DeepHealth: Geospatial and ML-based Approach to Identify Health Disparities and Determinants for Improving Pandemic Health Care

Jinwei Liu*, Rui Gong†, Long Cheng‡, Richard A. Alo*
*Department of Computer and Information Sciences, Florida A&M University, Tallahassee, FL 32307, USA
†Department of Informatics and Mathematics, Mercer University, Macon, GA 31207, USA
‡School of Control and Computer Engineering, North China Electric Power University, Beijing, China
*{jinwei.liu, richard.alo}@famu.edu, †gong_r@mercer.edu, ‡lcheng@ncepu.edu.cn

2023 32nd International Conference on Computer Communications and Networks (ICCCN) | 979-8-3503-3618-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCCN58024.2023.10230101

***Abstract*—The COVID-19 pandemic has exacerbated existing health disparities, and its impact has fallen disproportionately on disadvantaged and vulnerable communities. Racial and ethnic minorities such as Black Americans who are at a particular disadvantage are more likely to be the potential target of COVID-19 infection and are dying at alarmingly high rates. Despite a promising solution of the COVID-19 vaccination offers hope, equitable access to COVID-19 vaccines remains a challenge in the US, which has compounded the existing disparities in cases, hospitalizations, and deaths among racial and ethnic minority groups. The deep and pervasive history of medical racism in the US has led to the vaccine hesitancy in racial and ethnic minorities, and thereby caused the disparities. Although some studies examine determinants of health disparities (e.g., social health determinants), there is a shortage of studies examining the social, structural and constructural health determinants, either alone or in tandem with other determinants. Little research paid attention to leveraging geographic information to trace the social, structural and constructural health determinants, which can provide a lower level of granularity. In this paper, we propose DeepHealth, a geospatial and ML-based (machine learning based) approach to identify diverse determinants (including the social, structural, 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 based on multiple COVID-19 datasets and examine the social, structural, and constructural health determinants to assist in ascertaining why disparities (in racial and ethnic minorities who are particularly disadvantaged) occur in incidence and mortality rates due to COVID-19 pandemic. Extensive experimental results show the effectiveness of our approach. This research provides new strategies for health disparity identification and determinant tracking with a goal of mitigating health disparities and improving pandemic health care. The research suggests that policymakers should give attention to initiatives that will protect the health of populations (i.e., an upstream approach to reducing health disparities) rather than solely focusing only on providing health and social services.**

***Index Terms*—COVID-19, health care, health determinants, disparities, vaccine, pandemic, machine learning**

I. Introduction

The novel Coronavirus disease 2019 (COVID-19) pandemic has caused an immense impact on hospital systems, businesses, schools, and the economy. COVID-19 is now one of the most leading causes of death. Globally, as of April 12, 2023, the virus has resulted in 762,791,152 confirmed cases and 6,897,025 deaths [1]. Vulnerable populations around the world are disproportionately affected by COVID-19, and the inequities in the burden of COVID-19 are particularly striking in racial and ethnic minorities in the US [2], [3]. 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]. Based on 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]. In New York City, the mortality rates of African Americans and Latinos diagnosed with COVID-19 are 1.6 to 2 times higher than those of Whites [7]. In predominantly Black counties across the US, COVID-19 infections are threefold higher and mortality rates are sixfold higher than those in predominantly White counties [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], [8]–[12]. The causes of these inequities are multifold and involve differences in exposure, susceptibility, testing, and treatment [13]. The underlying causes of health disparities are complex and include social determinants such as unemployment, lack of adequate transportation, education and food, and housing insecurity, etc. [14], [15], structural determinants such as concentrated poverty, lack of wealth, racially segregated neighborhoods, etc. [16], [17], and constructural determinants such as influence of social and multimedia on behavior and potential modifications and adaptations, influence of diet or non-healthy diet [18]–[20].

Despite a promising solution of the COVID-19 vaccination offers hope, equitable access to COVID-19 vaccines remains a challenge in the US. Black communities have low vaccine uptake [21]. Unequal vaccine distribution has compounded the existing disparities in cases, hospitalizations, and deaths among racial and ethnic minority groups (e.g., Black Americans). Black Americans have a long-standing history of disadvantage, the deep and pervasive history of medical racism in the US has led to the vaccine hesitancy [22], and thereby caused the disparities.

There are some recent studies examining racial and ethnic health disparities related to COVID-19 [2], [23], [24], disparities in COVID-19 vaccination [25]–[27], social determinants [14], [28]–[30] and structural determinants [31], [32] of health disparities in COVID-19. Also, some studies [15] develop software/tools (e.g., chatbots) to mitigate health disparities among racial/ethnic minority groups. However, there is a shortage of studies examining the social, structural and constructural health determinants, either alone or in tandem with other determinants. Little research paid attention to leveraging geographic information to trace the social, structural and constructural health determinants, which can provide a lower level of granularity.

To address the problem, we propose DeepHealth 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. The main contributions of this paper are summarized as follows:

- 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 DeepHealth, a geospatial and ML-based approach to identify diverse determinants (including social, structural and constructural determinants) of health disparities in COVID-19 pandemic, which can help to improve pandemic health care.
- We examine the disparities of COVID-19 related vaccine (pneumonia vaccine and Flu vaccine), and analyze the correlation between COVID-19 vaccination status and the COVID-19 cases and deaths. We analyze the relationship among obesity, diabetes and poverty, and find that poverty is associated with increased risk of obesity and diabetes: poverty can cause obesity and thus cause diabetes.

The remainder of this paper is organized as follows. Section II presents the literature review. Section III describes multiple COVID-19 datasets. Section IV introduces the methods used in the paper and the design of DeepHealth. Section V presents the experiments and findings. Section VI concludes this paper with remarks on our future work.

II. Related Work

First we review prior work on COVID-19 health disparities and determinants. Then we review prior work on vaccine inequity. Finally we indicate the advantages of our proposed approach compared to previous studies.

A. COVID-19 Health Disparities and Determinants

COVID-19 pandemic has disproportionately affected the health of vulnerable populations. Many recent studies uncover health disparities in COVID-19. Laurencin *et al.* [23] presented the earliest available data on race and ethnicity in the peer-reviewed literature for those in America affected by COVID-19. Finally, they called for action to identify and address racial and ethnic health disparities in the COVID-19 crisis. Lopez *et al.* [33] revealed racial and ethnic health disparities related to COVID-19, and they provided suggestions for addressing health care disparities. Razai *et al.* [34] indicated ethnic disparities in COVID-19, and analyzed the causes of ethnic disparities in health outcomes. Finally, they gave suggestions for mitigating ethnic disparities in COVID-19 and beyond. Hasson *et al.* [35] revealed racial/ethnic disparities in COVID-19 diagnosis and physical activity disparities, and highlighted effective and feasible strategies that provide more equitable access to physical activity programs and spaces across the US. They called to action to increase equitable physical activity opportunities for all Americans for achieving health equity. Alcendor [36] indicated health disparities associated with COVID-19 mortality among underserved populations, and examined the underlying clinical implications that may predispose minority populations and the adverse clinical outcomes that may contribute to increased risk of mortality. Abrams *et al.* [28] examined social determinants of health (SDOH) and their effect on COVID-19 outcomes, and they emphasized that SDOH must be included as part of pandemic research priorities, public health goals, and policy implementation. Singu *et al.* [30] studied the social determinants of health, and how they impact disadvantaged populations during the COVID-19 pandemic. Debopadhaya *et al.* [14] examined how social determinants associated with COVID-19 mortality change over time, and they quantified the effect of 19 high-risk factors on COVID-19 mortality rate. McClure*et al.* [32] examined structural racism and they called on public health professionals to center structural determinants of health in etiological evaluation. However, the above studies do not provide a thorough analysis of health disparities and leverage geographic information to experimentally examine diverse determinants (social, structural and constructural determinants) of health disparities in COVID-19.

B. COVID-19 Vaccine Inequity

Racial and ethnic minority groups have low vaccine uptake due to vaccine hesitancy. Some studies examine the disparities in COVID-19 vaccination. Callaghan *et al.* [25] developed an original survey that was given to a national sample of 5009 Americans from May 28-June 8, 2020, and they also investigated the reasons why individuals did not intend to pursue COVID-19 vaccination. Razai *et al.* [26] tracked the reasons for COVID-19 vaccine hesitancy among ethnic minority groups. Painter *et al.* [27] described demographic characteristics of persons vaccinated during the first month of the COVID-19 vaccination program in the US. Also, the above studies do not provide an in-depth analysis of disparities and experimentally examine diverse determinants of the disparities.

Motivated by the problems in the existing literature, we propose a geospatial and ML-based approach to identify 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 COVID-19 datasets.

A. Dataset 1

We collected the data from KFF (a non-profit foundation that focuses on major health care issues in the US) [37]. This dataset provides information on COVID-19 cases, deaths, testing, and hospitalizations; COVID-19 vaccine allocation and administration; metrics by race/ethnicity; state policy actions; and related information. The category of COVID-19 disparities records COVID-19 cases by race/ethnicity, COVID-19 deaths by race/ethnicity, COVID-19 vaccinations by race/ethnicity, percent of total population that has received a COVID-19 vaccine by race/ethnicity; the category of mental illness during the COVID-19 pandemic records adults reporting symptoms of anxiety or depressive disorder during COVID-19 pandemic, unmet need for counseling or therapy among adults reporting symptoms of anxiety and/or depressive disorder during the COVID-19 pandemic; the category of economic effects records unemployment claims and percent change in state tax revenue, etc.

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 May 20, 2022. 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 and North Carolina. In the experiments, we extracted the daily COVID-19 cases and deaths and the population in each county of Florida and North Carolina from January 22, 2020 to May 20, 2022.

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 states Florida and North Carolina 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.

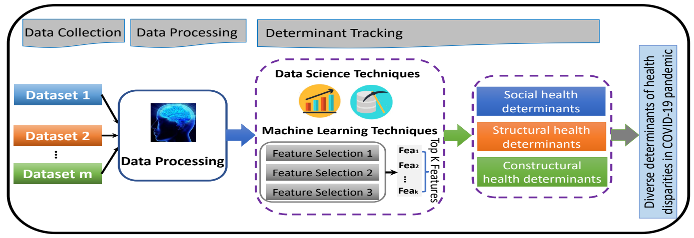


Fig. 1: The framework of DeepHealth.

D. Dataset 4

Also, we collected the COVID-19 data maintained by Our World in Data (https://github.com/owid/covid-19-data). The dataset records confirmed cases and deaths, hospitalizations and intensive care unit (ICU) admissions, vaccinations against COVID-19, testing for COVID-19, etc. The data about the confirmed cases and deaths comes from COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU); the data about the hospitalizations and intensive care unit (ICU) admissions was collected from official sources and collated by Our World in Data; the vaccination data was collected by the Our World in Data team from official reports [38]; the testing data was collected by the Our World in Data team from official reports.

E. Dataset 5

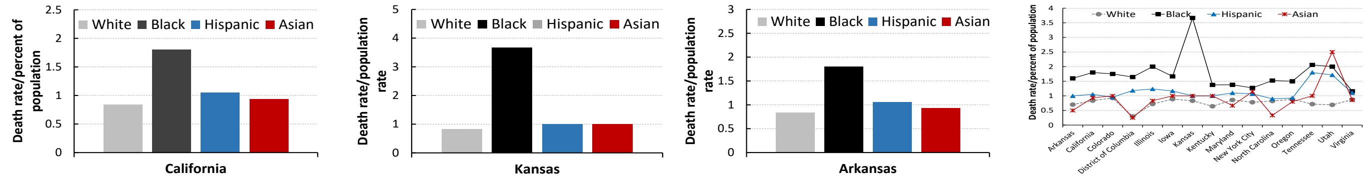
To further trace the determinants of health disparities, we also collected the data from the Imperial College London YouGov Covid 19 Behavior Tracker Data Hub (https://github.com/YouGov-Data/covid-19-tracker) [39]. The questions in the dataset, led by the Institute of Global Health Innovation (IGHI), 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 US. We used the attribute i3_health to identify the people who tested positive for COVID-19.

IV. Methods

A. Feature Selection Methods

To identify diverse determinants (e.g., social vulnerability index measures), we propose DeepHealth, 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.



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

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

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

(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.

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 [40], [41]. 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 [42]. 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 DeepHealth

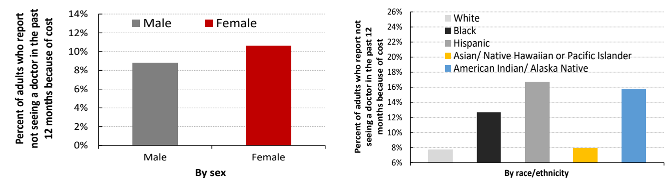
Figure 1 shows the framework of the DeepHealth. DeepHealth consists of three phases: data collection, data processing, determinant tracking. In data collection phase, DeepHealth collects the data with daily COVID-19 cases and deaths and population in each county of each state in the US, the data with 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, DeepHealth processes the data and extracts the information that is needed for determinant tracking. In determinant tracking phase, DeepHealth 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 multiple datasets, and trace the diverse determinants of health disparities.

A. Experimental Analysis Based on Dataset 1

Below we present the analysis results based on Dataset 1. We focus on analyzing how COVID-19 has disproportionately impacted different communities by analyzing COVID-19 deaths in different communities.



(a) By sex

(b) By race/ethnicity

Fig. 3: Percent of adults who report not seeing a doctor in the past 12 months because of cost.

Figure 2(a) shows the ratio of death rate and percent of population of different communities in California. 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 Kansas. In Figure 2(b), we also see the ratio of death rate and percent of population follows Black > Hispanic > Asian > 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 percent of population of different communities in Arkansas. 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, followed by Hispanic, Asian and White. 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.

Figure 3(a) shows the percent of adults who report not seeing a doctor in the past 12 months because of cost by sex. In Figure 3(a), we see that the percent of female adults who report not seeing a doctor in the past 12 months because of cost is higher than that of male adults. This shows that female adults has less chance of seeing a doctor due to the cost, suggesting the health disparities caused by gender. Figure 3(b) shows the percent of adults who report not seeing a doctor in the past 12

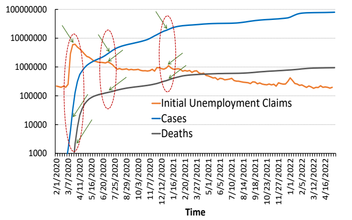


Fig. 4: Relationship among initial unemployment claims, cases and deaths.

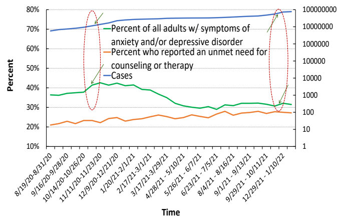
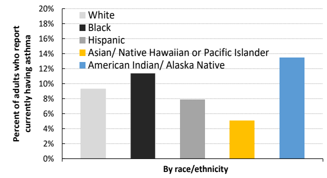
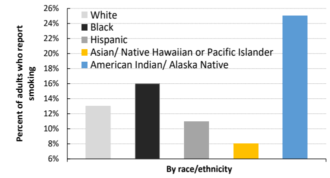


Fig. 5: Relationship among initial unemployment claims, cases and deaths.

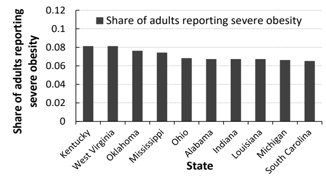


(a) Percent of adults who report currently having asthma

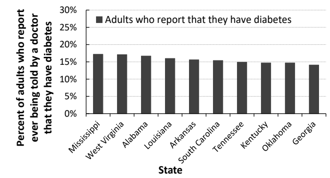


(b) Percent of adults who report smoking

Fig. 6: Percent of adults who report currently having asthma and smoking by race/ethnicity.



(a) Share of adults reporting severe obesity



(b) Percent of adults who report ever being told by a doctor that they have diabetes

Fig. 7: Share of adults reporting severe obesity and percent of adults who report ever being told by a doctor that they have diabetes.

months because of cost by race/ethnicity. In Figure 3(b), we see that the community White has the lowest percent of adults who report not seeing a doctor in the past 12 months because of cost. The minorities such as Black, Hispanic and American Indian/ Alaska Native have a higher percent, suggesting the health disparities caused by race and ethnicity.

Figure 4 shows the relationship among initial unemployment claims (measuring the number of people filing unemployment insurance claims during the past week and reflecting emerging unemployment), confirmed cases and deaths. In Figure 4, we see that the confirmed cases and deaths increase sharply when initial unemployment claims increase sharply (especially for the week of 4/4/2020). This shows that the COVID-19 cases and deaths are affected by employment status, which suggests that the determinant employment status significantly contributes to the health disparities in this pandemic.

Figure 5 shows the relationship among confirmed cases, percent of all adults with symptoms of anxiety and/or depressive disorder, and among adults with symptoms of anxiety and/or depressive disorder, percent who reported an unmet need for counseling or therapy. In Figure 5, we see that the percent of all adults with symptoms of anxiety and/or depressive disorder increases sharply between 10/14/2020 and 11/23/2020, and it decreases sharply between 1/20/2021 and 5/10/2021. Similarly, the confirmed cases increases sharply between 10/14/2020 and 1/18/2021, and it increases slowly between 1/18/2021 and 7/5/2021. The possible reason that the percent of all adults with symptoms of anxiety and/or depressive disorder decreases sharply between 1/20/2021 and 5/10/2021 and the confirmed cases increases slowly between 1/18/2021 and 7/5/2021 is that several COVID-19 vaccines were available, which helped remove symptoms of anxiety and/or depressive disorder of adults and reduce the COVID-19 incidence. Also, the percent of all adults with symptoms of anxiety and/or depressive disorder grows faster between 12/1/2021 and 1/10/2022, and it slightly decreases between 1/10/2022 and 2/7/2022. Similarly, the confirmed cases grows faster between 12/1/2020 and 1/10/2021, and it increases slowly between 1/10/2021 and 2/7/2021. The possible reason that the percent of all adults with symptoms of anxiety and/or depressive disorder grows faster between 12/1/2021 and 1/10/2022 and the confirmed cases grows faster between 12/1/2020 and 1/10/2021 is that Omicron had been detected in the US and continued to be the dominant variant in the US. This above result suggests that mental illness (e.g., anxiety and/or depressive disorder or hopeless) makes it easier for people to be infected with COVID-19, and it further indicates that the determinants such as anxiety and/or depressive disorder and the lack of access to health care significantly contribute to the health disparities in COVID-19 pandemic.

Figure 6(a) shows the percent of adults who report currently having asthma by race/ethnicity. In Figure 6(a), we see that the percent of adults who report currently having asthma in minorities such as Black and American Indian/ Alaska Native is higher than that in the other communities, which suggests the health disparities caused by race and ethnicity. Figure 6(b) shows the percent of adults who report smoking by race/ethnicity. Similarly, in Figure 6(b), we see that percent of adults who report smoking in minorities such as Black and American Indian/ Alaska Native is higher than that in the other communities, suggesting the health disparities caused by race and ethnicity. By examining Figures 6(a) and 6(b), we can find that tobacco smoke is a powerful trigger of asthma symptoms, irritating the lining of the airways. 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 affect people's health [43]–[45].

Figure 7(a) shows the share of adults reporting severe obesity in top 10 states with the highest share of adults reporting severe obesity. In Figure 7(a), we see that the share of adults reporting severe obesity follows Kentucky > West Virginia > Oklahoma > Mississippi > Ohio > Alabama > Indiana > Louisiana > Michigan > South Carolina. Figure 7(b) shows the percent of adults who report ever being told by a doctor that they have diabetes in top 10 states with the highest percent. In Figure 7(b), we see that the

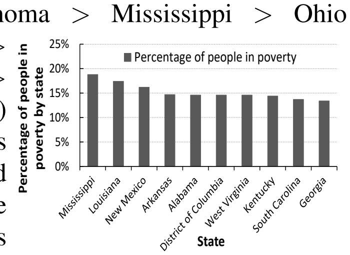
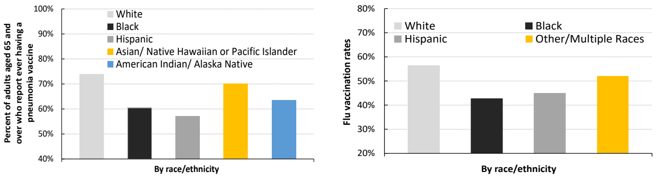


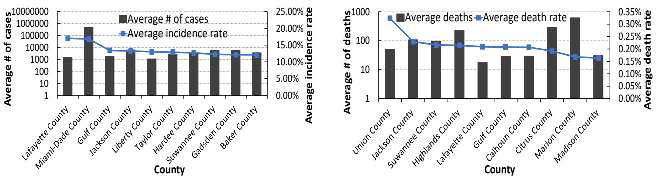
Fig. 8: Percentage of people in poverty by state in the US.



(a) Percent of adults aged 65 and over who report ever having a pneumonia vaccine

(b) Flu vaccination rates

Fig. 9: Percent of adults aged 65 and over who report ever having a pneumonia vaccine and the flu vaccination rates by race/ethnicity.



(a) Average number of cases and average incidence rate

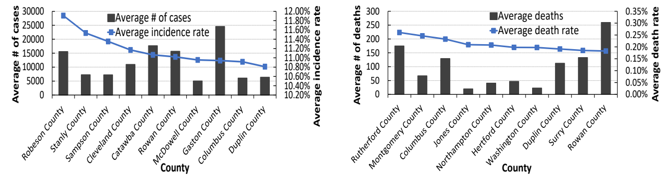
(b) Average number of deaths and average death rate

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

percent of adults who report ever being told by a doctor that they have diabetes follows Mississippi > West Virginia > Alabama > Louisiana > Arkansas > South Carolina > Tennessee > Kentucky > Oklahoma > Georgia. By examining Figure 7(a) and Figure 7(b), we find that 70% of the states Kentucky, West Virginia, Oklahoma, Mississippi, Alabama, Louisiana and South Carolina are in both lists of top 10 states, suggesting the geographical health disparities.

Figure 8 shows the top 10 US states with the highest percentage of people in poverty (the average percentage over three years from 2018 to 2020) based on the data collected from the US Census Bureau. In Figure 8, we see that the percentage of people in poverty follows Mississippi > Louisiana > New Mexico > Arkansas > Alabama > District of Columbia > West Virginia > Kentucky > South Carolina > Georgia. By examining Figures 7(a), 7(b) and 8, we find that 60% of the states Kentucky, West Virginia, Mississippi, Alabama, Louisiana and South Carolina are in all three lists of top 10 states, suggesting the geographical health disparities. Also, it indicates that when people are in poverty they are bound to buy junk food, which is the cheapest, which leads to obesity. When people are obese there is large chance of getting diabetes [46].

Figure 9(a) shows the percent of adults aged 65 and over who report ever having a pneumonia vaccine by race/ethnicity. In Figure 9(a), we see that the community White has the highest percent of adults aged 65 and over who report ever having a pneumonia vaccine. The minorities such as Black, Hispanic and American Indian/ Alaska Native have a lower percent, suggesting the health disparities caused by race and ethnicity. Figure 9(b) shows the flu vaccination rates by race/ethnicity. In Figure 9(b), we see that the community White has the highest flu vaccination rate, and the minorities such as Black and Hispanic have a lower flu vaccination rate. The community Black has the lowest flu vaccination rate, suggesting the health disparities caused by race and ethnicity.



(a) Average number of cases and average incidence rate

(b) Average number of deaths and average death rate

Fig. 11: Top 10 counties with the highest average incidence rate and average death rate in North Carolina.

B. Experimental Analysis Based on Datasets 2, 3 and 4

Figure 10(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 10(a), we see that the average number of cases follows Miami-Dade County > Jackson County > Gadsden County > Suwannee County > Baker County > Hardee County > Taylor County > Gulf County > Lafayette County > Liberty County, and the average incidence rate follows Lafayette County > Miami-Dade County > Gulf County > Jackson County > Liberty County > Taylor County > Hardee County > Suwannee County > Gadsden County > Baker County. Figure 10(b) shows the average number of deaths and average death rate in top 10 counties with the highest average death rate in Florida. In Figure 10(b), we see that the average number of deaths follows Marion County > Citrus County > Highlands County > Jackson County > Suwannee County > Union County > Madison County > Calhoun County > Gulf County > Lafayette County, and the average death rate follows Union County > Jackson County > Suwannee County > Highlands County > Lafayette County > Gulf County > Calhoun County > Citrus County > Marion County > Madison County.

Figure 11(a) shows the average number of cases and average incidence rate in top 10 counties with the highest average incidence rate in North Carolina. In Figure 11(a), we see that the average number of cases follows Gaston County > Catawba County > Rowan County > Robeson County > Cleveland County > Stanly County > Sampson County > Duplin County > Columbus County > McDowell County, and the average incidence rate follows Robeson County > Stanly County > Sampson County > Cleveland County > Catawba County > Rowan County > McDowell County > Gaston County > Columbus County > Duplin County. Figure 11(b) shows the average number of deaths and average death rate in top 10 counties with the highest average death rate in North Carolina. In Figure 11(b), we see that the average number of deaths follows Rowan County > Rutherford County > Surry County > Columbus County > Duplin County > Montgomery County > Hertford County > Northampton County > Washington County > Jones County, and the average death rate follows Rutherford County > Montgomery County > Columbus County > Jones County > Northampton County > Hertford County > Washington County > Duplin County > Surry County > Rowan County.

To analyze the relationship between factors such as social vulnerability index measures and COVID-19 incidence and

TABLE I: 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	0.1283	0.0087	**-0.3920**	**-0.3138**	**-0.5180**	**0.5103**
Death Rate	**0.3428**	0.0900	**-0.5321**	**-0.4694**	**-0.4700**	**0.3204**

	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	0.0365	**-0.3108**	-0.2122	**-0.2804**	**-0.3705**	-0.1034
Death Rate	-0.2140	**-0.4623**	**-0.5000**	**-0.4100**	**-0.5100**	**-0.3540**

	Total employment	Total employer establishments	Households w/ a computer	Households w/ a broadband Internet subscription	Mean travel time to work (16+ yrs)
Incidence Rate	0.1228	0.1394	**-0.3860**	**-0.3747**	0.1217
Death Rate	-0.2030	-0.1676	**-0.4370**	**-0.3385**	-0.0300

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

TABLE II: Significance test of the correlation between each factor and death rate in Florida.

Factor	**p-value (5% level)**	**Statistical Significance (5% level)**
With a disability (< 65 yrs)	0.0045	Significant
Persons W/o health insurance (<65)	0.4700	Not Significant
Percent of persons w/ bachelor's degree or higher (25+ yrs)	3.59E-06	Extremely Significant
Median household income	6.16E-05	Extremely Significant
Per capita income in past 12 mths	5.33E-05	Extremely Significant
Persons in poverty	0.0082	Significant
Housing units	0.0830	Not Significant
Median value of owner-occupied housing unit rate	8.19E-05	Extremely Significant
Median selected monthly owner costs-with a mortgage	2.34E-05	Extremely Significant
Median selected monthly owner costs-w/o a mortgage	0.0006	Extremely Significant
Median gross rent	8.26E-06	Extremely Significant
Building permits	0.0033	Significant
Total employment	0.0993	Not Significant
Total employer establishments	0.1753	Not Significant
Households w/ a computer	0.0002	Extremely Significant
Households w/ a broadband Internet subscription	0.0050	Significant
Mean travel time to work (16+ yrs)	0.7950	Not Significant

mortality, we calculate the Pearson correlation between each factor and incidence rate and mortality rate (death rate).

Table I shows the correlation between each factor and incidence rate and mortality rate, respectively based on the Florida dataset. In Table I, 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-w/o a mortgage', 'Median gross rent', '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 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 ('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-w/o a mortgage', 'Median gross rent', 'Households w/ a computer', 'Households w/ a broadband Internet subscription') have a relatively higher impact on COVID-19 incidence rate and mortality rate.

To test the significance of the correlation between the social vulnerability index measures and the COVID-19 incidence rate and mortality rate, respectively in Florida State. We tested the p-value of the correlation. Due to the page limit, we present the results of the significance of the correlation between the social vulnerability index measures and mortality rate. Table II shows the significance test of the correlation between the social vulnerability index measures and the COVID-19 mortality rate in Florida State. The experimental results show that there is a significant relationship between the 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 computer', 'Households w/ a broadband Internet subscription') and death rate. For most of the social vulnerability index measures, there is a significant relationship between them and the death rate.

Table III shows the correlation between each factor and incidence rate and mortality rate, respectively based on the North Carolina dataset. The results in Table III are similar to that in Table I, and they indicate that the 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', 'Households w/ a computer', 'Households w/ a broadband Internet subscription') have a relatively higher impact on COVID-19 incidence rate and mortality rate.

We also tested the significance of the correlation between the social vulnerability index measures and the COVID-19

TABLE III: Summary of correlation between each factor and incidence rate and death rate in North Carolina.

	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	0.1652	**0.2662**	**-0.4829**	**-0.4004**	**-0.5004**	**0.3656**
Death Rate	**0.4489**	0.1935	**-0.5821**	**-0.5839**	**-0.5585**	**0.5170**
	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	-0.0488	**-0.6424**	**-0.5882**	**-0.3376**	**-0.5054**	-0.0906
Death Rate	**-0.3093**	**-0.6858**	**-0.6215**	**-0.3332**	**-0.6192**	**-0.3384**
	Total employment	Total employer establishments	Households w/ a computer	Households w/ a broadband Internet subscription	Mean travel time to work (16+ yrs)	
Incidence Rate	-0.0339	-0.0605	**-0.3515**	**-0.3192**	0.1688	
Death Rate	**-0.2671**	**-0.2958**	**-0.6398**	**-0.6109**	-0.0554	

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

TABLE IV: Significance test of the correlation between each factor and incidence rate in North Carolina.

Factor	**p-value (5% level)**	**Statistical Significance (5% level)**
With a disability (< 65 yrs)	0.1006	Not Significant
Persons W/o health insurance (<65)	0.0074	Significant
Percent of persons w/ bachelor's degree or higher (25+ yrs)	3.61E-07	Extremely Significant
Median household income	3.66E-05	Extremely Significant
Per capita income in past 12 mths	1.15E-07	Extremely Significant
Persons in poverty	1.83E-04	Extremely Significant
Housing units	0.6296	Not Significant
Median value of owner-occupied housing unit rate	5.85E-13	Extremely Significant
Median selected monthly owner costs-with a mortgage	1.23E-10	Extremely Significant
Median selected monthly owner costs-w/o a mortgage	5.92E-04	Extremely Significant
Median gross rent	8.17E-08	Extremely Significant
Building permits	0.3701	Not Significant
Total employment	0.7379	Not Significant
Total employer establishments	0.5496	Not Significant
Households w/ a computer	0.0003	Extremely Significant
Households w/ a broadband Internet subscription	0.0012	Significant
Mean travel time to work (16+ yrs)	0.0931	Not Significant

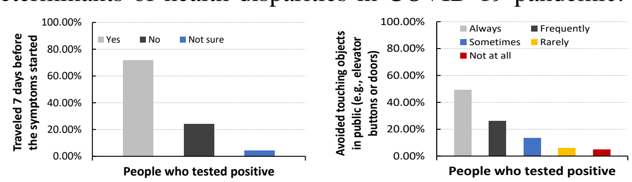
incidence rate and mortality rate, respectively in North Carolina State. Due to the page limit, we present the results of the significance of the correlation between the social vulnerability index measures and incidence rate. Table IV shows the significance test of the correlation between the social vulnerability index measures and the COVID-19 incidence rate in North Carolina. In Table IV, we observe similar results.

Table V shows the summary of the correlation between vaccination status and the confirmed cases/deaths in Florida. In Table V, we see that all the measures are positively correlated with both confirmed cases and deaths, and the correlation values are higher. The results suggest that vaccination can help reduce COVID-19 cases and deaths.

Table VI shows the summary of the correlation between vaccination status and the cases/deaths in North Carolina. In Table VI, we see that all the measures are positively correlated with both confirmed cases and deaths, and the correlation values are higher. The results further suggest that vaccination can help reduce COVID-19 cases and deaths.

C. Experimental Analysis Based on Dataset 5

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 [2]. A simulation study of agent-based influenza showed that small changes in behavior can have a significant effect on transmission patterns during epidemics [43], [47]. 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 [43], [44], [48]. Below we focus on people's behaviors in response to COVID-19 and examine the determinants of health disparities in COVID-19 pandemic.



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

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

Fig. 12: Status of traveling and touching objects in public (e.g. elevator buttons or doors).

Figure 12(a) shows whether a person traveled to a location where coronavirus has been reported 7 days before the person's symptoms started. We see that over 71% of the people who tested positive traveled to a location where coronavirus has 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 12(b) shows how frequently people who tested positive avoid touching objects in public (e.g. elevator buttons or doors). We see that around 50% of people always avoid touching objects in public, over 26% of the people frequently avoid touching objects in public. 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 can help prevent the spread of COVID-19.

D. Determinant Tracking using Machine Learning

1) County-Level Examination of Health Disparities and Determinant Tracking: Below we present the experimental results of the machine learning algorithms based on Dataset 2 and Dataset 3. In these datasets, We focus on the social vulnerability index measures with complete data for all counties in Florida and North Carolina. We used ClassifierAt-

TABLE V: Summary of correlation b/w vaccination status and cases/deaths in Florida.

	People vaccinated	People fully vaccinated	People vaccinated per hundred	People fully vaccinated per hundred	Total vaccinations	Total vaccinations per hundred
Cases	**0.8336**	**0.8215**	**0.8336**	**0.8215**	**0.9018**	**0.9018**
Deaths	**0.8710**	**0.8605**	**0.8710**	**0.8605**	**0.8031**	**0.8031**

Note: Bold numbers indicate higher absolute values of correlation.

TABLE VI: Summary of correlation b/w vaccination status and cases/deaths in North Carolina.

	People vaccinated	People fully vaccinated	People vaccinated per hundred	People fully vaccinated per hundred	Total vaccinations	Total vaccinations per hundred
Cases	**0.8785**	**0.7862**	**0.8785**	**0.7862**	**0.8850**	**0.8850**
Deaths	**0.9493**	**0.8836**	**0.9493**	**0.8836**	**0.9535**	**0.9535**

Note: Bold numbers indicate higher absolute values of correlation.

TABLE VII: Summary of top features with the highest rank for average incidence rate and mortality rate using ClassifierAttributeEval based on Florida dataset.

Ave. Incidence Rate	Ave. Death Rate
Households w/ a broadband	Households w/ a broadband
Internet subscription	Internet subscription
Age (<18 yrs)	Age (≥65 yrs)
Gender	With a disability, <65 yrs
With a disability, <65 yrs	Gender
Persons w/o health insurance, <65	Persons w/o health insurance, <65
Age (≥65 yrs)	Median household income
Age (<5 yrs)	Age (<5 yrs)
Median household income	Age (<18 yrs)
Hispanic or Latino	Hispanic or Latino
Black or African American alone	Black or African American alone

tributeEval, CorrelationAttributeEval and ReliefFAttributeEval to examine the determinants (e.g., social vulnerability index measures) of health disparities. Due to the page limit, we report the results based on ClassifierAttributeEval. Table VII shows the top features with the highest rank for average incidence rate and mortality rate based on the Florida dataset. In Table VII, we observe: the social vulnerability index measures "Households w/ a broadband Internet subscription", "Age (<18 yrs)", "Gender", "With a disability, <65 yrs", "Persons w/o health insurance, <65", "Age (≥65 yrs)", "Age (<5 yrs)", "Median household income", "Hispanic or Latino", "Black or African American alone" have a higher impact on COVID-19 incidence rate and mortality rate.

To further examine the determinants of health disparities, we also used ClassifierAttributeEval, CorrelationAttributeEval and ReliefFAttributeEval to examine the determinants of health disparities based on the North Carolina dataset. We obtained similar results (Due to the page limit, we report the results based on the last two feature selection methods). Table VIII and Table IX show the top features with the highest rank for average incidence rate and mortality rate obtained from the last two feature selection methods based on the North Carolina dataset.

VI. Conclusions

The COVID-19 pandemic has disproportionately affected vulnerable communities, and it has exacerbated existing health disparities. The racial and ethnic minorities such as Black Americans are more likely to be the target of COVID-19 infection due to the long-standing history of disadvantages. Examining health disparities and determinants (e.g. social, structural and constructural determinants) assist in ascertaining why disparities occur with higher incidence and mortality rates and in turn eventually help to mitigate health disparities and improve pandemic health care. By providing a thorough

TABLE VIII: Summary of top features with the highest rank for average incidence rate and mortality rate using CorrelationAttributeEval based on North Carolina dataset.

Ave. Incidence Rate	Ave. Death Rate
Age (<5 yrs)	American Indian & Alaska Native alone
Age (≥65 yrs)	Households w/ a broadband Internet subscription
Bachelor's degree or higher, percent of persons age 25 yrs+	Age (≥65 yrs)
With a disability, <65 yrs	Bachelor's degree or higher, percent of persons age 25 yrs+
White alone	With a disability, <65 yrs
Households w/ a broadband Internet subscription	Households w/ a computer
Owner-occupied housing unit rate	Black or African American alone
Persons in poverty	Persons in poverty
Gender	Hispanic or Latino
Households w/ a computer	Owner-occupied housing unit rate

TABLE IX: Summary of top features with the highest rank for average incidence rate and mortality rate using ReliefFAttributeEval based on North Carolina dataset.

Ave. Incidence Rate	Ave. Death Rate
Median value of owner-occupied housing units	American Indian & Alaska Native alone
Per capita income in past 12 mths	Median value of owner-occupied housing units
Median gross rent	Households w/ a broadband Internet subscription
Mean travel time to work (mins), workers age 16+	Median selected monthly owner costs-w/o a mortgage
Median selected monthly owner costs-with a mortgage	Households w/ a computer
Age (≥65 yrs)	Age (≥65 yrs)
Bachelor's degree or higher, percent of persons age 25 yrs+	Persons in poverty
Age (<5 yrs)	Per capita income in past 12 mths
Owner-occupied housing unit rate	Persons w/o health insurance, <65
Median household income	Black or African American alone

analysis of health disparities, we demonstrate that racial and ethnic minorities such as Black Americans have a higher incidence and mortality rates compared to other communities. 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, poverty, food sufficiency, obesity, disability, health insurance status, income, vaccination status, symptoms of anxiety, depressive disorder, tobacco use, diabetes, asthma, access to digital literacy, payment of housing units (or rent), etc. 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 (especially those suffering from obesity, experiencing food insecurity and living in food deserts). The virus itself does not discriminate, but America's history of discrimination creates potential longerterm scenarios. 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 the Faculty Research Awards Program at Florida A&M University. We would like to thank Mr. Trevor Taylor, Ms. Deyandra Burke, Raygen Bostick and Shamya Oscar for their help on this work.

References

[1] Covid-19 cases and deaths. https://covid19.who.int/ [accessed on 14 April 2022].

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

[3] D. B. G. Tai, A. Shah, C. A. Doubeni, I. G. Sia, and M. L. Wieland. The disproportionate impact of covid-19 on racial and ethnic minorities in the united states. *Clin Infect Dis.*, 72(4):703–706, 2021.

[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] New York City Health. Age-adjusted rates of lab confirmed covid-19 nonhospitalized cases, estimated non-fatal hospitalized cases, and patients known to have died 100,000 by race/ethnicity group as of april 16, 2020. https://www1.nyc.gov/assets/doh/downloads/pdf/imm/covid-19-deaths-race-ethnicity-04162020-1.pdf [accessed on 2 September 2022].

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

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

[10] 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].

[11] 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].

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

[13] P. Blumenshine, et al. Pandemic influenza planning in the united states from a health disparities perspective. *EID*, 14(5):709–715, 2008.

[14] S. Debopadhaya, J. S. Erickson, and K. P. Bennett. Temporal analysis of social determinants associated with covid-19 mortality. In *BCB*, 2021.

[15] J. Kim, J. Muhic, L. P. Robert, and S. Y. Park. Designing chatbots with black americans with chronic conditions: Overcoming challenges against covid-19. In *Proc. of CHI*, New Orleans, 2022.

[16] N. Thakur, S. Lovinsky-Desir, et al. The structural and social determinants of the racial/ethnic disparities in the u.s. covid-19 pandemic. what's our role? *Am J Respir Crit Care Med*, 202(7):943–949, 2020.

[17] C. N. Harrington, R. Garg, A. Woodward, and D. Williams. "it's kind of like code-switching": Black older adults' experiences with a voice assistant for health information seeking. In *Proc. of CHI*, 2022.

[18] L. D. Renzo, P. Gualtieri, F. Pivari, L. Soldati, A. Attinà, G. Cinelli, C. Leggeri, G. Caparello, L. Barrea, F. Scerbo, E. Esposito, and A. D. Lorenzo. Eating habits and lifestyle changes during covid-19 lockdown: an italian survey. *J Transl Med*, 18(229), 2020.

[19] P. L. Liu. Covid-19 information on social media and preventive behaviors: Managing the pandemic through personal responsibility. *Soc Sci Med*, 277:113928, 2021.

[20] J. Liu, L. Cheng, and H. Chi. Deepdive: Examining determinants of health disparities in covid-19 for improving pandemic health care. In *Proc. of ICVISP*, 2020.

[21] N. Restrepo and H. J. Krouse. Covid-19 disparities and vaccine hesitancy in black americans: What ethical lessons can be learned? *Otolaryngol–Head Neck Surg*, 166(6):1147–1160, 2021.

[22] L. L. Diamond, H. Batan, et al. The polyvocality of online covid-19 vaccine narratives that invoke medical racism. In *Proc. of CHI*, 2022.

[23] 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.

[24] S. J. Kim and W. Bostwick. Social vulnerability and racial inequality in covid-19 deaths in chicago. *Health Educ Behav*, 47(4):509–13, 2020.

[25] T. Callaghan, A. Moghtaderi, et al. Correlates and disparities of intention to vaccinate against covid-19. *Soc Sci Med*, 272:113638, 2021.

[26] M. S. Razai, T. Osama, D. G. J. McKechnie, and A. Majeed. Covid-19 vaccine hesitancy among ethnic minority groups. *BMJ*, 372:n513, 2021.

[27] E. M. Painter, E. N. Ussery, A. Patel, M. M. Hughes, E. R. Zell, et al. Demographic characteristics of persons vaccinated during the first month of the covid-19 vaccination program - united states, december 14, 2020-january 14, 2021. *MMWR Morb Mortal Wkly Rep*, 70(5):174–177, 2021.

[28] E. M. Abrams and S. J. Szefler. Covid-19 and the impact of social determinants of health. *Lancet Respir Med*, 8(7):659–661, 2020.

[29] B. Burström and W. Tao. Social determinants of health and inequalities in covid-19. *Journal of Public Health*, 30(4):617–618, 2020.

[30] S. Singu, A. Acharya, K. Challagundla, and S. N. Byrareddy. Impact of social determinants of health on the emerging covid-19 pandemic in the united states. *Frontiers in public health*, 8:406, 2020.

[31] S. Khalatbari-Soltani, R. G. Cumming, et al. Importance of collecting data on socioeconomic determinants from the early stage of the covid-19 outbreak onwards. *J. Epidemiol. Commun.*, 74:620-623, 2020.

[32] E. S. McClure, P. Vasudevan, Z. Bailey, S. Patel, and W. R. Robinson. Racial capitalism within public health—how occupational settings drive covid-19 disparities. *Am J Epidemiol*, 189(11):1244–1253, 2020.

[33] L. Lopez, L. H. Hart, and M. H. Katz. Racial and ethnic health disparities related to covid-19. *JAMA*, 325(8):719–720, 2021.

[34] M. S. Razai, H. K. N. Kankam, et al. Mitigating ethnic disparities in covid-19 and beyond. *BMJ*, 372:m4921, 2021.

[35] R. Hasson, J. F. Sallis, et al. Covid-19: Implications for physical activity, health disparities, and health equity. *AJLM*, 16(4):420–433, 2022.

[36] D. J. Alcendor. Racial disparities-associated covid-19 mortality among minority populations in the us. *J Clin Med*, 9(8):24–42, 2020.

[37] State health facts – kff. https://www.kff.org/state-category/covid-19/ [accessed on 15 June 2022].

[38] E. Mathieu, H. Ritchie, E. Ortiz-Ospina, et al. A global database of covid-19 vaccinations. *Nature Human Behaviour*, 5:947–953, 2021.

[39] 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.

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

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

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

[43] B. Yanti, E. Mulyadi, W. Wahiduddin, et al. Community knowledge, attitudes, and behavior towards social distancing policy as prevention transmission of covid-19 in indonesia. *JAKI*, 8(1):4–14, 2020.

[44] J. J. V. Bavel, et al. Using social and behavioural science to support covid-19 pandemic response. *Nat. Hum. Behav.*, 4:460–471, 2020.

[45] J. Liu, R. Alo, and Y. Parra Bautista. Deeptrace: Improving us pandemic health care through health disparity identification and determinant tracing. In *Proc. of IDSTA*, 2020.

[46] K. Sahoo, B. Sahoo, A. K. Choudhury, et al. Childhood obesity: causes and consequences. *J Family Med Prim Care*, 4(2):187–192, 2015.

[47] 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. *JCSMD*, 3(3):233–243, 2015.

[48] 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.