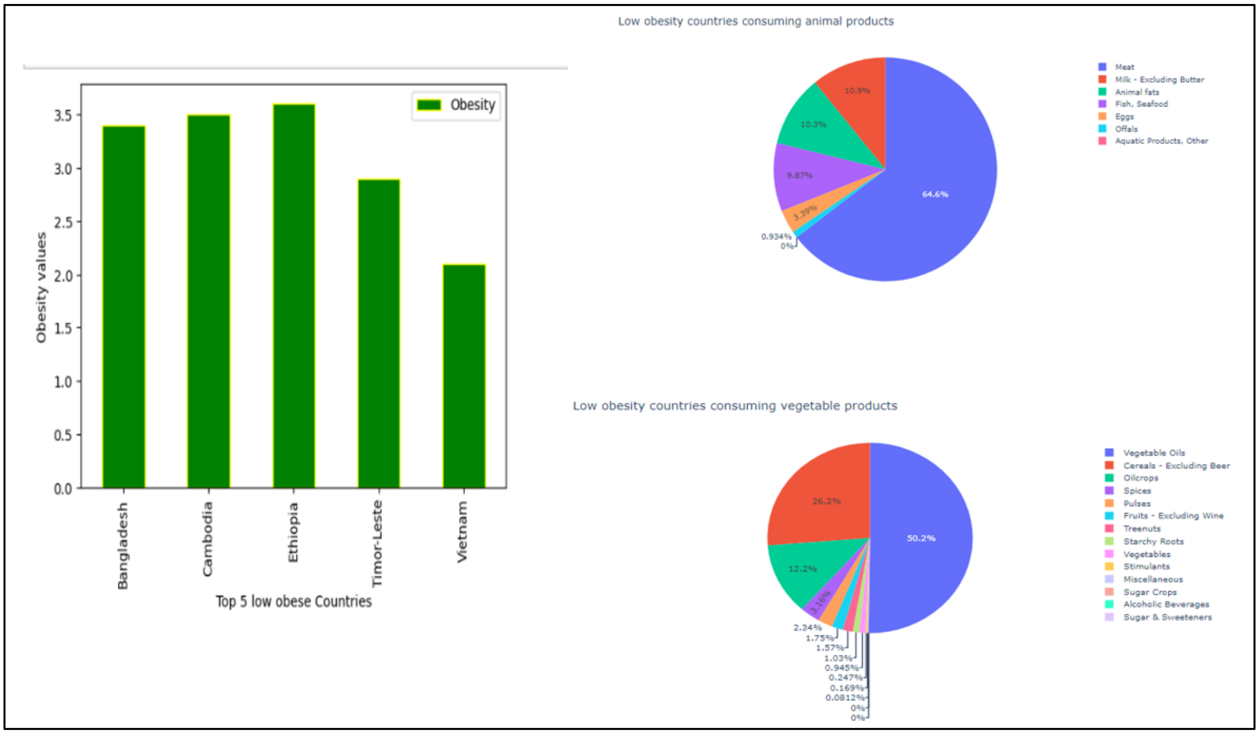


Fig (15) shows the top 5 countries in which obesity is low from the COVID-19 Healthy Diet dataset and shows the consumption of animal and vegetable products in those countries.

The top 5 low obesity countries include Vietnam,Timor-Leste,Bangladesh,Cambodia, and Ethiopia.

The relatively low levels of obesity observed in countries as shown in fig (15) such as Vietnam, Timor-Leste, Bangladesh, Cambodia, and Ethiopia can be ascribed to a combination of factors rooted in culture, diet, socioeconomic status, and lifestyle. In Vietnam, over the past 10 years, the economic situation has improved dramatically following the introduction of the social and economic policy reforms. It is considered that the resulting changes in the economy have in part contributed to changing morbidity patterns and lower cases of obesity in the country. In Timor-Leste,Bangladesh and Cambodia the 'obesity' factor only exists in the lower income groups particularly amongst the rural population and is less prevalent in the high-income groups due to urbanization and better access to healthy foods and healthcare facilities. People in these countries often lead less sedentary lives, with fewer occupations requiring prolonged periods of sitting and more physically active daily routines. Robust social networks and community support systems often encourage healthier behaviors, including cooking meals at home and engaging in physical activities [37].

3.Top 5 high COVID-19 confirmed cases countries from the COVID-19 healthy diet dataset.

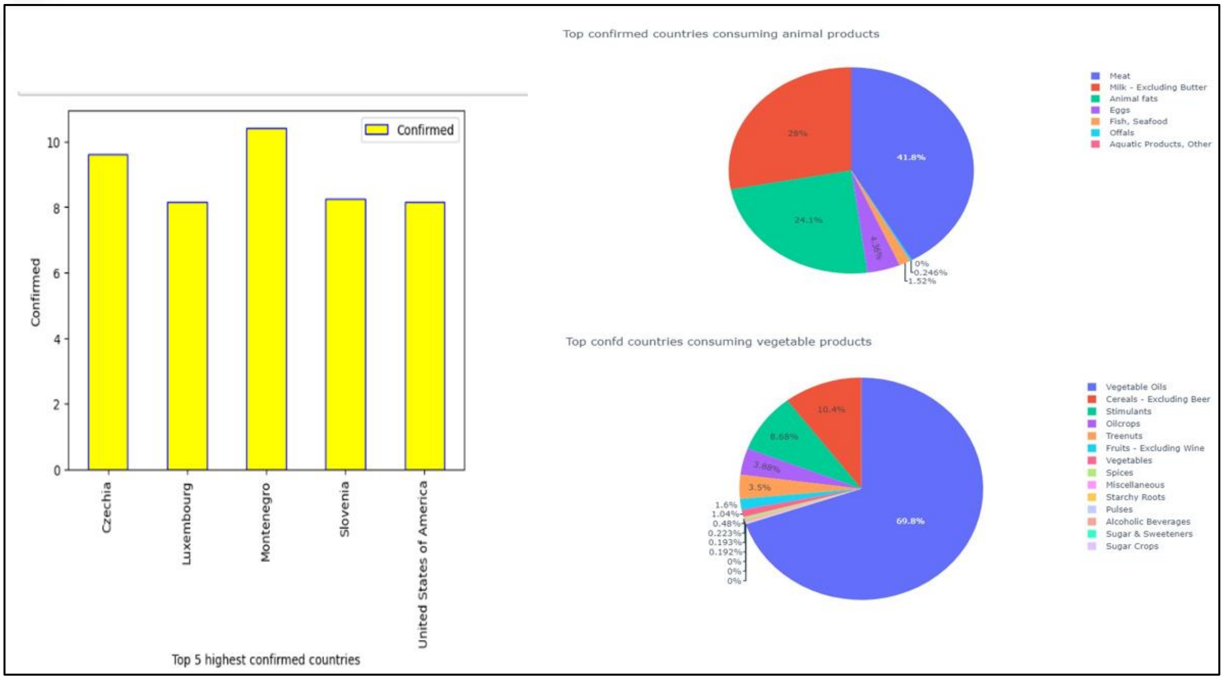


Fig (16) shows the Highest confirmed COVID19 cases in top 5 countries from the COVID-19 Healthy Diet dataset and shows the consumption of animal and vegetable products in those countries.

The top 5 confirmed countries include Montenegro, Czechia, Slovenia, United States of America, Luxembourg.

The number of COVID-19 cases remain high in these regions as shown in fig (16) mainly due to the lack of proper vaccination programs and campaigns that did not reach out to the vast population of people present in the countries which could have led to the surge as many people did not get the protection with the help of the COVID-19 vaccine and ended up spreading it to others. Secondly these countries imposed the COVID-19 restrictions very late as compared to other countries in the world such as Montenegro,Czechia,Slovenia, and Luxembourg due to lack of healthcare facilities or shortage of contact tracing for COVID-19 which made the situation much worse leading to a spike in the cases drastically over time. Whereas in USA due to the vast population the testing for COVID-19 began in very last stages and by then half of the population was influenced by COVID-19 indicating poor healthcare facilities in the country [38].

4.Top 5 low COVID-19 confirmed cases countries from the COVID-19 healthy diet dataset.

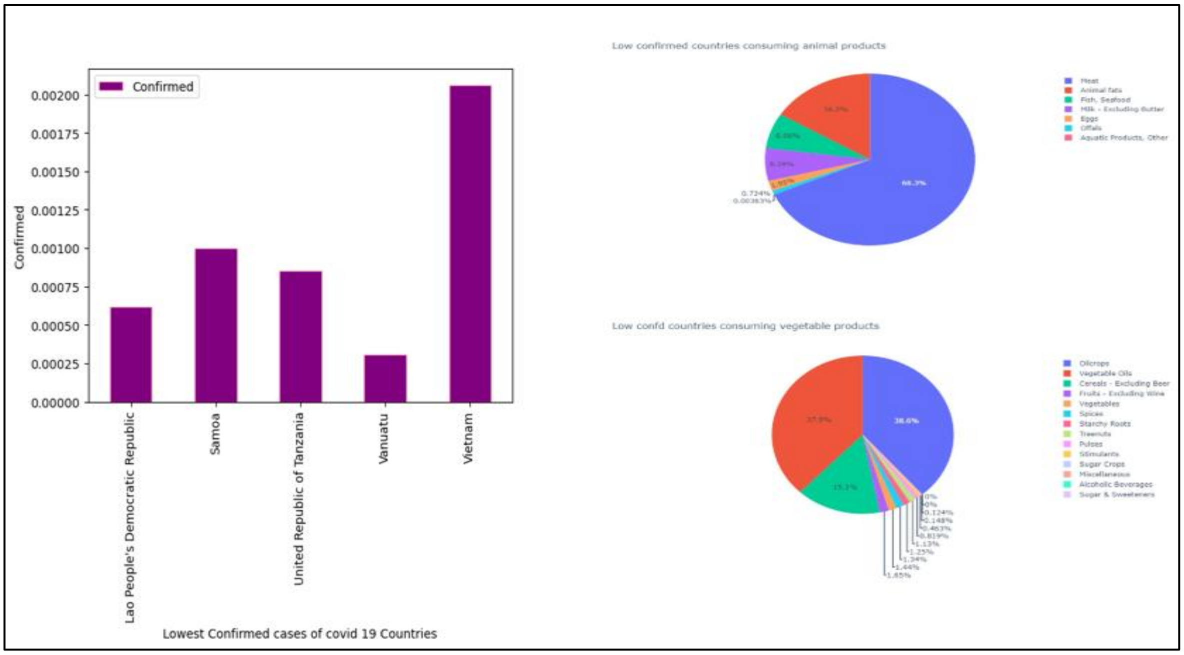


Fig (17) shows the Lowest confirmed, in top 5 countries from the COVID-19 Healthy Diet dataset and shows the consumption of animal and vegetable products in those countries.

The number of COVID-19 cases remained low in these regions. as shown in fig (17) as the governments imposed strict protocols in the country such as the Vanuatu government has moved quickly and has locked down the main island of Efate after recording its first case of COVID-19 after a citizen who was repatriated from the United States tested positive while in quarantine. Effective testing measure were in place such as vaccination campaigns that lead to a decrease in the COVID-19 cases and the government of Tanzania despite lacking the same testing capacity as other regions in the world, the low numbers of confirmed COVID-19 cases can be explained in part due to experiences in handling infectious diseases on the continent.

Firstly, resources meant for widespread HIV and tuberculosis testing were leveraged in the fight against COVID-19 [39].

Secondly, Governments were swift in imposing lockdowns, restricting movement, and setting up task forces to coordinate efforts [40]

Thirdly, though not backed by evidence, the issue of weather may have given Africa the much needed “lifeline” [41].compared to Europe, Africa had a lower importation risk of the virus based on the data on the volume of air travel from China to Africa [42] .

The Samoa Red Cross Society (SRCS) and the Adventists Disaster Relief Agency (ADRA) were some of the WHO initiatives to intensify public education on COVID-19.And Vietnam was named the gold standard country[44] when it came to handling COVID-19 mainly due to the government mostly mitigating the prevention strategies and adopting strict contact tracing and rigorous quarantine made it one of the lowest COVID-19 country when compared to other countries in the world [43].Due to countries following tight regulations and thorough testing techniques, the number of COVID-19 cases in these countries has remained low. Tanzania’s administration, on the other hand, lacks the same testing capabilities as other areas ,while the Samoa Red Cross Society (SRCS) increased public awareness of COVID-19.

CHAPTER V

5.1Web application development

Software used.

• Streamlit is a Python open-source framework for developing interactive web apps for data science and machine learning projects. It enables developers to quickly convert data scripts into shareable web apps. While Streamlit is essentially a Python library, it may be used in conjunction with a variety of other tools and services to build more complex applications. We have used streamlit to code our web application 'OBC HEALTH'.

•Vscode (Visual studio code) has gained immense popularity in the software development community due to its combination of powerful features, flexibility, and a vast ecosystem of extensions. It is widely used for web development, mobile app development, data science, and a broad range of programming tasks. We have used vscode for editing and viewing our code.

Sources of information.

•The BMI dataset at the backend for the BMI calculator as the BMI calculator uses the gender, height, and weight present in the dataset as inputs and gives the output as the BMI score from the information mentioned in the dataset. The XGB classifier explained in Chapter IV is used for the development of the BMI calculator in the Web application.

• A survey is required to assess the changes in individual's food and lifestyle habits before and after COVID-19.This survey was conducted using google forms and was circulated to individuals before making the Web application. The responses from this survey helped in the curation of the suitable diets on the Web application for people to lead a healthy lifestyle. The survey form is shown in Appendix B

• User testing and evaluation was also conducted using a feedback form which was given to 5 students of MSc in Artificial intelligence and computer since after they watched the video of the full Web application and were required to pen down their suggestions through the feedback form made with the help of google forms.

Challenges.

There are several challenges associated with this web application:

•There may not have been enough time to implement all the features I have wanted to add in the webpage as I juggled to balance between he machine learning models and the development of the Web application I focused only one aspect for example dedicated all the time to the Web application I would have been able to add more interactive features in the Web application and increased my software knowledge to the fullest.

•As we have just transited towards the post COVID-19 phase, and it is all new for people to adapt to the post-pandemic changes it was quite hard for me to gather information regarding how the food and lifestyle habits have changed as many people lacked the expertise mainly because of their inability to predict immediately and to judge themselves well.

•Deploying the Web application on the cloud was costly and required various membership. Hence finding the right framework to build my Web application which was easily accessible to all was challenging and time consuming.

5.2 Design and implementation .

Survey implementation.

The survey is designed to know how the diet and lifestyle patterns changed of individuals when transiting from COVID-19 to post COVID-19 period. The survey is aimed to analyze the responses of individuals and built-up recommended diets on the Web application OBC HEALTH based on their interests and the change in their diet and lifestyle pattern observed. This will help us in reaching our goal of promoting healthy lifestyle, educating individuals on diets that they can incorporate in their daily life leading to striving for a balance when transiting from COVID to post COVID times.[45]

Participants.

Inclusion criteria were people of ages ranging from 25-60.The individuals that participated in the survey consisted mainly of working professionals and university students. Children and adolescents were excluded from the study as child or adolescent obesity wasn't evaluated for the survey. This may be one of the limitations but if children and adolescents were included the scope of the survey would be vast, making the interpreting of results more challenging and difficult. Consent was not required for collecting the responses of the survey. As no personal information such as name, contact number or email was required from the participants. Participants were informed in the beginning of the survey that their responses will only be recorded for the study purpose and will not be used for medical purposes.

Ethical considerations were considered. The participant's identity was kept anonymous. Respondents were not allowed to record their responses more than once to validate the authenticity of the survey and they were not allowed to change their responses once submitted, this ensured transparency of the survey [46]

Data were collected through a digital google form which was shared to the participants through social media platforms (Watsapp,Instagram,Linkedin etc.) [47].The form was shared using the following link https://forms.gle/i4JtQ2QjXVm7siCz8.The survey contains 12 questions in total that analyzed various aspects of the individual including their COVID-19 status, diet and lifestyle patterns. It asks users of their contraction with COVID-19,the change in their food habits, the health metrics, Their knowledge on other aspects related to obesity and assess what type of dietary advice they rely on. This survey forms the backbone of the Web application OBC HEALTH. It acts as a catalyst to help individuals to adapt to a healthy lifestyle and healthcare professionals to identify the change in the people's attitude and behaviors towards food from COVID to post-pandemic period.

A sample of the survey form is shown in Appendix B.

Challenges for the Survey.

There was time crunch in collecting the responses. Due to the nature of the project and the duration given to complete it, I was only able to dedicate two weeks to the responses collection and the next one week was spent based on analyzing the responses in form of pie charts. Because of this bias may be present [48]

5.3 Survey Analysis.

The survey consisted of eleven closed ended questions and one open ended question that asked respondents how their food habits changed during the pandemic. The other questions mainly targeted on knowing the pandemics effects on the respondents' behaviors towards adopting to the transition from pandemic to post-pandemic [49].

100 individuals ranging from 25-60 years old responded to the survey out of which 56% of them said they contracted COVID-19 in the past three years, with 44% saying no. This suggests that many individuals, particularly those over 25 years old, may have underlying diseases that make them more susceptible to the virus.55% of the respondents reported that their food habits did not change during the pandemic, however the ones that responded 'yes' (45%) said that it lead to unhealthy eating habits and decreased intake of healthy food choices Travel restrictions (39.4%), home quarantine(28.3%), and lockdown measures (26.3%) disrupted their lives, impacting vacations, work-from-home routines, social isolation, and psychological well-being. Most respondents believed that heart disease (38%) is the top risk factor linked to COVID-19, followed by lung disease (25%) and hypertension (17%). This highlights the importance of public health education and communication in educating individuals at higher risk about the COVID-19's susceptibility and severity. Many respondents also preferred annual body checkups, with a significant portion undergoing them twice a year. The high frequency of checkups aligns with the global shift towards a more proactive and holistic approach to healthcare and well-being of individuals Majority of respondents exercise 2-3 days a week (29.9%), with a smaller group exercising rarely. This pattern may be due to personal motivation, pandemic-related restrictions, and evolving health and fitness perceptions. During the pandemic, individuals adjusted their exercise routines to lockdowns and social distancing, leading to a moderate exercise frequency. As restrictions ease, there's potential for increased exercise levels, potentially contributing to healthier BMI outcomes. However, individual behaviors and lifestyle choices will vary, necessitating public health initiatives to encourage regular physical activity as part of a holistic health and wellness approach.

The survey found that individuals' weight changes during and post-pandemic were varied, with the most common responses being "a few months back"(36.1%) and "a few days back"(28.9%). This highlights the dynamic nature of weight management which can be influenced by factors like pandemic-related stressors and lifestyle changes.

A significant number of respondents do not feel worried about weight changes, highlighting the diverse attitudes and emotional responses people have towards weight fluctuations, particularly during and post-pandemic times. This understanding can help shape recommended diets and public health strategies in the post-pandemic era.

The survey reveals a diverse dietary preference among respondents, with non-vegetarians(56.6%), omnivores(27.3%), and vegetarians(16.2%) being the majority. This underscores the importance of respecting individual needs, especially in a post-pandemic. A personalized, flexible approach is crucial for effective diet plans to cater to diverse dietary preferences, promoting a balanced, nutrient-rich diet for overall health.

The survey shows that many people rely on self-assessment(54%), medical advice(26%), and internet resources(20%) for health-related information. This indicates the need for accessible and reliable health information platforms, like the Web application being developed to assess BMI status and COVID-19 risks. A user-friendly Web application can provide accurate and personalized information, saving time and resources.

The visual results of the survey are shown in Appendix C.

5.4 Critical evaluation of Survey.

The shift from normalcy to the pandemic hindered activities worldwide, especially food accessibility. The survey aims to understand the different reasons behind the transition of COVID-19 and how it affected the lifestyle and food habits of individuals during the transitional period. However, a potential resurgence of less nutritious choices is highlighted from the results with the help of which health experts can design evidence-based strategies to support healthier diets and improve long-term health outcomes. The findings suggest that public health initiatives should focus on promoting healthy and sustainable lifestyle choices tailored to individuals' evolving needs and circumstances. Interventions focusing solely on weight loss may not resonate with everyone, suggesting the need for focused approaches to health that prioritize overall well-being, self-acceptance, and healthy habits. Public health initiatives can adopt inclusive strategies that encourage positive health choices both during and after the pandemic[50]. 'OBC HEALTH' a user-friendly web application, is designed at the end of this survey to empower individuals to take charge of their well-being and seek medical advice when needed.

5.5 System Analysis.

Existing works.

Although web applications have been effectively used to manage chronic illnesses [51], the continuing COVID-19 epidemic has highlighted the need for web app solutions to limit the danger of long-term effects of the COVID-19 and educate individuals about their health metrics for promoting a balanced lifestyle in the long run. In the recent months the web applications for COVID-19 were created to widely attempt to “flatten the curve” of the rising number of COVID-19 cases by offering knowledge and information to individuals about the sysmptoms,causes and long-term effects caused by it.

Most of the previous studies have focused on mobile applications rather than web applications for addressing the COVID-19 situations. One such study shown in [52] Application of mobile health to support the elderly during the COVID-19 outbreak shows the role of mobile application in addressing the COVID-19 issues in elderly people by targeting their physical and mental health and addressing their COVID-19 concerns These included app features such as video consultations, telephone outreach, to reach the older population often in need of communication mental health problems especially during quarantine [52].

Despite the already existing solutions proposed previously as a part of COVID-19 response plans there exist major knowledge gaps exist about educating the individuals about the status of their health metrics in the post-pandemic and how much prone are individuals to be inflicted by the long-term COVID-19 effects along with how they can tailor their diets to suit the post-pandemic situation while transiting from the pandemic times.

Hence to solve this issue we propose a Web application called “OBC HEALTH” that caters to the post-pandemic situations where individuals need to know how their health metric status has changed, what changes they need to make in their diets and how much prone they are to be getting inflicted by the long-term COVID-19 effects in the long run.

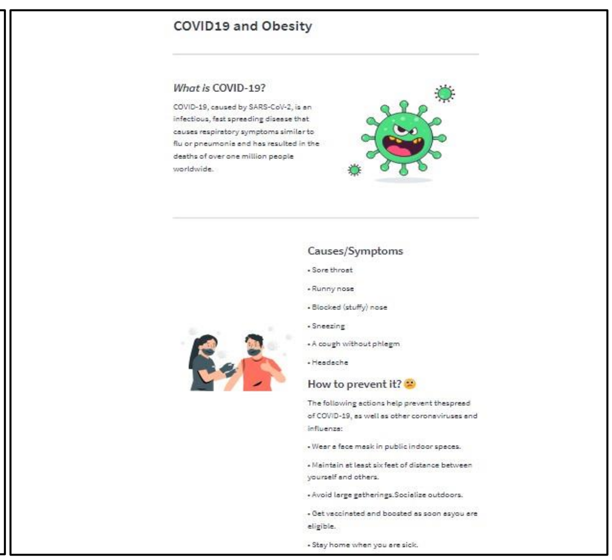
5.6 Web Application Architecture.

Interface Design

The front end of the web application is designed to be responsive and attractive to display with the usage of animations and interactive images. Information is presented in a clear and organized manner. Python is used for developing the web application with the help of streamlit, an open-source framework built for making interactive web interfaces.intratcive features of the Web application including a pre-post survey to assess the food and lifestyle habits of individuals during and post COVID-19,a BMI calculator which use machine learning models at the backend. BMI calculator is made using the XGB classifier model which takes gender, height and weight as input and gives the BMI categories and the BMI score as output. The COVID-19 assessment tool used XGB regressor model and is aimed to analyze the potential score of individuals for long-term effects of COVID-19 on their health. It asks individuals various question on their lifetstyle,diet and curates their potential risk on the scale of 1-20,if the risk is in between 1-10 it indicates that the patient is less likely to get long-term effects of COVID-19 but if the potential score is above 10 the it indicates a higher risk of the individual to develop the long-term effects of COVID-19 and the individual should look into ways to combat it efficiently with the help of recommended diets listed .To evaluate the user validating and testing a feedback form is given to five individuals doing MSc In Artificial intelligence and Computer science form university of Birmingham, Dubai after they have seen the web application in the form of video sent to them by email. Lastly the important World Health organization updates regarding COVID-19, and a contact form is at the end of the web application will help us to collect user information to contact them in future.

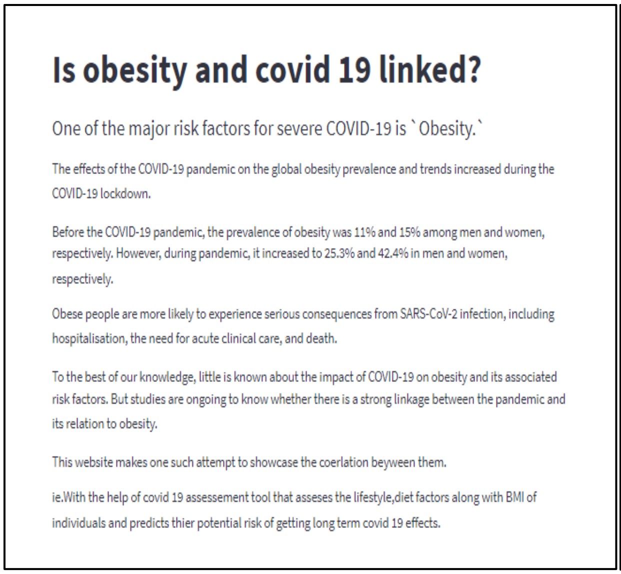
The instructions on how to run the web application 'OBC HEALTH are given in Appendix A.

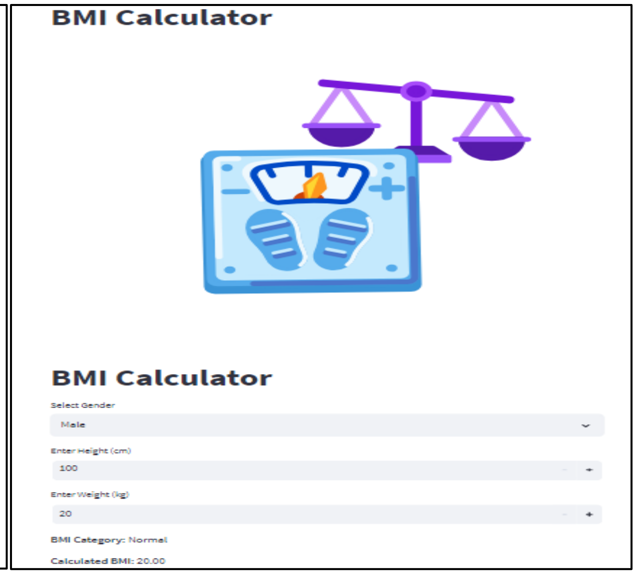
The screenshots of the web application pages are shown below:



Page 1 shows the homepage of the web application.

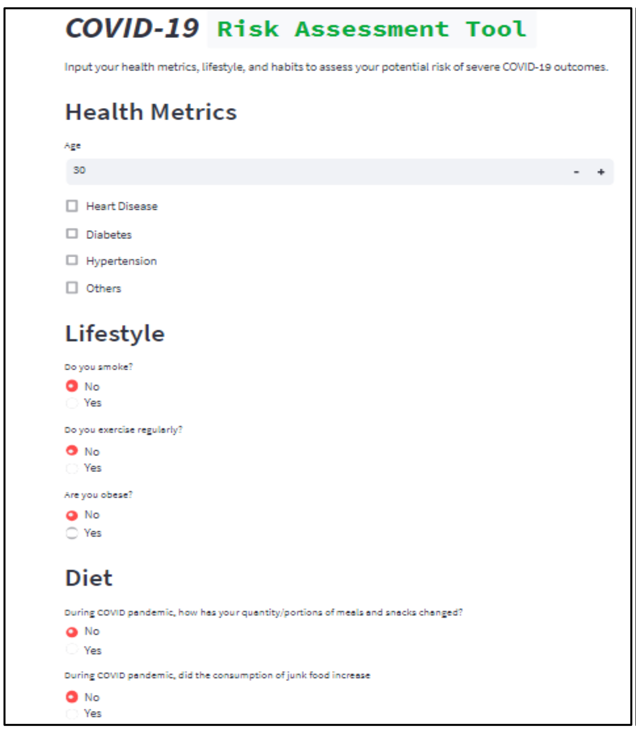
Page 2 explains obesity and COVID 19 in detail.

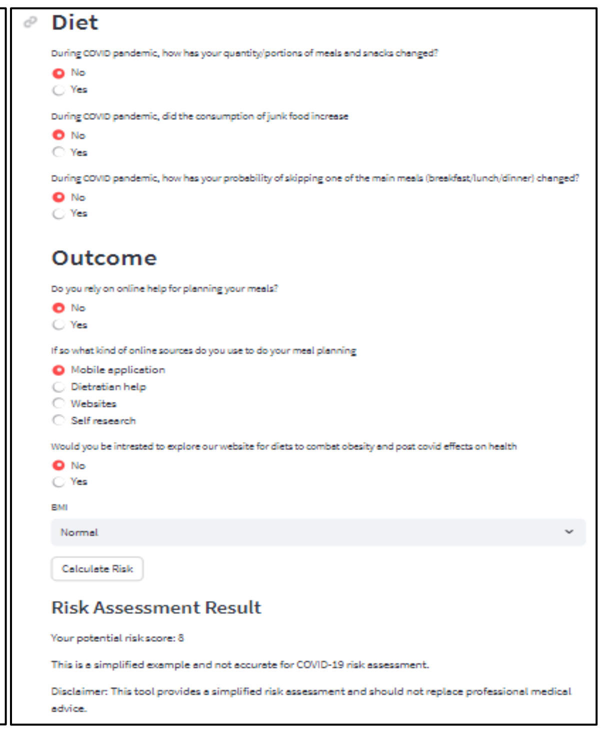




Page 3 talks about the link between COVID 19 and obesity.

Page 4 shows the BMI calculator which calculates the BMI category and the BMI score using XGB Regressor machine learning model at the backend.

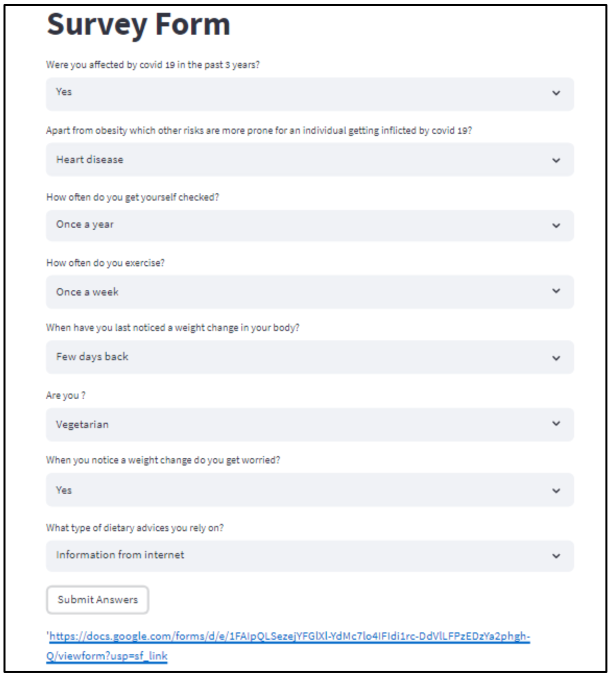




Page 5 shows the risk assessment tool that shows the potential risk for an individual having long term effects of covid 19 in the range of 1-20.(Potential risk score)

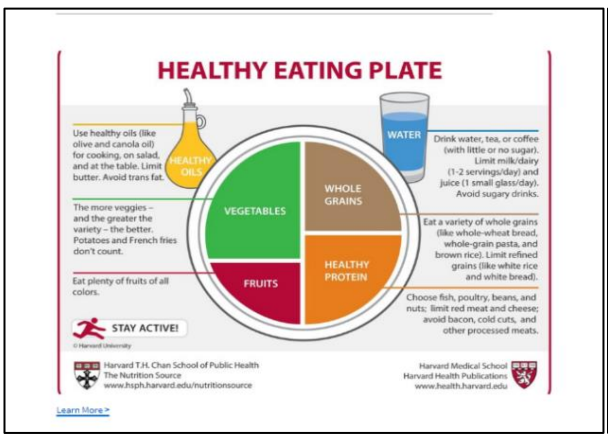
(1-10)-Low risk of long term COVID 19 .

(10-20)-High risk of long term COVID 19 effects.



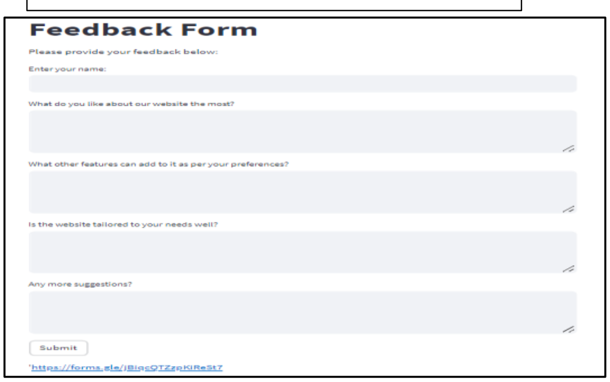
Page 6 shows a survey form to analyze the changes in food and lifestyle habits of individuals while transiting to post-pandemic from COIVD 19.

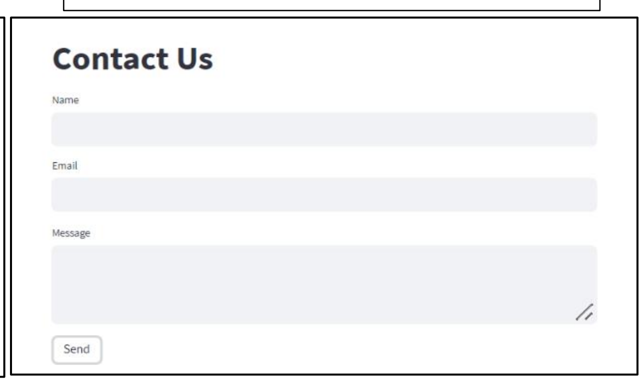
Page 7 shows the recommended diets for people affected by COVID 19 and obesity as well as to those who are transiting in the post-pandemic stage.



Page 8 acts as guidance page to promote healthy food habits and lifestyle amongst individuals.

Page 9 shows the latest news of COVID 19.





Page 10 is used for knowing the web application's performance by its users in the form of feedback form.

Page 11 is used to gather the information of the users that visited our web application.

5.7 Testing.

Web application testing is an important stage in the development of any web-based program. It entails carefully testing a web application's functionality, performance, security, and usability to verify its dependability and user-friendliness. Web application testing tries to find and correct any problems or inadequacies before the application goes live using a combination of human and automated testing methodologies. This thorough testing procedure is critical for providing a seamless and secure user experience while reducing potential risks and preserving the application's integrity.

There are four types of web application testing:

1.Unit testing- Individual units or components of the program, such as functions or methods, are tested independently during unit testing in web application development to verify they execute as intended. It aids in the early detection and correction of problems, hence improving code quality and overall program reliability.

2.Integration testing- In web application development, integration testing evaluates how different components or modules of the program interact with one another. It guarantees that the integrated elements work together as a whole, identifying any problems that may develop when these components are merged and assisting in ensuring a seamless user experience.

3.System testing- System testing for web apps involves evaluating the complete program to ensure that it satisfies the defined requirements and performs as intended in a real-world setting. It entails thoroughly evaluating all integrated components, features, and interactions to find any anomalies or flaws, frequently utilizing test cases that imitate user scenarios.

4.Acceptance testing- Acceptance testing for web apps determines if the application satisfies the expectations and needs of the user or client. End-users or stakeholders generally execute real-world scenarios to ensure that the application operates well and meets their expectations before it is deployed. Black box testing for acceptance in web applications evaluates the application's functioning and usability from the user's perspective without requiring knowledge of its internal code. Testers are responsible for testing user requirements, interface design, and overall user experience to ensure that the program fulfills user expectations.

For ‘OBC HEALTH’ web application we will be using the acceptance testing method to evaluate the application's functionality and usability from the user's point of view.

A comprehensive table of test cases is provided in Appendix D

5.8 Potential User Testing And Validation.

To evaluate the final web application five students from the MSc Artificial intelligence and Computer Science (Conversion) course were shown a five-minute video demonstration of the application.

The video compassed all working functional aspects of the web application, and after showing them all the features of the Web applications. A copy of the feedback form was shared to them via google forms to know their opinions of the working of the web interface. The link for the feedback form is https://forms.gle/zE1sSBmVqhmuv1SW6.

- The responds were impressive. According to the feedback form all the students liked the front-end design of the Web application which was developed using streamlit and streamlit Lottie for animations.

- 2 out of 5 respondents stated that they wanted to see recommended exercises along with the recommended diets' features.

- 3 out of 5 respondents pointed out that the Web application name was very impressive, and it integrates well with the topic i.e., obesity and COVID-19 linked.

- All 5 respondents liked the overall features and were satisfied with them.

- 2 out of 5 respondents stated that the Web application had features like most of the competitor mobile applications in the market. A similar feature they noticed was the BMI calculator. The unique feature which they did not find in any of the competitor applications in the market was the COVID-19 risk assessment tool to predict the risk score of individuals getting inflicted by long-term effects of COVID-19.

- Lastly the users recommended that the web application should include a bot feature which can answer user related queries instantly and for a virtual health coach that would help them find the best suitable diet plans and exercises regimes according to their fitness goals. And a frequently asked questions (FAQS) page for the users helps while navigating through the web application.

The feedback form and the results of the feedback form are shown in Appendix E and Appendix F.

CHAPTER VI

6.1 Future works.

Clearly, we are facing an unpredictable interaction between two pandemics, COVID-19, and obesity. An important limitation of our study is that it only analyzed food categories vs COVID-19 variables ,other variables that may be more closely related to obesity and COVID-19 weren't addressed in detail. Studies on multivariate variables can be done for example genetics,behavioural and psychological factors like mental health can be adressed.This will increase the scope for our study and give us multidimensional insights on combating these global challenges effectively. Secondly the study only focuses on countries on a whole and individuals, it does not take into account societies ,cross countries or cross regional comparison in specific this is a true limitation as there might exist some inequalities in societies such as disproportionately increased risk of the two diseases due to socioeconomic factors in various regions or countries that can affect the lifestyle and diet of the people on a larger scale leading to higher risks of obesity and COVID-19 [53].Hence future studies should focus on all multi-disciplinary aspects as well. Lastly even though BMI is taken into account it doesn't address the change in trends age wise that took place during and shortly after the pandemic which was declared as ended on May 5th 2023 as per The World Health Organization, Future works should focus on how the BMI changed over time in different age groups individually ranging from adolescents to elderly which can help the healthcare professionals with valuable insights to observe obesity trends effectively over time and suggest suitable recommendations. No one will ever forget the years with a legacy from COVID-19 for decades to come. We have had our wake-up call; now it's time for us all to reshape the future with the impetus to develop new obesity and COVID-19 strategies in the future [54].

6.2 Limitations of the paper.

- The datasets used in the paper do not give us individual nutrients intake figures but just country-wise and that is one of the major drawbacks as it gets difficult to suit diets for an individual based on country-wise data.
- The relationship between obesity and confirmed showed poor accuracy, hence the model was extremely overfitted .Regularization methods must be used to reduce the overfitting to some extent.
- The paper only considers obesity on a whole realm, further discussions on how obesity affects the immunity of the individuals is not discussed .This is a limitation as showcasing immunity levels would have given us a clear picture of the heath state of the individuals better. The paper only talks about the COVID-19 variant overall it does not categorize the COVID-19 variants in all its forms. There should be future studies on showcasing how obesity is linked to different types of COVID-19 variants as well.

6.3 Conclusion.

To conclude the interdependency of the obesity epidemic and COVID-19 pandemic has resulted in a complex collision of global health challenges. By determining the impact of obesity in those affected by pandemic worldwide, our aim was to shed light on vital aspects that led to this intervention. The drastic change in dietary habits were observed with the help of datasets for country wise comparison and to accurately assess the obesity of individuals based on gender ,height and weight a BMI dataset was shown.

To predict the relationship between COVID-19 and food categories, linear regression model was used which was overfitting at first but when we trained and regularized it by reducing the loss function of the model, it resulted in promising accuracy .To predict the co-dependency of obesity with COVID-19,an XGB regressor was used in contrast to linear regression as the XGB regressor proved to be more advantageous due to its efficiency in dealing with outliers and capability of dealing with large and noisy data. The results of the XGB regressor showed promising accuracy scores when evaluated. The model had a low RMSE score (0.029) depicting better accuracy and high R^2values(99.98) indicating a better fit for the data on both training and testing datasets. For assessing the obesity classes in the BMI dataset, we have used XGB classifier as it's a classification problem with categorical data. The classification report showed various evaluation scores of the obesity classes and concluded that all the scores shown in the classification report indicated that the data was well fitted in the model and was accurate almost about 90%.

After developing the models, a web application OBC HEALTH was developed which analyzed changes in dietary habits of individuals during and after COVID-19 through a survey and recommended suitable diets .It also provided users with a BMI calculator to evaluate their obesity levels in the form of BMI score and find out in what BMI category they fall into. With the help of COVID-19 risk assessment tools individuals were able to predict their long-term COVID-19 effects risk score based on their answers to the questions mentioned in the tool in the range of 1-20 where 1-10 score was considered as a low risk and 10-20 as a higher risk of having long-term effects of COVID-19.

In future more emphasis will be required on studies that will capture and explore how diet and nutritional status can modify the immune system's response to an individual getting inflicted by long-term COVID-19 effects. This knowledge could help explain some of the variability in the diet and help health practitioners tailor their strategies in promoting a healthy liefetyle.Further research could also identify different groups of obesity and see how each group is susceptible to long-term effects of COVID-19.More information is needed on how the pandemic has changed the life cycles of individuals over time and the need for advanced models to be trained to predict the likeliness of the emergence of another pandemic which will be a combination of obesity along with the after effects of the COVID-19 in the due course of time.

REFERENCES

[1] Shams, M.Y. *et al.* (2021) 'Hana: A Healthy Artificial Nutrition Analysis Model during COVID-19 pandemic', *Computers in Biology and Medicine*, 135, p. 104606. doi:10.1016/j.compbiomed.2021.104606.

[2] *Obesity statistics* (2023) *Forbes*. Available at: https://www.forbes.com/health/body/obesity-statistics/#:~:text=Worldwide%2C%20more%20than%201%20billion,million%20children%2C%20according%20to%20WHO (Accessed: 16 September 2023).

[3] Kristensen, N.M. *et al.* (2022) 'Obesity augments the disease burden in Covid-19: Updated data from an Umbrella Review', *Clinical Obesity*, 12(3). doi:10.1111/cob.12508.

[4] González-Monroy, C. *et al.* (2021) 'Eating behavior changes during the COVID-19 pandemic: A systematic review of longitudinal studies', *International Journal of Environmental Research and Public Health*, 18(21), p. 11130. doi:10.3390/ijerph182111130.

[5] Rangelov, N. and Schlunegger, M. (2022) 'Altea: Improving long covid affected people's quality of life by knowledge and experience sharing', *01.04 - M-Health/e-health* [Preprint]. doi:10.1183/13993003.congress-2022.1824.

[6] Ziegler, S. *et al.* (2021) *Long covid citizen scientists – developing a needs-based research agenda by persons affected by Long Covid* [Preprint]. doi:10.1101/2021.12.08.21267181.

[7] Dicker, D. *et al.* (2020) 'Obesity and Covid-19: The two sides of the coin', *Obesity Facts*, 13(4), pp. 430–438. doi:10.1159/000510005.

[8] Lambert, H. *et al.* (2020) 'Covid-19 as a global challenge: Towards an inclusive and sustainable future', *The Lancet Planetary Health*, 4(8). doi:10.1016/s2542-5196(20)30168-6.

[9] Mohammad, S. *et al.* (2021) 'Obesity and covid-19: What makes obese host so vulnerable?', *Immunity & Ageing*, 18(1). doi:10.1186/s12979-020-00212-x.

[10] *Weekly epidemiological update on COVID-19 - 10 August 2023* (no date) *World Health Organization*. Available at: https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---10-august-2023 (Accessed: 16 September 2023).

_[11] Pandeya, P.R. *et al.* (2021) 'Efficacy of a novel herbal formulation (F2) on the management of obesity: In vitro and in vivo study', *Evidence-Based Complementary and Alternative Medicine*, 2021, pp. 1–14. doi:10.1155/2021/8854915.

[12] Bretón, I. *et al.* (2021) 'Obesity and covid-19. A necessary position statement', *Endocrinología, Diabetes y Nutrición (English ed.)*, 68(8), pp. 573–576. doi:10.1016/j.endien.2021.02.004.

[13] Choi, W.S. and Cheong, H.J. (2021) 'Covid-19 vaccination for people with comorbidities', *Infection & Chemotherapy*, 53(1), p. 155. doi:10.3947/ic.2021.0302.

[14] *Coronavirus on the border: Why Mexico has so few cases compared with the U.S.* (2020) *The Washington Post*. Available at: https://www.washingtonpost.com/world/the_americas/coronavirus-mexico-us-border-covid-19/2020/03/30/18ad79b2-7061-11ea-a156-0048b62cdb51_story.html (Accessed: 16 September 2023).

[15] Palacio-Mejía, L.S. *et al.* (2022) 'Leading causes of excess mortality in Mexico during the COVID-19 pandemic 2020–2021: A death certificates study in a middle-income country', *The Lancet Regional Health - Americas*, 13, p. 100303. doi:10.1016/j.lana.2022.100303.

[16] Weinglass, S. *et al.* (2020) *How did Israel keep its death toll so low, and does it now risk a new spike?, The Times of Israel*. Available at: https://www.timesofisrael.com/qa-how-did-israel-keep-its-death-toll-so-low-and-does-it-now-risk-a-new-spike/ (Accessed: 16 September 2023).

[17] Kristensen, N.M. *et al.* (2022) 'Obesity augments the disease burden in Covid-19: Updated data from an Umbrella Review', *Clinical Obesity*, 12(3). doi:10.1111/cob.12508.

[18] García-Ordás, M.T. *et al.* (2020) 'Evaluation of country dietary habits using machine learning techniques in relation to deaths from covid-19', *Healthcare*, 8(4), p. 371. doi:10.3390/healthcare8040371.

[19] Shams, M.Y. *et al.* (2021) 'Impact of covid-19 pandemic on diet prediction and patient health based on support vector machine', *Advances in Intelligent Systems and Computing*, pp. 64–76. doi:10.1007/978-3-030-69717-4_7.

[20] Naja, F. and Hamadeh, R. (2020) ‘Nutrition amid the COVID-19 pandemic: A multi-level Framework for Action’, *European Journal of Clinical Nutrition*, 74(8), pp. 1117–1121. doi:10.1038/s41430-020-0634-3.

[21] Stella, L. *et al.* (2022) ‘The role of asymptomatic infections in the COVID-19 epidemic via complex networks and stability analysis’, *SIAM Journal on Control and Optimization*, 60(2). doi:10.1137/20m1373335.

[22] Marti Mason, D. *et al.* (2020) ‘Impact of social distancing to mitigate the spread of COVID-19 in a virtual environment’, *26th ACM Symposium on Virtual Reality Software and Technology* [Preprint]. doi:10.1145/3385956.3422093.

[23] Verdhan, V. (2020) ‘Supervised learning for classification problems’, *Supervised Learning with Python*, pp. 117–190. doi:10.1007/978-1-4842-6156-9_3.

[24] . Malla, R.R. and Katneni, V. (2023) ‘Artificial Intelligence in breast cancer: An opportunity for early diagnosis’, *Computational Methods in Drug Discovery and Repurposing for Cancer Therapy*, pp. 73–89. doi:10.1016/b978-0-443-15280-1.00004-2.

[25] Samuels, P. (2014) *Simple Linear Regression* [Preprint].

[26] Shi, J. *et al.* (2023) ‘XGBoost-based multiparameters from dual-energy computed tomography for the differentiation of multiple myeloma of the spine from vertebral osteolytic metastases’, *European Radiology*, 33(7), pp. 4801–4811. doi:10.1007/s00330-023-09404-7.

[27] Jyothsna, C. *et al.* (2022) ‘Health Insurance Premium prediction using XGBoost regressor’, *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)* [Preprint]. doi:10.1109/icaaic53929.2022.9793258.

[28] Hillel, T. *et al.* (2021) ‘A systematic review of machine learning classification methodologies for Modelling passenger mode choice’, *Journal of Choice Modelling*, 38, p. 100221. doi:10.1016/j.jocm.2020.100221.

[29] Yashwanth, N. (2022) *Evaluation of a traditional regression model term selection process for balance calibration data ...* [Preprint]. doi:10.2514/6.2022-2055.vid.

[30] Esaki, T. (2021) ‘Appropriate evaluation measurements for regression models’, *Chem-Bio Informatics Journal*, 21(0), pp. 59–69. doi:10.1273/cbij.21.59.

[31Khan, M.A. *et al.* (2022) ‘Performance evaluation of regression models for COVID-19: A statistical and Predictive Perspective’, *Ain Shams Engineering Journal*, 13(2), p. 101574. doi:10.1016/j.asej.2021.08.016.

[32] Ampountolas, A. *et al.* (2021) ‘A machine learning approach for Micro-Credit scoring’, *Risks*, 9(3), p. 50. doi:10.3390/risks9030050.

[33] Gillespie, C. (2021) ‘Using the least absolute shrinkage and selection operator for variable selection and Model Building’, *Proceedings of the 2021 AERA Annual Meeting* [Preprint]. doi:10.3102/1689101.

[34] Butler, M.J. and Barrientos, R.M. (2020) ‘The impact of nutrition on covid-19 susceptibility and long-term consequences’, *Brain, Behavior, and Immunity*, 87, pp. 53–54. doi:10.1016/j.bbi.2020.04.040.

[35]Tong, T.J., Mohammadnezhad, M. and Alqahtani, N.S. (2022) ‘Determinants of overweight and obesity and preventive strategies in Pacific Countries: A systematic review’, *Global Health Journal*, 6(3), pp. 122–128. doi:10.1016/j.glohj.2022.07.005.

[36] Alnohair, S. (2014) ‘Obesity in gulf countries’, *International Journal of Health Sciences*, 8(1), pp. 79–83. doi:10.12816/0006074.

[37] Cuong, T.Q. *et al.* (2006) ‘Obesity in adults: An emerging problem in urban areas of Ho Chi Minh City, Vietnam’, *European Journal of Clinical Nutrition*, 61(5), pp. 673–681. doi:10.1038/sj.ejcn.1602563.

[38] Bergquist, S., Otten, T. and Sarich, N. (2020) ‘Covid-19 pandemic in the United States’, *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.3657404.

[39] Nkengasong, J. (2020) ‘Let Africa into the market for COVID-19 Diagnostics’, *Nature*, 580(7805), pp. 565–565. doi:10.1038/d41586-020-01265-0.

[40] Kuguyo, O., Kengne, A.P. and Dandara, C. (2020) ‘Singapore covid-19 pandemic response as a successful model framework for low-resource health care settings in Africa?’, *OMICS: A Journal of Integrative Biology*, 24(8), pp. 470–478. doi:10.1089/omi.2020.0077.

[41] Hopman, J., Allegranzi, B. and Mehtar, S. (2020) 'Managing covid-19 in low- and middle-income countries', *JAMA*, 323(16), p. 1549. doi:10.1001/jama.2020.4169.

[42] Gilbert, M. *et al.* (2020) 'Preparedness and vulnerability of African countries against importations of COVID-19: A modelling study', *The Lancet*, 395(10227), pp. 871–877. doi:10.1016/s0140-6736(20)30411-6.

[43] . Chitungo, I. *et al.* (2020) 'Covid-19: Unpacking the low number of cases in Africa', *Public Health in Practice*, 1, p. 100038. doi:10.1016/j.puhip.2020.100038.

[44]Abuza, Z. (2021) *What explains Vietnam's current covid-19 struggles?, – The Diplomat*. Available at: https://thediplomat.com/2021/06/what-explains-vietnams-current-covid-19-struggles/ (Accessed: 16 September 2023).

[45] Arulanandam, B., Beladi, H. and Chakrabarti, A. (2023) 'Obesity and covid-19 mortality are correlated', *Scientific Reports*, 13(1). doi:10.1038/s41598-023-33093-3.

[46] 'Risk prediction of covid-19 related death and hospital admission in adults after covid-19 vaccination: National Prospective Cohort Study' (2021) *BMJ* [Preprint]. doi:10.1136/bmj.n2300.

[47] Akter, T. *et al.* (2022) 'Impact of the COVID-19 pandemic on BMI: Its changes in relation to socio-demographic and physical activity patterns based on a short period', *PLOS ONE*, 17(3). doi:10.1371/journal.pone.0266024.

[48] Yates, T. *et al.* (2022) 'A population-based cohort study of obesity, ethnicity and covid-19 mortality in 12.6 million adults in England', *Nature Communications*, 13(1). doi:10.1038/s41467-022-28248-1.

[49] Marks, R. (2022) 'Covid-19 and childhood obesity', *Advances in Obesity, Weight Management & Control* , 12(2), pp. 45–50. doi:10.15406/aowmc.2022.12.00362.

[50]Kwok, S. *et al.* (2020) 'Obesity: A critical risk factor in the covid-19 pandemic', *Clinical Obesity*, 10(6). doi:10.1111/cob.12403.

[51] Agarwal, P. *et al.* (2021) 'Assessing the quality of mobile applications in Chronic Disease Management: A scoping review', *npj Digital Medicine*, 4(1). doi:10.1038/s41746-021-00410-x.

[52] Kondylakis, H. *et al.* (2020) 'Covid-19 mobile apps: A systematic review of the literature', *Journal of Medical Internet Research*, 22(12). doi:10.2196/23170.

[53Lange, S.J. *et al.* (2021) 'Longitudinal trends in body mass index before and during the COVID-19 pandemic among persons aged 2–19 years — United States, 2018–2020', *MMWR. Morbidity and Mortality Weekly Report*, 70(37), pp. 1278–1283. doi:10.15585/mmwr.mm7037a3.

[54] Zhang, X. *et al.* (2021) 'A systematic review and meta-analysis of obesity and COVID-19 outcomes', *Scientific Reports*, 11(1). doi:10.1038/s41598-021-86694-1.

Appendix A

Gitlab Repository link: https://git.cs.bham.ac.uk/projects-2022-23/sxp362.git

The git repository contains the following folders:

1.OBC HEALTH-This folder contains the webpages of the web application 'OBC HEALTH' along with the codes used for making the web application and the survey and feedback form used for user testing and validation.

2.CODE-This folder contains the jupyter notebook which should be downloaded and viewed in the Jupyter notebook from the browser.

3.README-This file contains all the information to run the application as well as install the required libraries.

If one does not have jupyter notebook installed please install using the link below:

Open the command prompt and type.

pip install notebook.

After pressing 'Enter' and to run the notebook type .

Jupyter notebook

How to run the web application:

Step 1: Download and install python version 3.9 or 3.10 a

Step 2: Download the folder 'OBC HEALTH' from GitLab.

Steps 3:Open command prompt and go to the project directory using cd command (e.g., use command: cd(space) path of the file).

The file name is 'OBC HEALTH' in git repository.

Step 4: Once entered in the project directory, then create Conda environment by running the below command(this will create the virtual environment).

conda create -n projectenv.

Step 5:Now active the Conda environment by running the below command:

conda activate projectenv.

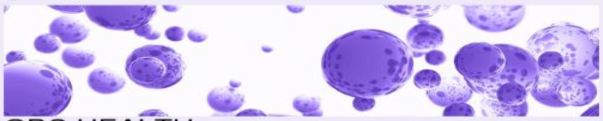
Step 6: Now install the required libraries (these requirement libraries are available as 'requirements.txt' file which is available at 'source code' folder).

conda list-export>requirements.txt

Step 7:Once all the required libraries are installed then run streamlit application using below command (this will open homepage of the application):

Streamlit run 1_😇_Homepage.py

Appendix B



OBC HEALTH

This survey is solely for research purposes and not for any medical use.

A post covid 19 analysis survey for observing changes in dietary patterns and lifestyle of individuals.

Were you affected by covid 19 in the past 3 years?

Yes
No

Did your food habits change drastically due to covid 19?

Yes
No

If you answered yes for the above question, please specify how did they change?

Your answer

Did the covid restrictions hinder your access to food?

Yes
No

Which of the covid restriction were you affected the most with?

Lockdowns
Home quarantine
Travel restrictions
Other:

Apart from obesity which other risks in your opinions is more likely to be linked with covid 19?

Heart diseases
Diabetes
Lung diseases
Hypertension
Other:

How often do you get a body checkup done?

Once a year
Once a month
Twice a year
Other:

How often do you exercise?

Once a week
5 days a week
2-3 days a week
Rarely

When have you last noticed a weight change in your body?

Few days back.
Few months back.
Yearly
Irregular patterns

When you notice a weight change do you get worried?

Yes
No

Are you?

Vegetarian
Non vegetarian
Omnivore
Other:

What type of dietary advice you rely on?

Information from internet
Medical advice from doctors
Self assesement

THANK YOU FOR COMPLETING THE SURVEY:)

Get link

Appendix C

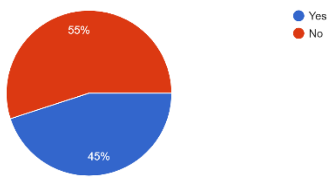
OBC HEALTH

100 responses

Publish analytics

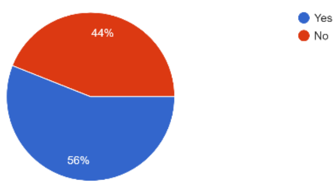
Did your food habits change drastically due to covid 19?

100 responses



Were you affected by covid 19 in the past 3 years?

100 responses



If you answered yes for the above question, please specify how did they change?

53 responses

-

Eating less and focused on more health related foods

Higher intake of processed foods during COVID-19 lockdown.

Tried to incorporate more herbal food items

Became a full time vegetarian

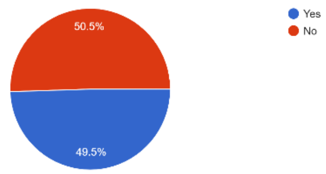
Junk food increawd

I stopped having less of chicken and outside food.

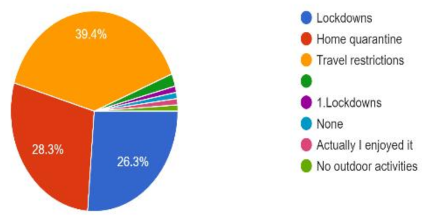
More junk

Started eating more junk

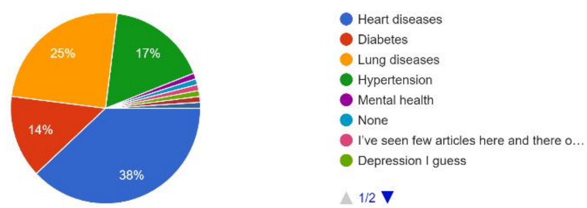
Did the covid restrictions hinder your access to food?
99 responses



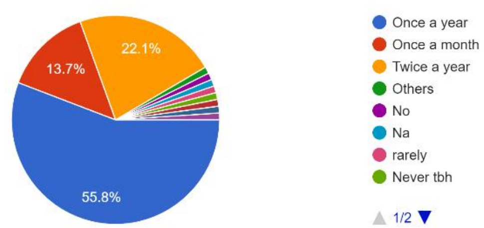
Which of the covid restriction were you affected the most with?
99 responses



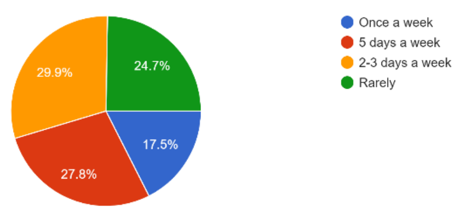
Apart from obesity which other risks in your opinions is more likely to be linked with covid 19?
100 responses



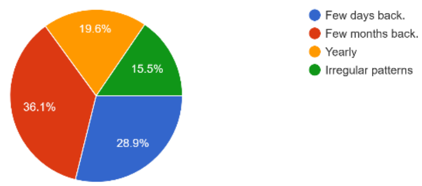
How often do you get a body checkup done?
95 responses



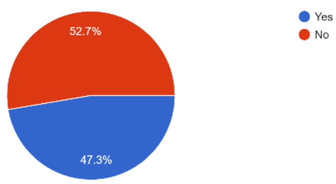
How often do you exercise?
97 responses



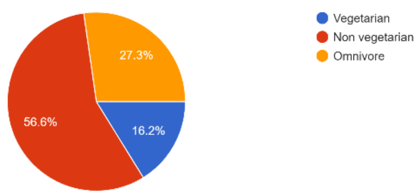
When have you last noticed a weight change in your body?
97 responses



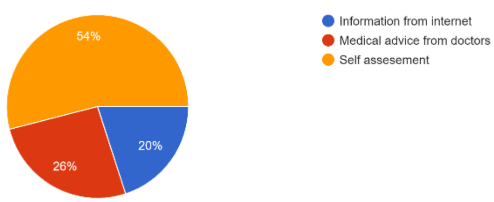
When you notice a weight change do you get worried?
93 responses



Are you?
99 responses



What type of dietary advice you rely on?
100 responses

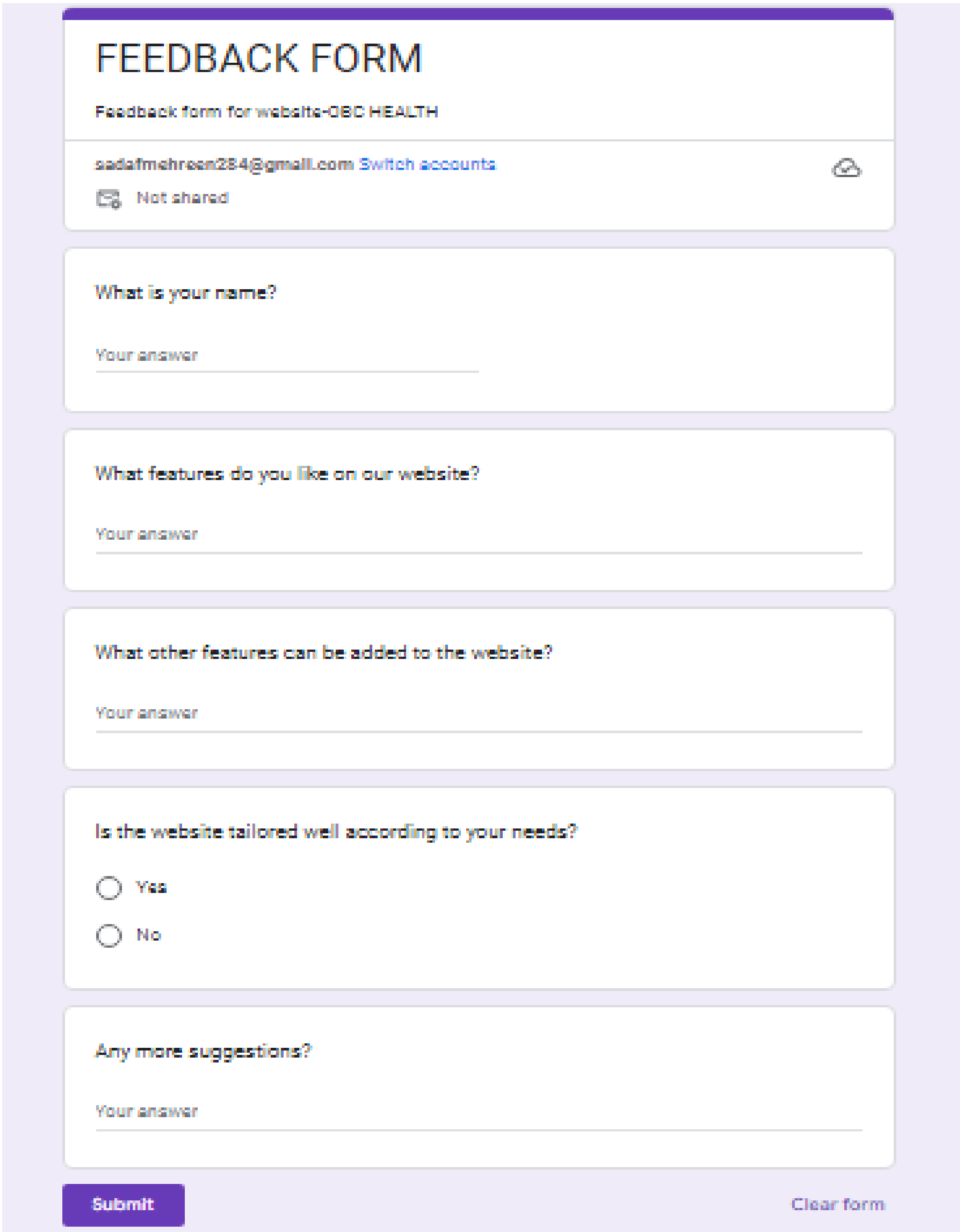


Appendix D

ID	TEST CASE	TEST STEPS	EXPECTED OUTCOME	PASS/FAIL
1.	The user should be able to open the homepage.	1.Open command prompt and change directory to the files containing the web tabs. 2. Open the web application by typing the following in the command prompt: streamlit run 1_😇_Homepage.py	The homepage opens showing the tabs of the web pages at the left side	PASS
2.	The user should be able to navigate through the web pages	1.The user should be able to click tabs and open the pages he/she wants to see in the web application in the vertical drop-down column at the left.	The web page which the user wants to see should open with a click	PASS
3.	The user should get the predicted BMI level after entering the inputs (Height,Weight,Age)	1..Select 'BMI page. Calculator' page. 2.Enter the inputs in the drop boxes 3..User clicks 'Enter.'	The user gets the BMI category and the BMI level based on the users' inputs and by using the following formula. Test status-PASS	PASS
4.	When the user does not enter either height or Weight in the BMI calculator.	1.User enters weight in 'kg.' 2.User forgets to enter Height in 'cm'	The BMI calculator displays the following 2 messages. 1.'Please enter a valid height greater than zero'. 2.'Please enter weight greater than zero'.	PASS
5.	If the BMI calculator is not able to predict the BMI score or user enters wrong values	1.User enters values for height and weight above 100. 2.The BMI calculator shows error.	The BMI calculator shows the following message. "Wrong values entered"	PASS
6.	The user should be shown his/her potential risk score of getting long-term effects of COVID-19.	1.User clicks on the 'Risk assessment tool' page. 2.User answers all questions in the drop boxes including the BMI category they have got from the BMI calculator.	Expected Result-The user gets the potential risk score ranging from 1-20 (1-10) No medical advice needed. (10-20) Medical guidance required.	PASS

7.	Users answer the Survey Form successfully.	1.User enters the information on the page. 2. After completing all the questions Users click on the 'Submit Answer' button.	Expected Result- If submitted successfully the messages appears in the text below as 'Answers submitted carefully.'	PASS
8.	Users answer the Feedback Form questions successfully.	1.Users enter the relevant information in the comment boxes as inputs. 2.After completing all the questions Users click on the 'Submit' button.	Expected Result- If submitted successfully the messages appear in the text below as 'Answers submitted carefully'.	PASS
9.	Users skip a question while answering the Survey (Verifying unsuccessful submission).	1. Users enter the relevant information in the comment boxes provided as inputs. 2.The User skips one question and clicks on the 'Submit' button.	Expected Result- If submitted unsuccessfully the messages appear in the text below as 'Please enter your name and answer at least one question before submitting'.	PASS
10.	User enters the contact information.	1.User enters the relevant contact information such as Name, Email, and message. 2.After entering the details in the comment boxes the users click 'Send.'	Expected Result- After submitting the user gets the message 'Message sent successfully!' .	PASS

Appendix E



FEEDBACK FORM

Feedback form for website-OBC HEALTH

sadafmehreen284@gmail.com Switch accounts

Not shared

What is your name?

Your answer

What features do you like on our website?

Your answer

What other features can be added to the website?

Your answer

Is the website tailored well according to your needs?

- Yes
- No

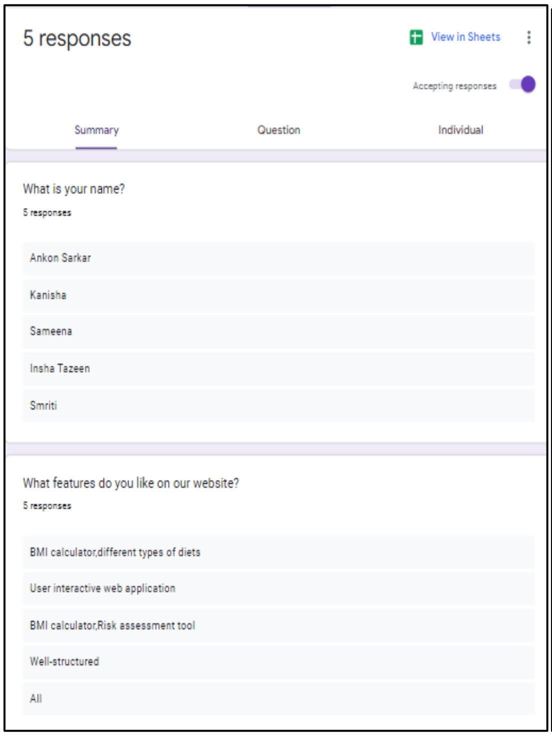
Any more suggestions?

Your answer

Submit

Clear form

Appendix F



5 responses

View in Sheets

Accepting responses

Summary | Question | Individual

What is your name?

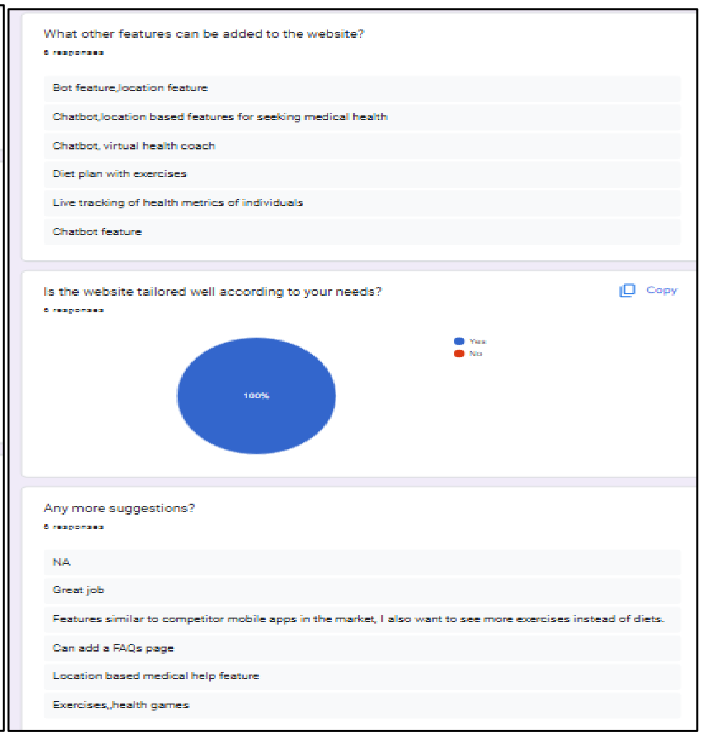
5 responses

- Ankon Sarkar
- Kanisha
- Sameena
- Insha Tazeen
- Smriti

What features do you like on our website?

5 responses

- BMI calculator,different types of diets
- User interactive web application
- BMI calculator,Risk assessment tool
- Well-structured
- All



What other features can be added to the website?

- Bot feature,location feature
- Chatbot,location based features for seeking medical health
- Chatbot, virtual health coach
- Diet plan with exercises
- Live tracking of health metrics of individuals
- Chatbot feature

Is the website tailored well according to your needs?

Copy

Yes

No

100%

Any more suggestions?

- NA
- Great job
- Features similar to competitor mobile apps in the market, I also want to see more exercises instead of diets.
- Can add a FAQs page
- Location based medical help feature
- Exercises,,health games

Appendix H

Fig No.	Fig Description.	Source
1	COVID-19 cases as per World Health Organization(WHO) Region wise as on 6th August 2023.	*Weekly epidemiological update on COVID-19 - 10 August 2023* [10]
2	Shows the Pearson correlation coefficient between obesity, confirmed cases and deaths due to COVID-19	Taken from codes file. Accessible in GitLab: https://git.cs.bham.ac.uk/projects-2022-23/sxp362.git
3	Shows the Mortality and confirmed rates in Mexico and Israel	Taken from codes file. Accessible in GitLab: https://git.cs.bham.ac.uk/projects-2022-23/sxp362.git
4	Explains different kinds of regressions and their properties in detail.	[25[Samuels, P. (2014) *Simple Linear Regression* [Preprint].
5	Shows the confusion matrix for the classification report.	[32] *'A machine learning approach for Micro-Credit scoring'*
6	Training the model with all variables of the COVID-19 Healthy Diet Dataset and confirmed cases of COVID-19.	Taken from codes file. Accessible in GitLab: https://git.cs.bham.ac.uk/projects-2022-23/sxp362.git
7	Shows the performance of the linear regression model.	Taken from codes file. Accessible in GitLab: https://git.cs.bham.ac.uk/projects-2022-23/sxp362.git
8	Shows the method to reduce the overfitting of the model.	Taken from codes file. Accessible in GitLab: https://git.cs.bham.ac.uk/projects-2022-23/sxp362.git
9	Which shows the variables in the dataset that have a positive and a negative correlation w9th the target variable ('Confirmed') in the dataset.	Taken from codes file. Accessible in GitLab: https://git.cs.bham.ac.uk/projects-2022-23/sxp362.git
10	Splitting and training the model based on obesity and confirmed variables.	Taken from codes file. Accessible in GitLab: https://git.cs.bham.ac.uk/projects-2022-23/sxp362.git
11	Evaluates the linear model for obesity and confirmed variables on training and testing sets.	Taken from codes file. Accessible in GitLab: https://git.cs.bham.ac.uk/projects-2022-23/sxp362.git
12	Shows the XGB regressor model.	Taken from codes file. Accessible in GitLab: https://git.cs.bham.ac.uk/projects-2022-23/sxp362.git
13	Shows the classification report.	Taken from codes file. Accessible in GitLab: https://git.cs.bham.ac.uk/projects-2022-23/sxp362.git
14	Shows the Obesity in top 5 countries from the COVID-19 Healthy Diet dataset and shows the consumption of animal and vegetable products in those countries.	Taken from codes file. Accessible in GitLab: https://git.cs.bham.ac.uk/projects-2022-23/sxp362.git

15	Shows the top 5 countries in which obesity is low from the COVID-19 Healthy Diet dataset and shows the consumption of animal and vegetable products in those countries.	Taken from codes file. Accessible in GitLab: https://git.cs.bham.ac.uk/projects-2022-23/sxp362.git
16	Shows the Highest confirmed COVID19 cases in top 5 countries from the COVID-19 Healthy Diet dataset and shows the consumption of animal and vegetable products in those countries.	Taken from codes file. Accessible in GitLab: https://git.cs.bham.ac.uk/projects-2022-23/sxp362.git
17	Shows the Lowest confirmed in top 5 countries from the COVID-19 Healthy Diet dataset and shows the consumption of animal and vegetable products in those countries.	Taken from codes file. Accessible in GitLab: https://git.cs.bham.ac.uk/projects-2022-23/sxp362.git