DeepTrack: An ML-based Approach to Health Disparity Identification and Determinant Tracking for Improving Pandemic Health Care

Jinwei Liu*, Long Cheng†, Ankur Sarker‡, Li Yan§, Richard A. Alo*
*Department of Computer and Information Sciences, Florida A&M University, Tallahassee, FL 32307, USA
†School of Control and Computer Engineering, North China Electric Power University, Beijing, China
‡Department of Electrical Engineering, University of California, Los Angeles, Los Angeles, CA 90095, USA
§Department of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China
*{jinwei.liu, richard.alo}@famu.edu, †lcheng@ncepu.edu.cn, ‡ankursarker@g.ucla.edu, §li.yan.88@xjtu.edu.cn

2021 IEEE International Conference on Big Data (Big Data) | 978-1-6654-3902-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/BigData52589.2021.9671860

***Abstract*—The Coronavirus disease 2019 (COVID-19) pandemic has severely impacted countries around the world with unprecedented mortality and economic devastation and has disproportionately and negatively impacted different communities—especially racial and ethnic minorities who are at a particular disadvantage. Black Americans have a long-standing history of disadvantage (e.g., long-standing disparities in health outcomes) and are in a vulnerable position to experience the impact of this pandemic. Some studies indicate high-risk and vulnerability of the elderly and patients with underlying co-morbidities, however, little research paid attention to leveraging geographic information and machine learning (ML) to track the social and structural health determinants, which can provide a lower level of granularity. In this paper, we propose DeepTrack, a geospatial and ML-based approach to identify diverse determinants (including the structural, social, and constructural determinants) of health disparities in COVID-19 pandemic, which provides a lower level of granularity. We provide a thorough analysis of health disparities and diets based on multiple COVID-19 datasets and examine the structural, social, and constructural health determinants to assist in ascertaining why disparities (in racial and ethnic minorities who are particularly disadvantaged) occur in infection and death rates due to COVID-19 pandemic. We track determinants of nutrition and obesity through diet examination. Extensive experimental results show the effectiveness of our approach. The research provides new strategies for health disparity identification and determinant tracking with a goal to improve pandemic health care.**

***Index Terms*—COVID-19, health determinants, disparities, pandemic, diets, nutrition, obesity, machine learning**

I. Introduction

The novel Coronavirus disease 2019 (COVID-19) was declared a pandemic by the World Health Organization (WHO) on March 11, 2020 [1], and it has severely impacted and devastated the world. Globally, as of September 1, 2021, the virus has resulted in 218,946,836 confirmed cases and 4,539,723 deaths [2]. COVID-19 pandemic has disproportionally and negatively impacted people. On April 8, 2020, the Center for Disease Control and Prevention (CDC) published surveillance data of laboratory-confirmed COVID-19-associated hospitalizations in 14 US states [3]. Among those with data on race/ethnicity ($n = 580$), African Americans account for 33.1%, although 18% of individuals in the catchment population were African American [3]. The government statistics from the cities in the US show similar disparities (racial disparities). African Americans in Chicago account for only 14.6% of Illinois' population, but as of April 9, 2020, 51.5% of COVID-positive patients and 67.3% ($n = 132$) of those who died were African American [4]. According to the 2019 US census, African Americans in Louisiana only account for 33% of the state's population. However, this community accounts for 55% of the COVID-19-related deaths [5]. In Michigan, black individuals account for 33% of the confirmed COVID-19 cases and 40% of attributed deaths despite making up only 14% of the state's population [6]. The states such as North Carolina, Alabama, and the cities St Louis and New York are other examples of disparities in COVID-19-related deaths and ethnicity. African Americans reported having higher death rates than Caucasians [5], [7]–[10]. Racial and ethnic minorities, such as Black/African Americans, are more likely to be the potential target of COVID-19 infection.

A healthy diet is essential to health and depends on one's access to affordable, healthy foods. Poor diet is the leading underlying cause of death, having surpassed tobacco use in related mortality in the US [11]. Around 11% of US households are affected by food insecurity, which is more common in Black, Latinx, and Native Americans [12]. These individuals may predominantly have access to low-cost, energy-dense processed foods. The disparities in nutrition are driven by socioeconomic, educational, and environmental disadvantages that have historically beset vulnerable communities and that persist today. The health disparities in nutrition and obesity are closely related to the alarming racial and ethnic disparities related to COVID-19 [13], [14].

Some recent studies indicate the high-risk groups (e.g., Black/African Americans) and the vulnerability of the elderly and patients with underlying co-morbidities of COVID-19. However, little research paid attention to tracking the social and structural determinants of health disparities in the COVID-19 pandemic for improving health care.

To address the problem, we propose DeepTrack to identify health disparities and track the diverse determinants (social, structural, and constructural determinants) of health disparities in the COVID-19 pandemic and experimentally examine diverse determinants for improving health care. We first provide a thorough analysis of health disparities based on multiple COVID-19 datasets; then, we examine the social, structural, and constructural determinants of health disparities in COVID-19 pandemic for improving health care. We summarize the contribution of this work below.

- We provide a thorough analysis of health disparities based on multiple COVID-19 datasets for deeply examining determinants of health disparities in COVID-19, and the analysis results confirm our conjecture.
- We propose DeepTrack, a geospatial and ML-based approach to identify diverse determinants (including the structural, social, and constructural health determinants) of health disparities in COVID-19 pandemic, which can help to improve pandemic health care.
- We provide an in-depth analysis of diet based on diet dataset of COVID-19. We analyze correlations between: diet and health disparities (in terms of infection and death rates). DeepTrack identifies the diet-related and obesity-related determinants (e.g., different food types and obesity) that contribute to the infection and death rates.

II. Related Work

COVID-19 pandemic has disproportionately impacted people's health. Some recent studies reveal the health disparities in COVID-19. Laurencin *et al.* [15] presented the earliest available data in the peer-reviewed literature on the racial and ethnic distribution of COVID-19-confirmed cases and fatalities in the state of Connecticut. They sought to explode the myth of Black immunity to the virus. Chowkwanyun *et al.* [16] contextualized the COVID-19 data incorporating the demographic detail, and provided an analysis of racial health disparities in COVID-19. Hooper *et al.* [17] indicated the health disparities and provided some explanation about the causes of health disparities. Gray *et al.* [18] examined social determinants of health and health outcomes during COVID-19, and they indicated that upstream social determinants of health (SDOH) are the root causes of health disparities at the population level. Finally, they provided three key strategies for achieving health equity in the US. However, the above studies do not provide a thorough analysis of health disparities and experimentally examine diverse determinants (social, structural and constructural determinants) of health disparities in COVID-19.

To better show health disparities, some recent studies show the health disparity by providing some data analysis. Azar *et al.* [19] conducted a retrospective cohort analysis of COVID-19 patients at Sutter Health, a large integrated health system in northern California, to measure potential disparities. They observed that compared with non-Hispanic white patients, non-Hispanic African American patients had 2.7 times the odds of hospitalization, after adjustment for age, sex, comorbidities, and income. They explored possible explanations for the observation. Selden *et al.* [20] examined historical patterns in health risk, job characteristics, and household composition in search of potential explanations for the large disparities being reported in COVID-19 outcomes. However, their data reflect employment before the onset of COVID-19, and their study therefore provides no insights into whether there have been racial/ethnic dimensions to the COVID-19 employment changes. Also, the above studies do not provide an in-depth analysis of health disparities and experimentally examine diverse determinants (social, structural, and constructural determinants) of health disparities in COVID-19.

Motivated by the problems in the existing literature, we propose an ML-based approach to identify the diverse determinants of health disparities in COVID-19 pandemic for improving health care. We provide a thorough analysis of health disparities based on multiple COVID-19 datasets for deeply examining determinants of health disparities in COVID-19.

III. Data Description

In this section, we describe multiple representative COVID-19 datasets.

A. Dataset 1

We collected the data from KFF [21], and the data were extracted from state websites on July 6, 2020. The dataset records the percent of cases and/or deaths of each race/ethnicity category in each state, and it also records the percent of population of each race/ethnicity category in each state. The total state population distribution by race/ethnicity is based on KFF analysis of 2018 American Community Survey. Percent of cases and/or deaths in each state may not sum to 100% due to rounding. Percentages are based on a total of all probable and confirmed cases and deaths for which race/ethnicity is known.

B. Dataset 2

We also collected the representative data for our experimental analysis from USAFacts (https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/). The dataset records the US COVID-19 cases and deaths by state. Specifically, it records the daily COVID-19 cases and deaths in each county of each state in the US from January 22, 2020 to April 12, 2021. The dataset also records the population of each county in each state in the US. The county-level tracker makes it easy to follow COVID-19 cases on a granular level, as does the ability to break down infections per 100,000 people. In this dataset, we focused on all counties in Florida. In the experiments, we extracted the daily COVID-19 cases and deaths and the population in each county of Florida from January 22, 2020 to April 12, 2021.

C. Dataset 3

To trace the determinants of health disparities, we also collected the data from the US Census Bureau and measured the household experience, distribution of communities, age periods and gender in all counties of the state Florida during the COVID-19 pandemic. The dataset records the experiences of individuals in terms of income & poverty, employment status, spending patterns, housing, physical health, economy accommodation and food services, transportation, access to health care, computer and Internet, and education during the COVID-19 pandemic.

D. Dataset 4

To further trace the determinants of health disparities, we also collected the data from the Imperial College London YouGov Covid 19 Behavior Tracker Data Hub (https://github.com/YouGov-Data/covid-19-tracker) [22]. The questions in the dataset, led by the Institute of Global Health

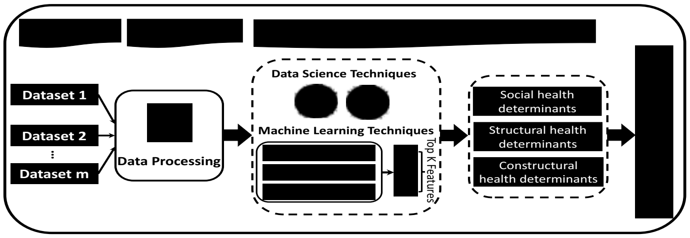


Fig. 1: The framework of DeepTrack.

Innovation (IGHI), cover data on testing, symptoms, self-isolating in response to symptoms and the ability and willingness to self-isolate if needed. It also looks at behaviors, including going outdoors, working outside the home, contact with others, hand washing and the extent of compliance with 20 common preventative measures. Contextual data includes: gender, age, region (within a country), number of people in the household, children in household, health conditions, working status and the date of the survey response. In the dataset, we focused on the data for the United States. We used the attribute i3_health to identify the people who tested positive for COVID-19.

E. Dataset 5

To learn more about how a healthy eating style could help combat the Corona Virus, we also collected diet data of COVID-19 from Kaggle (https://www.kaggle.com/mariaren/covid19-healthy-diet-dataset) for our experimental analysis. The dataset records different types of food, world population obesity and undernourished rate, and global COVID-19 cases count from around the world. The data for different food group supply quantities, nutrition values, obesity, and undernourished percentages are obtained from Food and Agriculture Organization of the United Nations FAO website. The data for population count for each country comes from Population Reference Bureau PRB website. The data for COVID-19 confirmed, deaths, recovered and active cases are obtained from Johns Hopkins Center for Systems Science and Engineering CSSE website. The USDA Center for Nutrition Policy and Promotion diet intake guideline information can be found in ChooseMyPlate.gov.

IV. Methods

A. Feature Selection Methods

To identify diverse determinants (e.g., social vulnerability index measures), we propose DeepTrack, a geospatial and ML-based approach, which leverages techniques of data science and machine learning (e.g., different types of feature selection methods). Below, we introduce the feature selection methods.

Filter Method of Attribute Selection: Filter-based feature selection methods use statistical measures to score the correlation or dependence between input variables that can be filtered to choose the most relevant features. We used ClassifierAttributeEval, a Weka implementation of Filter Method of Attribute Selection. ClassifierAttributeEval measures the significance of an attribute using a specified classifier.

Correlation-Based Feature Selection: Correlation-Based Feature Selection (CFS) ranks the feature subset according to the correlation with the class label and other features. The function evaluates subsets made of attribute vectors, which are correlated with the class label, but independent of each other. Subsets that show high correlation with the class label and less correlation with other features will be ranked a higher value. This method ignores irrelevant and redundant features from the dataset [23], [24]. We used the Weka implementation CorrelationAttributeEval to identify top features contributing to COVID-19 incidence and mortality.

Instance-Based Filter for Feature Selection: Instance Based Filters conduct search not only in the feature space but also in the Instance Space. The instance-based search aims to find the closest decision boundary to the instance under consideration and assign weight to the features that bring about the change. Relief algorithm is an Instance Based Filter [25]. In this paper, we used ReliefFAttributeEval which is instance-based: it samples instances randomly and checks nearby instances of the same and different classes.

B. The Design of DeepTrack

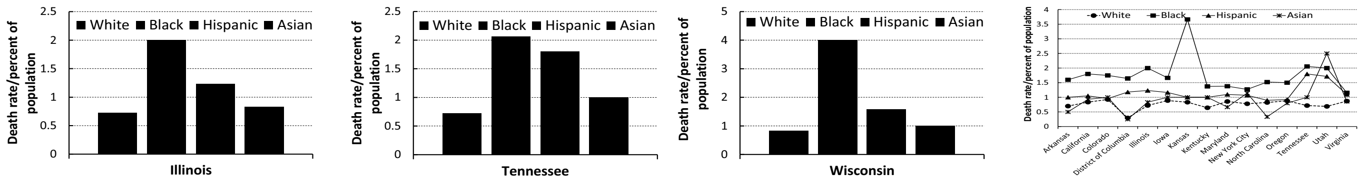
Figure 1 shows the framework of the DeepTrack. DeepTrack consists of three phases: data collection, data processing, determinant tracking. In data collection phase, DeepTrack collects the data with daily COVID-19 cases and deaths and population in each county, the data with household experience, distribution of communities, age periods, gender and the experiences of individuals in terms of income & poverty, etc., and the data with testing, symptoms, self-isolating in response to symptoms and behaviors including going outdoors, working outside the home, hand washing, etc. In data processing phase, DeepTrack processes the data and extracts the information that is needed for determinant tracking. In determinant tracking phase, DeepTrack uses the techniques of data science and machine learning (e.g., feature selection) to extract features and identify health disparities and determinants.

V. Experiments and Findings

In this section, we provide the experimental results based on the multiple datasets, and track the diverse determinants of health disparities.

A. Experimental Analysis Based on Dataset 1

Figure 2(a) shows the ratio of death rate and percent of population of different communities in Illinois. In Figure 2(a), we see that racial and ethnic minority such as Black has the highest ratio of death rate and percent of population among all the communities, followed by Hispanic and Asian, and the White community has the lowest ratio of death rate and percent of population among all the communities. Moreover, the ratio of death rate and percent of population in Black is much higher than that of White. Figure 2(b) shows the ratio of death rate and percent of population of different communities in Tennessee. In Figure 2(b), we also see that racial and ethnic minority such as Black has the highest ratio of death rate and percent of population among all the communities, followed by Hispanic, Asian and White. Moreover, the ratio of death rate and percent of population in Black is much higher than that of White. Figure 2(c) shows the ratio of death rate and



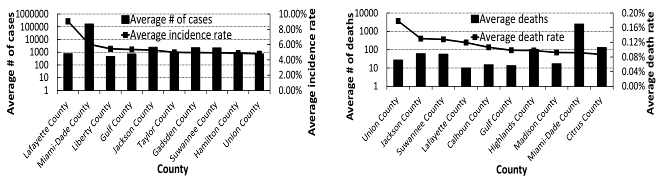
(a) Ratio of death rate and percent of population in Illinois

(b) Ratio of death rate and percent of population in Tennessee

(c) Ratio of death rate and percent of population in Wisconsin

(d) Ratio of death rate and percent of population in different states/cities

Fig. 2: COVID-19 deaths in different communities across different states/cities.



(a) Average number of cases and average incidence rate

(b) Average number of deaths and average death rate

Fig. 3: Top 10 counties with the highest average incidence rate and average death rate in Florida.

percent of population of different communities in Wisconsin. In Figure 2(c), we also find similar results. The results in these figures indicate that racial and ethnic minorities have a higher chance of being killed by COVID-19. Figure 2(d) shows the ratio of death rate and percent of population of different communities in different states/cities. In Figure 2(d), we see that racial and ethnic minority such as Black in general has the highest ratio of death rate and percent of population among all the communities, and the racial and ethnic minority Hispanic in general has the second highest ratio of death rate and percent of population. The result in Figure 2(d) further verifies that racial and ethnic minorities have a higher chance of being killed by COVID-19. The results in all these figures confirm our conjecture: racial and ethnic minorities are at a particular disadvantage, and they are more likely to be the potential target of COVID-19 infection due to the health disparities.

B. Experimental Analysis Based on Dataset 2 and Dataset 3

Figure 3(a) shows the average number of cases and average incidence rate in top 10 counties with the highest average incidence rate in Florida. In Figure 3(a), we see that the average number of cases follows Miami-Dade County > Jackson County > Gadsden County > Suwannee County > Taylor County > Lafayette County > Union County > Gulf County > Hamilton County > Liberty County, and the average incidence rate follows Lafayette County > Miami-Dade County > Liberty County > Gulf County > Jackson County > Taylor County > Gadsden County > Suwannee County > Hamilton County > Union County. Figure 3(b) shows the average number of deaths and average death rate in top 10 counties with the highest average death rate in Florida. Similarly, we can see the rank of each county by average number of deaths and the rank of each county by average death rate.

We also calculate the Pearson correlation between the percent of population of each race/ethnicity category and average incidence/death rate in Florida. Table I shows the correlation between the percent of population of each race/ethnicity category and average incidence/death rate. In Table I, we see

TABLE I: Summary of correlation b/w the percent of population of each race/ethnicity category and average incidence/death rate in Florida.

	White alone	Black or African American alone	American Indian & Alaska Native alone	Native Hawaiian & Other Pacific Islander alone
Incidence Rate	**-0.31591**	**0.3828**	0.1571	**0.2099**
Death Rate	**-0.2126**	**0.2271**	0.0488	**0.5507**

Note: Bold numbers indicate relatively higher absolute values of correlation.

that the percent of population of "White alone" is negatively correlated with both average incidence rate and average death rate in Florida. "Black or African American alone", "American Indian & Alaska Native alone" and "Native Hawaiian & Other Pacific Islander alone" are positively correlated with both average incidence rate and average death rate in Florida. The results indicate that racial and ethnic minorities such as Black or African Americans have a higher rate of infections and deaths and they are more susceptible to COVID-19.

To analyze the relationship between factors such as social vulnerability index measures and COVID-19 incidence and mortality, we calculate the Pearson correlation between each factor and incidence rate and mortality rate (death rate).

Table II shows the correlation between each factor and incidence rate and mortality rate, respectively based on the Florida dataset. In Table II, we see that the absolute value of the Pearson correlation between social vulnerability index measures ('Percent of persons w/ bachelor's degree or higher (25+ yrs)', 'Median household income', 'Per capita income in past 12 mths', 'Persons in poverty', 'Median value of owner-occupied housing unit rate', 'Median selected monthly owner costs-with a mortgage', 'Median selected monthly owner costs-w/o a mortgage', 'Median gross rent', 'Building permits', 'Households w/a computer', 'Households w/ a broadband Internet subscription') and incidence rate is relatively higher than the Pearson correlation between other social vulnerability index measures and incidence rate. We also see that the absolute value of the Pearson correlation between social vulnerability index measures ('With a disability (< 65 yrs)', 'Percent of persons w/ bachelor's degree or higher (25+ yrs)', 'Median household income', 'Per capita income in past 12 mths', 'Persons in poverty', 'Median value of owner-occupied housing unit rate', 'Median selected monthly owner costs-with a mortgage', 'Median selected monthly owner costs-w/o a mortgage', 'Median gross rent', 'Building permits', 'Households w/ a

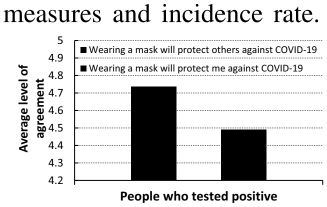
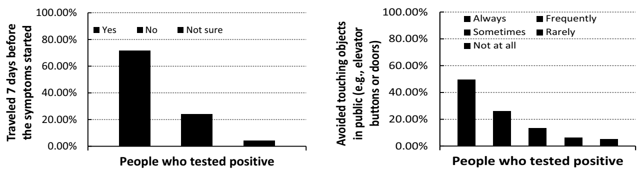


Fig. 4: Ave. level of agreement on mask-wearing policy based on the people who tested positive in the last seven days

TABLE II: Summary of correlation between each factor and incidence rate and death rate in Florida.

	With a disability (< 65 yrs)	Persons W/o health insurance (<65)	Percent of persons w/ bachelor's degree or higher (25+ yrs)	Median household income	Per capita income in past 12 mths	Persons in poverty
Incidence Rate	0.1691	0.1135	**-0.4480**	**-0.3646**	**-0.5214**	**0.5663**
Death Rate	**0.2405**	0.1170	**-0.4609**	**-0.4003**	**-0.4100**	**0.2912**
	Housing units	Median value of owner-occupied housing unit rate	Median selected monthly owner costs-with a mortgage	Median selected monthly owner costs-w/o a mortgage	Median gross rent	Building permits
Incidence Rate	-0.1102	**-0.3630**	**-0.2713**	**-0.3370**	**-0.4819**	**-0.2466**
Death Rate	-0.1300	**-0.3832**	**-0.3900**	**-0.3100**	**-0.4500**	**-0.2785**
	Total employment	Total employer establishments	Households w/ a computer	Households w/ a broadband Internet subscription	Mean travel time to work (16+ yrs)	
Incidence Rate	-0.0228	-0.0084	**-0.4817**	**-0.4333**	0.0428	
Death Rate	-0.1225	-0.0827	**-0.4112**	**-0.3081**	-0.0200	

Note: A bold number indicates relatively higher correlation.



(a) Whether a person traveled to a location where coronavirus had been reported 7 days before the person's symptoms started

(b) Avoided touching objects in public (e.g. elevator buttons or doors)

Fig. 5: Status of traveling and touching objects in public (e.g. elevator buttons or doors).

computer', 'Households w/ a broadband Internet subscription') and death rate is relatively higher than the Pearson correlation between other social vulnerability index measures and death rate. This indicates that the social vulnerability index measures (the factors with bold numbers for both Incidence Rate and Death Rate in Table II) have a relatively higher impact on COVID-19 incidence rate and mortality rate.

C. Experimental Analysis Based on Dataset 4

1) Examining People's Behaviors in Response to COVID-19: The underlying causes of health disparities are complex and include social and structural determinants of health, racism and discrimination, economic and educational disadvantages, health care access and quality, individual behavior, and biology [17]. A simulation study of agent-based influenza showed that small changes in behavior can have a significant effect on transmission patterns during epidemics [26], [27]. People's behaviours could result in non-communicable diseases such as diabetes, hypertension, etc. when they engage in health-risking behaviour (e.g., smoking, substance abuse, not exercising or not eating correctly). People's behavior can indicate and affect people's health [26], [28], [29]. Below we focus on people's behaviors in response to COVID-19 and examine the determinants of health disparities in COVID-19 pandemic.

Figure 4 shows the average level of agreement on mask-wearing policy based on the people who tested positive (for COVID-19) in the last seven days. The level ranges from 1 to 7. In Figure 4, we see that both average levels of agreement are higher than 4.425, suggesting that people believe that wearing a mask will protect them and others against COVID-19.

Figure 5(a) shows whether a person traveled to a location where coronavirus had been reported 7 days before the person's symptoms started. In Figure 5(a), we see that over 71% of the people who tested positive traveled to a location where coronavirus had been reported, which suggests that traveling can facilitate the spread of COVID-19 and increase the probability of a person being infected with COVID-19, and travel restrictions can be useful for preventing the spread of COVID-19. Figure 5(b) shows how frequently people who tested positive avoid touching objects in public (e.g. elevator buttons or doors). In Figure5(b), we see that around 50% of people always avoid touching objects in public (e.g. elevator buttons or doors), over 26% of the people frequently avoid touching objects in public (e.g. elevator buttons or doors). This indicates that most of the people believe that touching objects in public can increase the risk of being infected with COVID-19. Therefore avoiding touching objects in public (e.g. elevator buttons or doors) can help to prevent the spread of COVID-19.

TABLE III: Summary of correlation between Fruits (or Vegetables) Intake and COVID-19 Incidences (or Deaths)

	Fruits Intake	Vegetables Intake
COVID-19 Incidences	-0.028	**-0.214**
COVID-19 Deaths	-0.035	**-0.274**

Note: Bold numbers indicate relatively higher absolute values of correlation.

D. Determinant Tracking using Machine Learning

1) Tracking Determinants of Nutrition and Obesity through Diet Examination: We also provide the experimental analysis based on the diet dataset (Dataset 5), and track the health determinants. We also selected the top eleven developed countries in the world that eat less fruits and vegetables and the top seventeen developing countries in the world that consume the most vegetables, and analyze the effects of some food types (e.g., fruits and vegetables) on COVID-19 incidences (confirmed cases) and deaths.

To analyze the relationship between fruits intake and COVID-19 incidences (or deaths) and the relationship between vegetables intake and COVID-19 incidences (or deaths), we calculate the Pearson correlation between fruits intake and COVID-19 incidences (or deaths) and the Pearson correlation between vegetables intake and COVID-19 incidences (or deaths) based on the data from the above developed and developing countries. Table III summarizes the correlation between fruits (or vegetables) intake and COVID-19 incidences (or deaths). In Table III, we see that fruits intake is negatively correlated with both COVID-19 incidences and COVID-19 deaths, which suggests fruits intake can help to combat the Corona Virus. We also see that vegetables intake is negatively correlated with both COVID-19 incidences and COVID-19 deaths, which suggests vegetables intake can help to combat the Corona Virus. Moreover, we find that the absolute value of the correlation between vegetables intake and COVID-19

TABLE IV: Summary of top factors with the highest rank of COVID-19 incidences and deaths based on ClassifierAttributeEval

COVID-19 Incidences (Confirmed Cases)	Deaths
Alcoholic Beverages	Alcoholic Beverages
Animal Products	Animal Products
Aquatic Products, Other	Animal fats
Cereals - Excluding Beer	Aquatic Products, Other
Eggs	Fish, Seafood
Fish, Seafood	Cereals - Excluding Beer
Stimulants	Eggs
Sugar Crops	Offals

TABLE V: Summary of top factors with the highest rank of COVID-19 incidences and deaths based on CorrelationAttributeEval

COVID-19 Incidences (Confirmed Cases)	Deaths
Obesity	Obesity
Eggs	Animal Products
Animal Products	Eggs
Stimulants	Animal fats
Milk - Excluding Butter	Milk - Excluding Butter
Meat	Meat
Treenuts	Treenuts
Animal fats	Stimulants

incidences (or deaths) is higher than that of the correlation between fruits intake and COVID-19 incidences (or deaths), which indicates that vegetables intake is more effective/helpful to combat the Corona Virus.

We used ClassifierAttributeEval to examine the determinants of health disparities. Table IV shows the top features with the highest rank for rate of confirmed cases and deaths. In Table IV, we see that the food types “Alcoholic Beverages”, “Animal Products”, “Aquatic Products, Other”, “Cereals - Excluding Beer”, “Eggs”, “Fish, Seafood”, “Stimulants” and “Sugar Crops” have a higher impact on rate of confirmed cases, and the food types “Alcoholic Beverages”, “Animal Products”, “Animal fats”, “Aquatic Products, Other”, “Fish, Seafood”, “Cereals - Excluding Beer”, “Eggs” and “Offals” have a higher impact on rate of deaths. Moreover, some food types such as “Alcoholic Beverages”, “Animal Products”, “Aquatic Products, Other”, etc. have a higher impact on rate of confirmed cases and deaths.

We also used CorrelationAttributeEval to further examine the determinants of health disparities. Table V shows the top features with the highest rank for rate of confirmed cases and deaths. In Table V, we observe: “Obesity” has a higher impact on rate of confirmed cases and deaths, and it has the highest rank. We also see that the food types “Eggs”, “Animal Products”, “Stimulants”, “Milk - Excluding Butter”, “Meat”, “Treenuts” and “Animal fats” have a higher impact on rate of confirmed cases and deaths. The results suggest that people with obesity and people who eat food belonging to these types are more susceptible to COVID-19.

Also, we used ReliefFAttributeEval to further examine the determinants of health disparities. Table VI shows top features with the highest rank for the rate of confirmed cases and deaths. In Table VI, we see that “Stimulants” has a higher impact on rate of confirmed cases and deaths, and it has the highest rank. We also see that the other food types “Eggs”, “Miscellaneous”, “Offals”, “Animal fats”, “Meat” and “Fish, Seafood”, and “Obesity” have a higher impact on the rate of confirmed cases. The result indicates that people with obesity are more susceptible to COVID-19, and people who eat food belonging to these types are more susceptible to COVID-19. Similarly, we can find the top features that contribute most to the death rate. By examining the these top features, we find that some food types such as “Stimulants”, “Eggs”, “Animal fats”, “Meat”, “Offals” and “Obesity” have a higher impact on both COVID-19 infection and deaths.

TABLE VI: Summary of top factors with the highest rank of COVID-19 incidences and deaths based on ReliefFAttributeEval

COVID-19 Incidences (Confirmed Cases)	Deaths
Stimulants	Stimulants
Eggs	Animal fats
Obesity	Eggs
Miscellaneous	Meat
Offals	Aquatic Products, Other
Animal fats	Obesity
Meat	Treenuts
Fish, Seafood	Offals

VI. Conclusions

The COVID-19 pandemic has disproportionately and negatively impacted different communities. Examining health determinants (e.g. structural, social and constructural health) will assist in ascertaining why disparities occur with higher infection and death rates and in turn eventually help to improve pandemic health care. From a thorough analysis of health disparities we demonstrate that racial and ethnic minorities [such as Black Americans] have a higher rate of infections and deaths in comparison to other communities. From the in-depth analysis of diets we demonstrate that the disparities in nutrition and obesity play a crucial role in the health inequities unfolding during the pandemic and that race and ethnicity play a pivotal role in determining how and when care is accessed, and what the outcome might be. Our analysis highlights the fact that race and ethnicity play a pivotal role in determining how and when care is accessed, and what the outcome might be. Our findings suggest that the determinants are diverse. Determinants such as age, employment status, food sufficiency, diet and obesity, health insurance status, access to medical care, education, household income, housing units, poverty significantly contribute to the health disparities in this pandemic and hence COVID-19, as a disease, may potentially have devastating effects on communities of color. For now we provide new strategies for determining diverse health determinants, and new findings on health determinants for understanding health disparities. In the future, we will compare our approach with state-of-the-art to fully verify the performance of our approach. Also, we will collect the health data prior to COVID-19, and compare the experimental results based on the COVID-19 data and that of the health data prior to COVID-19 to further verify that the disparities are showing due to COVID-19. Finally, we will expand the dataset to further improve the performance of our approach.

Acknowledgment

This research was supported in part by a charitable contribution from Google LLC, COVID-19 Artificial Intelligence and Data Science for the Social Good Fund, and by the Faculty Research Awards Program at Florida A&M University.

References

[1] Who pandemic announcement. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020 [accessed on 31 October 2021].

[2] Covid-19 cases and deaths. https://covid19.who.int/ [accessed on 6 November 2021].

[3] S. Garg, L. Kim, M. Whitaker, A. ÓHalloran, C. Cummings, R. Holstein, M. Prill, S. Chai, P. Kirley, N. Alden, B. Kawasaki, K. Yousey-Hindes, L. Niccolai, E. Anderson, K. Openo, A. Weigel, M. Monroe, P. Ryan, J. Henderson, S. Kim, K. Como-Sabetti, R. Lynfield, D. Sosin, S. Torres, A. Muse, N. Bennett, L. Billing, M. Sutton, N. West, W. Schaffner, H. Talbot, C. Aquino, A. George, A. Budd, L. Brammer, G. Langley, A. Hall, and A. Fry. Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019–covid-net, 14 states, march 1-30, 2020. Technical report, MMWR Morb Mortal Wkly Rep. 2020;69:458-464, 2020.

[4] M. Shah, M. Sachdeva, and R. P. Dodiuk-Gad. Covid-19 and racial disparities. *J Am Acad Dermatol*, 83(1):e35, 2020.

[5] Alireza Hamidian Jahromi and Anahid Hamidianjahromi. Why african americans are a potential target for covid-19 infection in the united states. *J Med Internet Res*, 22(6):e19934, 2020.

[6] C. W. Yancy. Covid-19 and african americans. *JAMA*, 323(19):1891–1892, 2020.

[7] Louisiana department of health covid-19. http://ldh.la.gov/coronavirus/ [accessed on 6 November 2021].

[8] Nyc health covid-19: Data. https://www1.nyc.gov/site/doh/covid/covid-19-data.page [accessed on 6 November 2021].

[9] F. Echols. The st. louis american all 12 covid-19 deaths in the city of st. louis were black. https://www.stlamerican.com/your_health_matters/covid_19/all-12-covid-19-deaths-in-the-city-of-st-louis-were-black/article_da7ed56c-79d1-11ea-85bc-7b8539eaf346.html [accessed on 6 November 2021].

[10] Alabama public health characteristics of laboratory-confirmed cases of covid-19. https://www.alabamapublichealth.gov/covid19/assets/cov-al-cases-041520.pdf [accessed on 6 November 2021].

[11] A. Mokdad, K. Ballestros, and M. Echko. The state of us health, 1990-2016: Burden of diseases, injuries, and risk factors among us states. *JAMA*, 319(14):1444–1472, 2018.

[12] Key statistics & graphics. https://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us/key-statistics-graphics.aspx [accessed on 6 November 2021].

[13] M. Belanger, M. Hill, A. Angelidi, M. Dalamaga, J. Sowers, and C. Mantzoros. Covid-19 and disparities in nutrition and obesity. *N. Engl. J. Med.*, 2020.

[14] J. Liu, R. Alo, and Y. Parra Bautista. Deeptrace: Improving pandemic health care by identifying disparities and determinants. In *Proc. of IEEE BigComp*, pages 46–47, 2021.

[15] C. Laurencin and A. McClinton. The covid-19 pandemic: a call to action to identify and address racial and ethnic disparities. *Journal of Racial and Ethnic Health Disparities*, 7:398-402, 2020.

[16] M. Chowkwanyun and A. Reed. Racial health disparities and covid-19 – caution and context. *N. Engl. J. Med.*, May 6 2020.

[17] M. Hooper, A. Nápoles, and E. Pérez-Stable. Covid-19 and racial/ethnic disparities. *JAMA*, 323(24):2466–2467, 2020.

[18] D. M. Gray, A. Anyane-Yeboa, S. Balzora, R. B. Issaka, and F. P. May. Covid-19 and the other pandemic: populations made vulnerable by systemic inequity. *Nat Rev Gastroenterol Hepatol*, 17:520–522, 2020.

[19] K. Azar, Z. Shen, R. Romanelli, S. Lockhart, K. Smits, S. Robinson, S. Brown, and A. Pressman. Disparities in outcomes among covid-19 patients in a large health care system in california. *Health Affairs*, 39(7), 2020.

[20] T. M. Selden and T. A. Berdahl. Covid-19 and racial/ethnic disparities in health risk, employment, and household composition. *Health Affairs*, 39(9):1624–1632, 2020.

[21] Covid-19 deaths by race/ethnicity – kff. https://www.kff.org/other/state-indicator/covid-19-deaths-by-race-ethnicity/ [accessed on 30 July 2020].

[22] S. P. Jones. Imperial college london yougov covid data hub, v1.0. Technical report, Imperial College London Big Data Analytical Unit and YouGov Plc, April 2020.

[23] M. Hall. *Correlation based feature selection for machine learning*. PhD thesis, The University of Waikato, 1999.

[24] M. Hall. Correlation-based feature selection for discrete and numeric class machine learning. In *Proc. of ICML*, 2000.

[25] K. Kira and L. Rendell. A practical approach to feature selection. In *Proc. of ICML*, 1992.

[26] B. Yanti, E. Mulyadi, W. Wahiduddin, R. G. H. Novika, Y. M. D. Ariana, N. S. Martani, and Nawan. Community knowledge, attitudes, and behavior towards social distancing policy as prevention transmission of covid-19 in indonesia. *JAKI*, 8(1):4–14, 2020.

[27] K. A. Pawelek, C. Salmeron, and S. D. Valle. Connecting within and between-hosts dynamics in the influenza infection-staged epidemiological models with behavior change. *J Coupled Syst Multiscale Dyn*, 3(3):233–243, 2015.

[28] J. J. V. Bavel, K. Baicker, P. S. Boggio, V. Capraro, A. Cichocka, M. Cikara, et al. Using social and behavioural science to support covid-19 pandemic response. *Nature Human Behaviour*, 4:460–471, 2020.

[29] J. Liu, H. Shen, H. Narman, W. Chung, and Z. Lin. A survey of mobile crowdsensing techniques: A critical component for the internet of things. *ACM Trans. Cyber-Phys. Syst.*, 2(3):18:1–18:26, 2018.