IMPACT OF SDOH ELEMENTS IN VALUE-BASED CARE MODEL TO DRIVE BETTER CLINICAL OUTCOME FOR US HEALTH PLANS

by

Dhanasekaran Radhakrishnan

DISSERTATION

Presented to the Swiss School of Business and Management, Geneva

In Partial Fulfilment

Of the Requirements

For the Degree

DOCTOR OF BUSINESS ADMINISTRATION

SWISS SCHOOL OF BUSINESS AND MANAGEMENT GENEVA

MAY 2025

IMPACT OF SDOH ELEMENTS IN VALUE BASED CARE MODEL TO DRIVE BETTER CLINICAL OUTCOME FOR US HEALTH PLANS

by

Dhanasekaran Radhakrishnan

Supervised By

Prof. David Annan

APPROVED BY

____

Dissertation Chair

RECEIVED BY

Renee Goldstein Osmic

SSBM Representative

Dedication

I dedicate this work to my family, whose unwavering support and encouragement have been my foundation. To Prof. David Annan, for his selfless follow-ups and guidance. And to my entire product management team, led by Akram, Amrita, and Chenny, whose collaboration and dedication have inspired me throughout this journey.

Acknowledgements

I would like to express my deepest gratitude to the entire working colleagues for their invaluable support, collaboration, and encouragement throughout the course of this research. This work would not have been possible without the expertise, dedication, and passion each team member brought to our shared goals.

Working alongside such a talented and driven group has been both an honor and a learning experience that has profoundly shaped my scientific journey. I look forward to carrying forward the lessons, values, and spirit of collaboration fostered within this team.

Thank you for being an essential part of this accomplishment.

ABSTRACT

IMPACT OF SDOH ELEMENTS IN VALUE-BASED CARE MODEL TO DRIVE BETTER CLINICAL OUTCOME FOR US HEALTH PLANS

Dhanasekaran Radhakrishnan

2025

Supervisor

Prof. David Annan

The transition to value-based care (VBC) models in the United States has highlighted the critical role of Social Determinants of Health (SDOH) in achieving better clinical outcomes. This paper explores how addressing SDOH elements—such as socioeconomic status, education, housing stability, food security, and access to transportation—can drive improved patient outcomes and reduce overall healthcare costs for U.S. health plans. Through a review of emerging strategies and case studies, and using a quantitative methodology, the study examines how health plans are integrating SDOH data into care management, provider partnerships, and risk adjustment methodologies. The findings demonstrate that proactive SDOH interventions not only enhance member engagement and care quality but also support the financial sustainability of VBC arrangements. As U.S. health plans continue to refine their approaches, the alignment of clinical care with social supports will be essential to maximize outcomes and equity in healthcare delivery.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviations	Full Form
SDOH	Social Determinants of Health
ACOs	Accountable Care Organisations
PSE	Policies, Systems, And Environments
IOM	Institute of Medicine
VBHC	Value-Based Healthcare
ML	Machine Learning
SVMs	Support Vector Machines
ACS	American Community Survey
SVI	Social Vulnerability Index
KFF	Kaiser Family Foundation
EHRs	Electronic Health Records
AHIP	America's Health Insurance Plans
US	United States
ANN	Artificial Neural Network
DL	Deep Learning
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
VBP	Value-Based Payment
SPSS	Statistical Packages for Social Sciences
ACS	American Community Survey
SVI	Social Vulnerability Index
CDC	Centers for Disease Control and Prevention
USDA	United States Department of Agriculture

CHAPTER I:

INTRODUCTION

1.1 Introduction

There is a mountain of convincing research that has built up over the past two decades showing that social variables, aside from medical treatment, significantly impact health in a variety of contexts, populations, and health indices. The data presented here do not disprove that medical care affects people's well-being. Instead, it shows that healthcare is not the only thing affecting people's well-being, and it implies that healthcare's impacts may not be as big as previously believed, especially when it comes to figuring out who becomes ill or hurt (Braveman, Egerter and Williams, (2011); Adler and Stewart, (2010); Braveman *et al.*, (2011); McGinnis and Foege, (1993) . There are ongoing debates over the quality of the evidence indicating that some social elements have a causal impact on health, and the links among social factors and health are not straightforward. All the while, scholars are questioning whether traditional criteria are enough for assessing the data (Kelly et al., 2007; Glasgow & Emmons, 2007).

McKeown et al., (1975) showed the limits of medicine by researching the mortality data of England and Wales from the mid-nineteenth century to the early 1960s. Prior to the invention of contemporary medical procedures like antibiotics and critical care units, he discovered that mortality rates had been decreasing consistently for decades. According to McKeown, the main causes of the sharp rise in life expectancy during the 19th century include better living circumstances, such as clean water, hygienic environments, and proper diet (McKeown et al., 1975).

Most writers think that nonmedical variables, such problems falling within the conventional public health scope, were probably more important in boosting living standards, however public health nursing and its advocacy role could have contributed significantly(Grundy, 2005). Increasing inequalities in mortality rates across UK socioeconomic groups in the decades after the launch of the National Health Service in 1948 (making healthcare available to everyone) is another illustration of the limitations of medical treatment. Martinson discovered, using more current data, that although overall health was better in the UK than in the US, which does not have universal coverage, health inequalities by wealth were comparable in the two nations (Martinson, 2012). Despite increased access to healthcare for everyone, large health disparities by socioeconomic status have been recorded time and time again throughout European nations (Mackenbach et al., 1997; Mackenbach et al., 2008).

The United States has a relative ranking that has been declining over the years, and it routinely ranks last or near last among rich countries on important health metrics including life expectancy and infant mortality. Despite investing more money in healthcare per person than any other industrialized country, this remains the case. When compared to other developed nations, the United States' death and morbidity rates are shockingly high. This disadvantage persists across most age groups except for those over 75 years old, and it is true for both wealthy and poor Americans, as well as for non-Latino whites when these factors are considered independently. In the United States, for instance, it has been noted that despite an increase in prenatal care for African American women due to Medicaid maternity care expansions in the 1990s, racial disparities in the two most important birth outcomes—low birthweight and premature delivery—did not narrow (Ananth et al., 2001). Traditional clinical prenatal care is crucial for mother health, but it hasn't been shown to enhance baby outcomes in most cases.

Social Determinants of Health

With the rise of accountable care organizations (ACOs) and other programs aimed at bettering the health of populations, social determinants of health (SDOH) have been hot topics in health policy debates as of late (Hacker & Walker, 2013). The need for health care professionals to demonstrate their effect via improved population health is growing. Because medical treatment impacts such a small percentage of overall health, there are substantial barriers to enabling healthcare providers to enhance public health via ACOs and value-based payment models (Marmot, 2005; McGinnis et al., 2002).

Figure 1.1: Social Determinants of Health (PEI, 2024)

Numerous research have pitted health care services, genetics, behaviors, the environment, and social factors against one another in an effort to improve health and reduce early death (Prus, 2011). Social, behavioral, and environmental determinants of health, which are not directly related to medical issues, have regularly been shown to have a much greater impact on health results than medical variables. Even while the relative contributions could differ by 5-10% from one health result to another, similar trends do hold for certain health outcomes, including costly and onerous illnesses like diabetes, heart disease, stroke, and vascular disease(Platz et al., 2000; Hu et al., 2001).

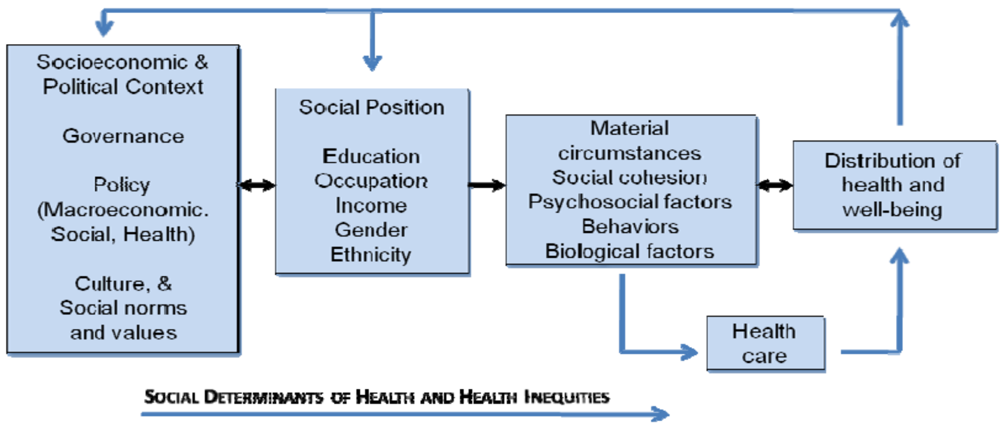

Figure 1.2: Social Determinants conceptual framework (Andvik, 2010)

Despite the data, one of the most persistent problems with the research on socioeconomic determinants of health is how to turn its findings into practical suggestions. Numerous studies, some of which go back to the 1970s, have shown that negative social determinants of health have negative effects on health both immediately and over time. On the other hand, there is mounting evidence that positive social conditions positively affect health outcomes (Myers et al., (2014); Freedman et al., (2011); Braveman et al., (2010); Kumar et al., (2012); Loucks et al., (2015); Carroll-Scott et al., (2013)).

Unfortunately, there has not been a comprehensive literature study to compile an evidence-based summary of the most effective strategies to tackle socioeconomic determinants of health that may lead to positive health outcomes while maintaining or reducing health care expenses. This study sets out to gather all the information we have so far on how social service interventions affect healthcare spending and health outcomes in the hopes of identifying programs and policies that can do both.

1.2 Research problem

In the US, medical professionals are making great strides towards patient-centered care, but it is not enough to only give top-notch clinical treatment; we must also address the non-medical aspects that impact people's health. Social determinants of health, which include economic and social variables, are responsible for as much as 75% of health outcomes, according to research(Ferrer, 2023).

Social and environmental variables, including economic stability, education, food security, and housing, are responsible for at least 20% of the premature deaths in the US; this is particularly the case for those facing disadvantages, as shown by statistics from the KFF. Health care professionals must priorities patient well-being by delivering comprehensive treatment that takes into account and resolves SDOH, social risk factors, and social needs. Providers have a great chance to do this via VBP, which prioritizes community health and person-centered care.

Due to their insufficient reimbursement of providers for providing treatment outside of their own premises, traditional Fee for Service (FFS) models fail to provide the appropriate incentives to address non-medical problems. VBP's continued success in paying for health rather than amount of services provided bodes well for the healthcare industry's efforts to address an impact of socioeconomic factors on community and patient health(Malamou, 2015).

Transportation services, like on-demand businesses, may be partnered with a provider taking part in a value-based payment risk agreement to guarantee that at-risk patients have the means to get to their appointments. Payers and providers are also collaborating with neighborhood groups that combat food poverty and spending more in housing for the homeless and low-income. Individuals' health may be hindered or threatened by social risk factors and unmet social needs; VBP promotes and incentivizes clinicians and payers to proactively address these issues. More and more people are starting to notice that SDOH is a problem. In fact, a recent poll found that 80% of payers think that improving their population health programs would include tackling the SDOH of their beneficiary populations(Hahn, 2021).

Addressing SDOH, unmet social needs, and promoting health equality may be achieved via value-based care approaches. Eighty percent (80%) of hospitals surveyed by Deloitte said their leadership is dedicated to creating and implementing procedures to systematically address social needs as an aspect of clinical care, proving that hospitals and health systems are making investments in health-related social needs with strong support from leadership. There are still gaps in linking programs that enhance health outcomes or save costs, and a lot of activity is ad hoc, according to our study, meaning it only happens sometimes and only reaches part of the target population.

1.3 Purpose of research

This study aims to investigate how health plans and provider organizations may better use data to address SDOH. This study aims to assess the necessity for additional data collection related to SDOH, identifying gaps and opportunities that could enhance the understanding of these determinants in specific populations.

Furthermore, it seeks to identify effective strategies for leveraging existing data across value-based programs, which could lead to improved resource allocation and better-informed decision-making.

Additionally, the research will investigate available open-source resources that can be utilized to generate valuable insights related to SDOH, ensuring that organizations can access cost-effective tools for data analysis. It will also examine what additional insights can be derived beyond SDOH, providing a more comprehensive view of the various factors influencing health outcomes.

A key component of our research is to identify the best ways to convey these findings to healthcare practitioners and health insurance. This includes evaluating whether a separate report is necessary for clarity and actionability. Finally, the research will explore methods for achieving synergy among value-based programs, facilitating collaboration and integration across different health initiatives.

Ultimately, this study aims to provide actionable recommendations that will empower organizations to utilize data-driven approaches in improving health care delivery, enhancing clinical outcomes, and promoting health equity within diverse populations. By bridging the gap between data and practice, the research aspires to contribute meaningfully to public health initiatives and foster a more equitable healthcare system.

1.4 Significance of the study

This study is significant for several reasons. First, it addresses a critical gap in understanding how health plans and provider organizations can effectively utilize existing data to generate insights related to SDOH. By investigating the necessity for additional data collection, the research will inform stakeholders about potential areas for enhancement in their data strategies, ultimately leading to more comprehensive insights into the factors affecting health outcomes.

Moreover, the study's focus on integrating existing data across value-based programs is crucial for optimizing resource allocation and improving care delivery. By examining how to leverage existing datasets, organizations can identify patterns and correlations that may have previously gone unnoticed, thereby enhancing the efficacy of their interventions.

The exploration of available open-source resources for generating SDOH insights will empower health plans and provide organizations to utilize cost-effective tools for analysis, promoting efficiency and reducing barriers to access. This democratization of data resources can lead to more equitable health outcomes, particularly for underserved populations.

Additionally, by investigating the potential for generating insights beyond SDOH, the study will contribute to a more holistic understanding of the various determinants of health. This broader perspective allows for more targeted and multifaceted interventions, ultimately improving clinical outcomes.

The development of effective communication strategies to disseminate these insights will ensure that findings are not only accessible but also actionable for health plans and providers. Understanding the best methods for sharing insights will facilitate the implementation of data-driven strategies in practice, enhancing collaboration among stakeholders.

Finally, by fostering synergy among value-based programs, this research aims to create a more cohesive approach to healthcare delivery. The insights generated from this study can lead to improved clinical outcomes and greater health equity within diverse populations. By bridging the gap between data and practice, the research holds the potential to drive meaningful changes in public health initiatives, contributing to a healthier and more equitable society.

1.5 Research purpose and hypothesis

The main goal of this research is to assist health plans and provider organizations in identifying existing data within their systems and leveraging it to generate meaningful insights related to the SDOH of their populations, ultimately aimed at improving clinical outcomes. To achieve this, the following research questions will be addressed:

- What gaps exist in current data collection efforts related to Social Determinants of Health (SDOH), and is there a need for additional data capture?
- In what ways can existing SDOH data be optimally leveraged to enhance value-based care programs?

- Which open-source tools and datasets are available to support the generation of actionable insights on SDOH?
- Beyond SDOH-specific factors, what additional insights can be derived from integrated clinical, claims, and social data sources?
- What are the most effective strategies for communicating SDOH-related insights to health plans and providers, and would a separate, dedicated reporting mechanism improve decision-making?

Research Hypothesis

- **H1:** Current data collection efforts inadequately capture critical SDOH factors, leading to incomplete patient profiles and missed opportunities for intervention.
- **H2:** Leveraging existing SDOH data through advanced analytics and integration into care models significantly improves clinical outcomes and cost savings in value-based care programs.
- **H3:** Open-source tools and publicly available datasets can meaningfully supplement proprietary health plan data to generate actionable SDOH insights.
- **H4:** Integrating SDOH with clinical and claims data yields richer, more predictive insights than using clinical data alone, enabling better population health management.
- **H5:** Dedicated and tailored reporting of SDOH insights to providers and health plans enhances clinical decision-making and member engagement more effectively than embedding SDOH information within general reporting structures.

CHAPTER II:

REVIEW OF LITERATURE

2.1 Theories and conceptual framework

The transformation of the U.S. healthcare system from a fee-for-service model to a value-based care (VBC) framework has ushered in a renewed focus on outcomes, patient-centered care, and cost-efficiency. Within this evolving paradigm, Social Determinants of Health (SDOH), including factors such as income, education, housing, transportation, food security, and social support have emerged as critical drivers of both clinical outcomes and health equity. The growing body of literature recognizes that up to 80% of health outcomes are influenced by non-medical factors, underscoring the urgency for health plans and care providers to address SDOH in care delivery models.

As U.S. health plans strive to align financial incentives with quality care, the integration of SDOH into value-based care strategies has become increasingly prominent. Stakeholders are exploring how targeted interventions ranging from community partnerships to predictive analytics can improve population health, reduce disparities, and ultimately drive better clinical outcomes. However, despite growing interest, there remain significant gaps in understanding the most effective ways to incorporate SDOH into VBC frameworks.

This literature review examines the current evidence on the impact of SDOH elements within value-based care models, evaluating how these factors influence health outcomes, utilization patterns, and payer strategies. It also explores best practices, policy implications, and key challenges in operationalizing SDOH-informed interventions across U.S. health plans.

Insufficiencies of Medical Care (DOH)

Medical care is thought to account for just 10–20% of the modifiable variables that lead to a healthy population (Hood et al., 2016). The remaining 80 to 90% may be attributed to health-related behaviors, socioeconomic factors, and environmental variables, which are frequently referred to as the SDoH. Despite health expenditures accounting for a higher share of GDP in the US compared to other industrialized countries, comparing expenditure on the SDoH is more difficult. We are aware that compared to the United States, many industrialized nations spend a disproportionate amount more on social services (Bradley et al., 2011).

Influence of Policies and Programs

There is a strong correlation between policies, systems, and environments (PSE) and SDoH. Figure 2.1 is a visual representation of the link among health results, the SDoH, and policies and activities provided by County-Based Health Rankings and Roadmaps. For instance, although the availability of cessation clinics and quit lines is important, the price of cigarettes and smoke-free communal spaces have a greater impact on reducing tobacco use. Tobacco use is associated with a decline in life expectancy and other negative health effects.

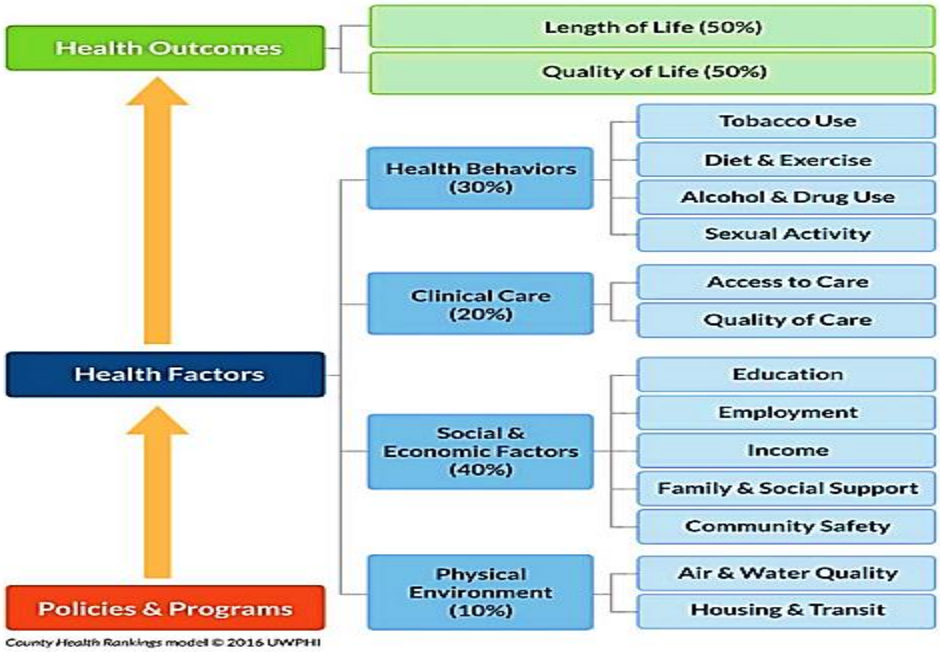

Figure 2.1: County Health Rankings & Roadmaps

New Payment Models Are Prompting Interest in the SDoH

Payment for results rather than process measurements and standards for "total cost of care" are becoming more prevalent in new value-based payment models like Medicare Shared Savings, ACOs, patient-centered medical homes, and alternative payment models. Can we expect payment structures to change in a way that rewards communities and health care organizations for shared goals like reducing the incidence of tobacco, obesity, and diabetes or increasing the number of high school graduates, given the strong correlation between better SDoH performance and improved health outcomes?

Emerging Frameworks for Integrating SDoH

There are data-frameworks that aim to capture SDoH domains in EHRs and integrate SDoH into primary care. An method that takes into account both individual and community-driven data for primary care is one way to address the persistent worries about effectiveness (DeVoe et al., 2016). The framework doesn't mention how the data can be utilized in combination with community partnerships to make data collecting more impactful. The creation of screening tools has been funded by two entities: a pediatric emergency department that serves low-income families and a responsible health community initiative (L. Gottlieb et al., 2014). Emerging strategies for following up on screening data include "clinic-to-community treatment models" for children from food-insecure homes, for example (Barnidge et al., 2017).

The EHR should include social and behavioral health areas, according to the Institute of Medicine (IOM) (Palermo & Beals-Erickson, 2015). The viability of implementing SDoH into EHRs has been examined, along with the incentive, training, and privacy challenges (L. M. Gottlieb et al., 2015). Importantly, more people disclosed sensitive information about themselves during electronic screenings than during in-person ones, especially when it came to topics like drug abuse and violence. Clinical trials were recently recommended after an evaluation of the practicality, dependability, and validity of the domains suggested by the IOM (with the exception of income) (Giuse et al., 2017; Prather et al., 2017).

Experiments at the Local and Federal Level

Models for state innovation include looking at how social services, health care, and certain SDoH are all interconnected. ACOs are addressing patients' non-medical needs including food, housing, and transportation in the hopes that this would lead to better results at lower costs. Further trials should be conducted to determine the best use of treatment funds, as the authors of a randomized pediatric intervention of in-person navigation services at two safety-net hospitals suggest that children reported overall health status improved, and families' reports of social needs decreased. The CMS launched ACH as the first innovation-center method to prioritize community resources while dealing with the demands of a population (here, CMS enrollees). The five-year ACH model involves a thorough review strategy and is divided into two tracks: the assistance track, which aids in navigating community services, and the alignment track, which promotes partner alignment to guarantee responsive and accessible services. Experiments like this will give additional proof that these methods work to improve results, user experience, and cost-effectiveness.

Impact of SDOH Interventions in Healthcare Systems

Unmet societal demands are being addressed by recent large-scale efforts. PREPARE is a social needs screening instrument that was created by the National Association of Community Health Centers for use by health center patients. In the meantime, the Accountable Health Communities experiment was initiated by the Centers for Medicare and Medicaid to study how healthcare costs are affected when social needs that go unmet are addressed for Medicare and Medicaid recipients living in the community. "S&R" stands for screening and referral for unmet social needs, and it is being promoted or required by most state Medicaid ACOs and Medicaid managed care organizations. These and similar smaller-scale initiatives have been put in place to tackle unmet social needs, but there is still no comprehensive theoretical framework that describes the potential ways in which healthcare-based interventions could enhance health results.

For instance, recent research showed that unmet social need S&R was linked to lower blood pressure and cholesterol, although it was unclear how the intervention produced these results. Similar studies that found that interventions targeting housing and economic requirements had favorable health benefits could not explain how these interventions produced these favorable results (Herd et al., 2008; Ludwig et al., 2011). As a result, it is unknown which aspects of an SDoH intervention caused the impact and why; thus, it is uncertain what can be duplicated or changed. Considering this, a framework would be useful in directing and educating future initiatives to identify and solve unmet societal needs, as well as to assess the results of such initiatives.

Upstream and Downstream Social Determinants of Health

Sometimes called "social determinants of health," these factors include everything other than medical care that has an impact on people's well-being, such as their knowledge, perspective, and actions (such as smoking). But these characteristics are influenced by things farther upstream, and they only reflect things further downstream in the chain of events that impact health. The upstream/downstream metaphor may be seen in the case of riverbank residents who become ill after consuming water polluted with harmful chemicals produced by a business located upstream(Braveman et al., 2011).

Although polluting one's water supply is the most direct source of illness, the most basic (and maybe less obvious, considering the distance in time and space) cause is the dumping of chemicals upstream. Some individuals may be advised to buy water filters to make the polluted water safe to drink. This might lead to socioeconomic disparities in illness rates, since the wealthy would likely be able to afford these water purification options. The factory's dumping would stop with the upstream approach, which would target the source of pollution. Even while these concepts can seem straightforward, the causal chains that connect upstream determinants to downstream determinants and, eventually, to health are usually complex and lengthy, sometimes comprising many intervening and perhaps interacting variables. Usually, this complexity makes it simpler to investigate and deal with downstream factors, but at the risk of ignoring underlying problems(Castañeda et al., 2015).

The most promising avenues for enhancing health and decreasing health inequalities are those that have a more basic causal influence. The conceptual basis for the work of the RWJF Commission is shown in Figure 2.2. This simplified schema emphasizes several fundamental concepts, even if the linkages are more complicated. The first important finding is that it establishes a direct correlation between health-related activities and the adoption of prescribed medical treatments. Instead, these elements are impacted by more systemic elements related to people's living and working environments, which can have a detrimental effect on health in two ways: directly, through stressful situations or toxic exposures, and indirectly, through the choices people make about their own and their families' health (P. Braveman et al., 2011).

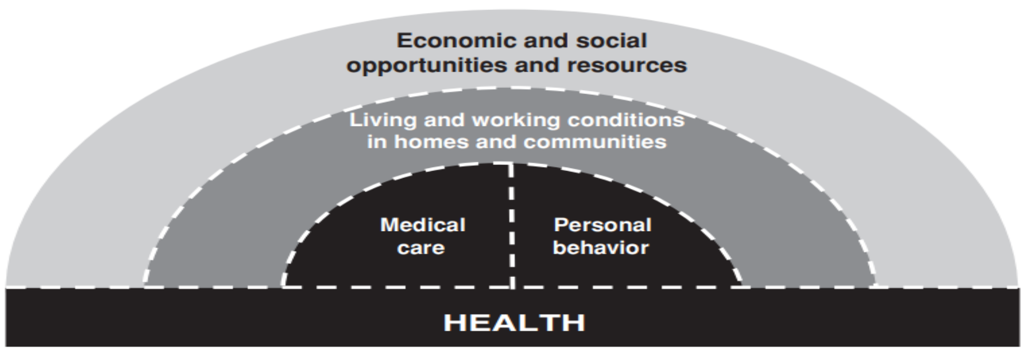

Figure 2.2: Upstream and downstream determinants

The graphic shows that a person's health is affected by a variety of factors, including their living and working conditions, but also by antecedent variables, which are more abstract and reflect a person's social and economic resources and opportunities.

Understanding Health Inequities, SDOH and Health outcomes

The term "health inequities" describes unequal and preventable disparities in health outcomes caused by systemic disparities in the opportunity that various groups have to attain optimum health. The racially and ethnically marginalized, the LGBTQ community, the disabled, the economically disadvantaged, and the rural poor are disproportionately impacted by these inequalities. As a result of SDOH, health inequalities are defined as disparities by the National Academy of Sciences.

A person's social and environmental circumstances, including their access to food, education, housing, transportation, and social support, make up their SDOH. A person's capacity to get healthcare and maintain adherence to treatments critical to their health is profoundly affected by these factors. Approximately 80% of a person's health is determined by SDOH, according to the research. Among the most robust indicators of the general population's health is SDOH and its corresponding variables. Several studies have shown that SDH may be responsible for 30–55 percent of health outcomes. It is imperative that civil society and all other parts of society work together to combat SDH in a manner that improves health and reduces long-term health disparities(Krause et al., 2021).

According to the available evidence, people from lower socioeconomic backgrounds, with less education, who reside and work in economically disadvantaged areas, are more prone to a multitude of risk factors that heighten their vulnerability to chronic diseases like diabetes and CVD (An et al., 2016). Also, dietary deficits, frequent in homes experiencing food insecurity, might raise the risk of developing chronic diseases. In addition to this, social isolation and material deprivation are both exacerbated by low income. It is more challenging for people to engage in cultural, educational, and recreational activities when they lack financial resources.

In addition, disparities in healthcare access stem from other sources, such as inadequate funding for medical research and services and inadequate health insurance. Major health problems and chronic diseases are more likely to go untreated among the uninsured, according to research (Cole et al., 2018; Seo et al., 2019).

Numerous research has shown a link between mortality rates, educational attainment, and health issues. With the financial security that comes from well-paying, stable jobs, families are better equipped to save for medical care. Less educated adults are more likely to smoke, eat poorly, and lead sedentary lives. Many significant health problems, like diabetes, liver disorders, CVD, and mental health difficulties, including sadness, anxiety, and worthlessness, are linked to a lack of knowledge, according to a study. Environmental factors impact every person and every community on their health.

Communities with lower incomes are disproportionately affected by environmental dangers such as air pollution, harmful agricultural chemicals, and water contamination.

Natural disasters may have different impacts on different populations due to differences in exposure and susceptibility, which can be explained by variables such as socioeconomic situation, level of education, gender, age, class, handicap, and health condition. If populations are uprooted or forced to evacuate, these injustices are magnified. Research demonstrates that wealth inequality, particularly in connection to race, education, and homeownership, increases over time as damage due to natural disasters rises (Laub, 1999). Barriers to high-quality healthcare are also correlated with geography, and these barriers might affect health outcomes. All five of the nation's top killers—cardiovascular disease, stroke, cancer, accidental injury, and chronic obstructive pulmonary disease—were more common in rural areas.

All underserved populations face challenges in getting the proper care they need because of the time, money, and inconvenience of travel. Due to the higher prevalence of rural poverty compared to urban poor, people residing in rural areas may face additional difficulties while attempting to travel large distances for treatment. People risk their lives attempting to seek the attention they need in a potentially fatal situation when they have to travel greater distances to get emergency medical care. Telehealth can aid in easing the difficulties associated with transportation in remote locations.

Initiatives to Address Social Determinants of Health

Efforts to address the social determinants of health have recently increased. There are efforts to have the health care system deal with larger social and environmental elements that impact health, and there are others afoot to have the health care system bring more attention to health in non-health areas.

Non-health sector policies and practices impact health and health equality. Public transportation's accessibility and availability, for instance, influence access to work, reasonably priced, healthful food, medical treatment, and other crucial factors that influence well-being and health. There is a great deal of untapped potential for health improvement in nutrition policy and initiatives.

Gardening programs in schools and communities, better options in low-income corner shops, and other initiatives to increase access to and consumption of nutritious food are just a few examples (Franck et al., 2013). Aside from helping children from low-income families and minority groups succeed academically, improving the health of students from low-income backgrounds, and promoting health equity, early childhood education has other positive effects.

The "Health in All Policies" strategy works to include health concerns into policy and industry decision-making. Understanding the effects of different sectors' activities on health and how enhanced health might contribute to the goals of these sectors is essential to a health-first policymaking strategy. Health, fairness, and sustainability are just a few of the many goals it seeks to achieve. Other goals include bolstering the agricultural sector, increasing educational attainment, creating jobs, stabilising the economy, improving transportation, and facilitating mobility.

Health in All Policies is being implemented at the state and municipal levels via several workgroups and task forces that aim to unite community and agency leaders in a shared commitment to health and health equality. The National Prevention Strategy was developed via a collaborative effort between the public, stakeholders, and the Prevention Advisory Group, which was formed by the ACA, as well as the leadership of several government departments, agencies, and offices.

To enhance health in areas or neighbourhoods that have poor health outcomes, place-based programs use cross-sector solutions. In recent years, a person's zip code has surpassed their genetic code as a predictor of their health, and the correlation between a person's area and their well-being is still widely recognised. Poor health outcomes and health inequalities are common in neighbourhoods where social, economic, and environmental obstacles are present, and many projects aim to address this by coordinating actions across sectors.

An excellent illustration of this is the HCZ program, which seeks to enhance the quality of life for children living in a 100-block area of Central Harlem that has much higher rates of infant mortality, chronic illnesses, poverty, and unemployment when contrasted with the rest of the city. HCZ's mission is to strengthen families, strengthen communities, and promote health, education, and economic opportunities for all members of the community via a variety of initiatives.

Value-based healthcare (VBHC)

Michael Porter, a professor at Harvard University, invented the acronym VBHC to describe value-based healthcare. Redefining Health Care: Creating Value-Based Competition on Results was released in 2006 by him and Elizabeth Teisberg (Gray, 2006). Reorganizing healthcare to prioritize competition and better patient outcomes was their recommended solution. For progress to be made, there must be some degree of competition. Competition is what moves knowledge forward and improves customer value in other domains of skill, including technology. This kind of rivalry is still present in the healthcare industry, but it is dysfunctional and fails to provide value to patients. Dr. Porter argues that value is best understood from the perspective of the consumer and that this is best achieved by maximizing health benefits while minimizing expenditure. Maximum health benefit with little expenditure is how Conrad defines health (Conrad, 2015).

Physicians must spearhead the transition from the current paradigm to the value-based approach, which is founded on three tenets: First, importance must be placed on value. Second, medical practice needs to be structured according to medical conditions and the care cycle. Third, outcomes have to be quantified (Porter & Teisberg, 2007).

While shifting to a value-based structure is no easy feat, it is doable, and improving results is the surest method to keep expenses in check (Porter, 2009). The foundation of a long-term health-care system is an emphasis on value (Bozic & Wright, 2012). It is less expensive to attain and maintain excellent health than to treat poor health. A value-based approach is being adopted by the industry, not only by doctors. Orthopaedics, for instance, uses value-based implants (Lybrand & Althausen, 2018). They produce single-use kits to save the expenses of sales agents and sterilization. The surgeon's conflict of interest with the company is one of the obstacles to the use of these implants, but they are surmountable.

A value-based strategy has been gradually replacing our current fee-for-service reimbursement paradigm in recent years. This new model seeks to link payment to quality and value (Ray & Kusumoto, 2016). The challenge in putting it into practice is figuring out how to measure value and quality. Different initiatives are being developed by professional groups in an effort to identify high value.

We need tools that can measure our progress towards our objectives. In VBHC, quality metrics are necessary for assessing the efficacy of care delivery in terms of patient benefit by quantifying health-care procedures, outcomes, patient experiences, and organizational systems (Debaun et al., 2019). Value and a successful conclusion might vary from person to person and from situation to circumstance. Developing a single tool that works for all conditions is challenging. Nevertheless, in order to improve patient value, how can we put this concept into action and move towards a value-based approach? This chapter aims to address that question. The issue is rather straightforward, but the solution is intricate. A small number of hospitals in the US and other countries are following this kind of patient-centered treatment. A handful of them will be reviewed, along with their implementation. While cost-effectiveness, patient-centered care, and evidence-based medicine are not the same thing, they may be combined to create value-based healthcare.

Machine Learning Application in Healthcare

Numerous benefits accrue to patients and doctors alike from the growing use of ML in healthcare. The most common uses of ML in healthcare now include automated healthcare billing, medical decision assistance, and the creation of standards for medical treatment. Multiple examples of healthcare models using ML may be found in the medical domain. ML has several current uses in healthcare, such as improved radiation, personalized treatment plans, smart health records, crowdsourcing data gathering, medical imaging diagnostics, clinical trials, ML-based behavioural change, and research (Sutabri et al., 2019). Patients getting radiation treatment for head and neck cancers may now be prepared for potentially serious adverse effects because to the first medical ML system (Bak et al., 2022).

Medical DL automatically detects intricate radiological patterns and aids radiologists in making educated assessments of images from conventional radiography, PET, MRI, CT scans, and radiology reports. When compared to radiologists, Google's ML apps for healthcare had an accurate rate of 89% in diagnosing breast cancer. All of these applications of ML in healthcare are only the tip of the iceberg. Machine learning has made it possible to extract medical data components from patient files, including medical diagnoses, treatment plans, and prescription drugs.

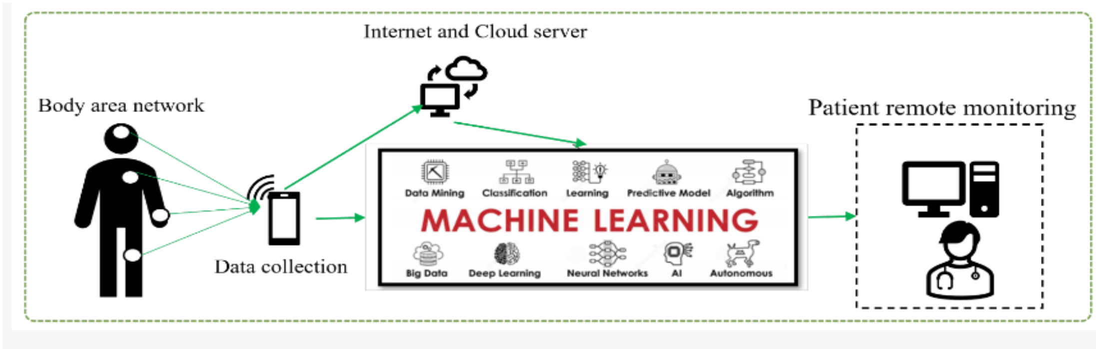

Figure 2.3: Concept of machine learning in healthcare area

Using ML with medical data aims to train computers to understand doctors' speech patterns and understand the perspective (negation, hypothetical) of key medical terms. Strong negation engines may distinguish between four basic negation categories: history, family history (mother, husband), negative (denies), and history again. The system can achieve 97 percent accuracy with over 500 negation phrases (Shukla et al., 2024). Because they enable companies to operate their systems without having to invest much in infrastructure expansion, ML applications are becoming more and more popular in the sector (Shahid et al., 2019)

ML eliminates the need for domain-specific code by solving problems using a combination of statistical methods and algorithmic models (Samuel, 2000). There is a lot of processing and feature extraction that happens before the data is given into the algorithm as most ML models just have one layer (Bellodi et al., 2022). To prevent the training dataset from being over- or underfitted and to provide reliable predictions even without further layers, many ML approaches need extensive data preparation. The majority of ML and AI algorithms are based on some kind of learning approach. As seen in Figure 2.4, supervised and unsupervised ML are the two primary forms of the field.

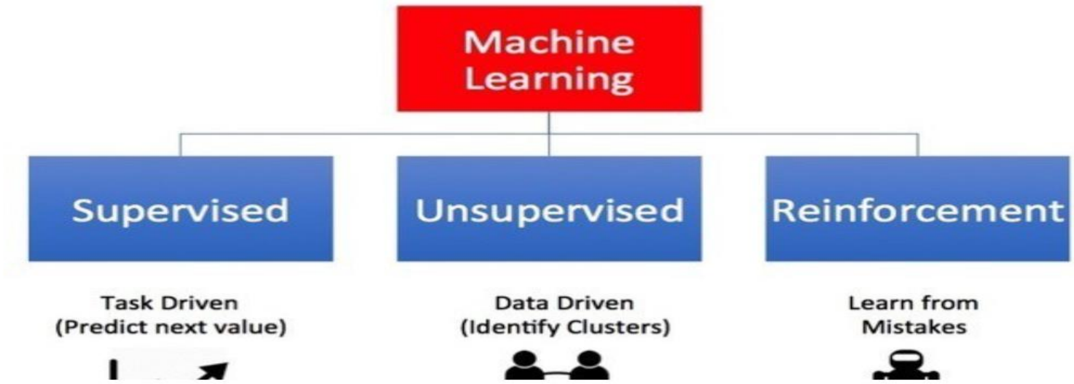

Figure 2.4: Different types of machine learning

In supervised ML, algorithms are taught to predict future outputs using data that already exists for both inputs and outputs. By exploring the input data for latent patterns or underlying structures, unsupervised ML finds these things. Unsupervised ML focusses on clustering, while supervised ML can handle classifications and regression (Ciaburro et al., 2021).

2.2 Value-Based Payment Models

This is a glaring example of how carelessness in data collecting and consent may lead to the financial and emotional exploitation of user data. Threats to privacy and information management can have detrimental effects, such as data breaches, monetary losses, and a decline in customer confidence in businesses. As a result of surveillance capitalism, Amal feels as though his every move is being monitored and that he is being paid high fees without giving his permission.

The only way for physicians to improve patient health outcomes is to address social determinants of health (SDOH), social risk factors, and social needs in a comprehensive manner. VBP, which is rooted in community wellness and person-centered care, has the ability to greatly assist providers in this endeavor. Because they do not sufficiently compensate physicians for treating patients outside of their office, tenured fee-for-service (FFS) models should be reevaluated if non-medical concerns are to be addressed. Since

VBP pays for health rather than the number of treatments delivered, it will continue to be a useful tool in pushing the healthcare system to consider the impact of social determinants on patient and community health(Hatef et al., 2021).

By forming partnerships with on-demand transportation services and other transportation providers, participating providers in value-based payment risk agreements can ensure that patients who require transportation can get to their appointments on time. Additionally, payers and providers are collaborating with neighborhood organizations that address food insecurity and spend more on housing for the homeless and low-income. VBP encourages payers and providers to proactively identify unmet social needs and social risk factors that could endanger or impede an individual's health. There is a rising realization of the need to address SDOH, as seen by recent research that shows eighty percent of payers think it will be a vital technique to strengthen their population health initiatives. This finding highlights the critical importance of addressing social determinants of health (SDOH) within healthcare systems. Unmet health needs that arise from economic instability, lack of education, poor housing conditions, and limited access to nutritious food or transportation are disproportionately experienced by underserved and marginalized populations. These unmet needs are deeply rooted in structural inequities and are strongly linked to poorer health outcomes, contributing to long-standing health disparities.

In this context, value-based care (VBC) models offer a transformative framework. Unlike traditional fee-for-service models that focus on the volume of care, VBC emphasizes quality, outcomes, and equity, aligning payment structures with the goal of improving patient health holistically. Integrating social justice principles into VBC means not only treating medical conditions, but also proactively identifying and mitigating social risk factors that hinder individuals from achieving optimal health.

VBC approaches can drive investment into community partnerships, preventive interventions, and wraparound services that address the root causes of poor health. By embedding SDOH screening into routine clinical workflows and allocating resources toward social support (e.g., housing assistance, food programs, transportation services), providers can close care gaps and foster equitable health outcomes.

Therefore, tackling SDOH through a value-based care lens is not just a healthcare strategy it is a moral and public health imperative that ensures a fairer, more just system where every individual has the opportunity to achieve their full health potential.

Research from Deloitte indicates that hospitals and health systems are making investments in social needs related to health, and that there is strong leadership support for these efforts: According to 80% of hospital responders, hospital leadership is dedicated to creating and refining procedures that offer clinical care that systematically addresses social needs. Nevertheless, the results also show that there are still gaps in linking programs that enhance health outcomes or lower costs, and that a large portion of activity is still ad hoc—as defined in our survey—and only reaches a portion of the target population.

Health equity plans and solutions have yet to effectively put this knowledge to use, despite growing understanding of how social and individual difference factors lead to disparities. However, there is currently greater attention on health equity than ever due to the growth of value-based care arrangements and required reporting criteria tied to health gaps.

Due to the fragmented approach of the healthcare system, both plans and providers lack some solid support from data front and SDOH insights. My research is filling that gap and assisting health plans and providers to address their populations SDOH concerns.

2.3 Domains of SDOH

Economic Stability

Employment:

Every day, many people are either working or actively seeking employment 17, 18, 16, and 15 Factors connected to one's workplace, such as duties, pay, working conditions, and job security, may have an effect on one's health.

Psychosocial stress is one of the harmful job situations that might raise the risk of adverse health effects (Shain & Kramer, 2004). In 2019, 5,333 people lost their lives and 2.8 million were injured on the job (Bureau of Labor Statistics, 2017). If a worker's employment involves repeatedly lifting, hauling, or pushing big things, they are more likely to get sick or injured. inadequate workplace supplies, such as chairs and keyboards prolonged exposure to dangerous substances such asbestos, insecticides, lead, and aerosols(National Institute for Occupational Safety and Health, 2004) or a noisy work environment(Hager, 2002).

Additionally, factors of psychosocial stress at work include extremely demanding occupations and a lack of control over daily tasks (Shain & Kramer, 2004). High levels of interpersonal conflict are among the additional causes of stress at work Schieman & Reid (2009), working more than eight hours a day, taking on numerous jobs, and working evening shifts (Caruso et al., 2004). These pressures put people at an increased risk of death Sabbath et al. (2015) and depression Simmons & Swanberg (2009), in addition to the possibility that they are linked to the increase in parental disengagement and conflict (Repetti & Wang, 2014). Smoking and alcoholism are two bad coping mechanisms that people with very demanding occupations may employ (Hoel et al., 2001).

Food Insecurity:

The social and economic condition of food insecurity occurs when families do not have enough food to eat (U.S. Department Of Agriculture, 2024). In 2020, 13.8 million families faced the challenge of food insecurity (U.S. Department Of Agriculture, 2024). There are other causes of hunger besides food insecurity, However, one consequence of food instability is hunger (Carlson et al., 1999).

The following are the two ways that food insecurity is classified by the United States Department of Agriculture (USDA) (U.S. Department Of Agriculture, 2024):

Insufficient food security: Claims that the diet is less appetizing, varied, or of low quality. Very little evidence of a decrease in caloric consumption.

Extremely poor food security: "There have been several reports of indications of disturbed eating patterns and decreased food consumption."

Temporary or long-term food insecurity is possible (Jones et al., 2013). Employment, income, race/ethnicity, and disability are all potential determinants of this. Increased likelihood of food insecurity occurs when purchasing food is financially difficult or impossible (Sharkey et al., 2011; Nord, 2007). While 10.5% of American families experienced food insecurity in 2020, 28.6% of low-income households did. A family's ability to afford food is another area that can suffer when breadwinners are out of work. It is more difficult for households to meet their basic food demands in low-income areas due to the high unemployment rate (Nord, 2007). Children whose parents are jobless are more likely to go hungry than children whose parents have jobs (Nord, 2009). Due to reduced spare income due to health care costs and less career options, people with disabilities may be more likely to experience food poverty (Huang et al., 2010). When it comes to food insecurity, there are racial and ethnic disparities.

Local conditions may affect the physical accessibility of food (Zenk et al., 2005). People living in low-income areas, rural areas, and even certain urban centers may not always have easy access to grocery stores or full-service supermarkets (Ver Ploeg, 2009). Compared to mostly white and non-Hispanic neighborhoods, predominantly Black and Hispanic neighborhoods could have fewer full-service supermarkets (Powell et al., 2007). Compared to supermarkets and grocery stores, convenience stores typically have a worse assortment of lower-quality foods, as well as greater prices. Due to limited transportation alternatives and lengthy commutes, people are even less inclined to eat healthily (Ver Ploeg, 2009).

Housing Instability

Some examples of housing instability include having trouble paying rent, living in confined quarters, relocating often, or housing taking up a disproportionate amount of a household's income. An individual's physical health may suffer as a result of these occurrences, and access to necessary medical care may be hindered (Kushel et al., 2006).

For a household to be classified as cost burdened, housing expenses must take up more than 30% of their income. If housing costs go above 50% of their income, they are classified as severely cost burdened (Bailey et al., 2016). Households with high expenses have little money left over each month for other essentials like clothing, food, electricity, and health care (Kushel et al., 2006). Cost stress is nearly twice as common in Black and Hispanic homes as it is in white households.

In 2019 JCHS Harvard University (2020): Of the 37.1 million households that were affected by costs, 17.6 million were severely burdened. Cost pressures affected 83.5% of the households with annual incomes below $15,000.

Since there aren't enough low-cost rental apartments, people in the lowest income brackets may have to settle for substandard housing that poses health and safety risks like mould, rodents, water leaks, and insufficient HVAC systems (JCHS Harvard University, 2020). On top of that, they may have little choice but to share housing, which could increase traffic (Crowley, 2003). It is called overcrowding when there are more than two people sharing a bedroom or when there are multiple families living in a same property (Kevin S. Blake et al., 2007). Relationships, sleep, stress, and mental health are among areas that can be affected by overcrowding, in addition to the increased risk of infectious diseases (Gove et al., 1979).

Poverty

To determine if a family or individual is poor in the US, their income is compared to a federally established benchmark. If a family of four has an annual income of $26,500 or less, or if an individual's income is $12,880 or less, then that person is deemed poor in 2021. The United States' poverty rate rose to 11.4% in 2020, impacting 37.2 million individuals, after declining for five years in a row (Shrider et al., 2021).

Mental illness, chronic diseases, early mortality, and a decreased life expectancy are all more common among the poor (Singh & Siahpush, 2006) Children are the most vulnerable demographic in terms of poverty (Kingdom U, 2023). Delays in development, toxic stress, chronic disease, and dietary deficiencies are among the many health issues linked to childhood poverty (Eamon, 2001).

Education Access and Quality

Enrollment in Higher Education

Overall, a lower risk of premature death and improved health and well-being can be the outcome of more knowledge. Compared to high school graduates, college graduates self-report greater health. Goesling (2007), Depression, anxiety, diabetes, hypertension, and heart disease are less commonly reported among persons with greater levels of education (Cutler & Lleras-Muney, 2006). One other thing: people with greater knowledge tend to be healthier overall. They exercise more, drink less, and get preventative medical care when they need it (Cutler & Lleras-Muney, 2006).

Health Care Access and Quality

Access to Health Services

Obtaining healthcare is defined as "timely use of personal health services to achieve the best possible health outcomes" by the National Academies of Sciences, Engineering, and Medicine (previously the Institute of Medicine) (Smedley et al., 2003)."

Inadequate health insurance coverage is a major barrier to seeking medical treatment, and health inequities are a direct result of improper coverage distribution (Call et al., 2014). People could put off or even avoid getting the medical treatment they need because of the costs associated with paying for it out of pocket Pryor & Gurewich (2004), both those with and without health insurance, rack up substantial amounts of medical debt (Pryor & Gurewich, 2004). Lack of health insurance is common among low-income individuals (Hadley et al., 2003), Also, minority groups account for almost 50% of the uninsured (Majerol et al., 2015).

Health Literacy

Poor health outcomes and healthcare utilization are correlated with low levels of personal health literacy, which is a social risk factor (Berkman et al., 2011). Those who struggle to understand their own health information are more likely to make mistakes with their treatment or become overwhelmed by the maze of healthcare options available to them.

Organizational health literacy.

The level of service offered and, consequently, health outcomes can be affected by living in places where health care organizations operate but where organizational health literacy is poor. People residing in areas serviced by groups with low health literacy skills may face greater challenges in obtaining assistance due to the increased likelihood of misinterpretation. Low organizational health literacy can have negative consequences for even those with excellent personal health literacy.

Organizational health literacy is still a relatively new idea. There are characteristics of a health-literate organization that go beyond just knowing how to become one (Brach et al., 2012). There are no metrics to measure the level of organizational health literacy in the country, despite the fact that its numerous components have been assessed multiple times (Kripalani et al., 2014). Organizational health literacy research has mostly been descriptive, with little effects documented (Weaver et al., 2012). There has to be more research done on the impact of organizational health literacy.

Neighborhoods and Built Environment

Crime and Violence

Crime and violence can have an impact on everyone, either directly or indirectly, for example, by causing them to witness or hear about property crimes or other forms of violence in their society (Hartinger-Saunders et al., 2012). Although all communities are vulnerable to crime and violence, some are more impacted than others. One example is the persistently higher national homicide rate among Black teens and young adults when contrasted with white kids (Sheats et al., 2018).

Communities with lower incomes tend to have higher rates of crime and property crime compared to those with higher incomes (Kang, 2016). There are many different kinds of violence, including but not limited to: assaults on children and the elderly, violent relationships, sexual assault, and violence with weapons. Those who manage to pull themselves together after becoming victims of violent crimes may endure excruciating agony, emotional distress, and even permanent disability or death. They might also experience mental anguish and a diminished standard of living (Krug et al., 2002). The negative health effects of being in an environment with high levels of crime and violence include, but are not limited to, asthma, high blood pressure, cancer, stroke, and mental disease.

Environmental Conditions

Everyday environmental factors, such as environmental quality and status, can have an effect on human health. Variations in weather, air, and water quality, among other environmental factors, can be substantial between geographical areas and demographic groups. Historical, economic, and sociopolitical factors often impact the quality and impacts of the environment, even if many environmental circumstances are naturally occurring (Schaider et al., 2019).

Drinking, bathing, and cleaning are just a few of the many common places used for water. While most American water supplies are safe to drink, they are susceptible to contamination from things like sewage overflows, some agricultural practices, and even naturally occurring pollutants. Every year, over 7.15 million cases of waterborne diseases caused by microorganisms alone occur in the United States (Kunz et al., 2024).

Additionally, air is essential for the preservation of health and life. Whatever the case may be, between 100,000 and 200,000 Americans die each year as a direct result of air pollution (Tessum et al., 2019). Gases like carbon monoxide, ozone, and nitrogen oxides are examples of air pollutants, whereas dust, smoke, and liquid droplets are examples of air particles. Vehicles, factories, and even wildfires can release these particles into the atmosphere. Deteriorating air quality is linked to a host of health problems, one of which is lung cancer (Turner et al., 2011) and heart disease (Alexeeff et al., 2021). Pollen and other airborne particles can aggravate preexisting conditions like asthma and allergies, which has far-reaching consequences for public health (Saha et al., 2021).

Quality of Housing

Numerous factors, such as affordability, stability, quality, and safety, as well as the surrounding community, influence people's health in relation to housing (Swope & Hernández, 2019). Both mental and physical health may be impacted by the way a home is designed and constructed, which has a big impact on housing quality (Weich et al., 2002). Overcrowding, poor air quality, lead, mould, or asbestos, and other unhealthy conditions can wreak havoc on people's health and increase their risk of developing chronic diseases and injuries (Krieger & Higgins, 2002). For instance, exposure to lead through paint, pipes, and faucets can have permanent negative health impacts (Gostin, 2016). Lead exposure, even at low levels, can have detrimental consequences on children's behavior and health, including the development of their neurological systems and cognitive abilities (Schnoor, 2016).

Social and Community Contexts

Discrimination

The interpersonal and structural impacts of prejudice can affect everyone. Residential segregation is an example of a structural discriminatory policy that limits the "opportunities, resources, and well-being" of marginalized groups (Lukachko et al., 2014). Negative interactions between individuals in their institutional roles (e.g., a healthcare practitioner and a patient) or as private or public persons (e.g., a salesperson and a client) due to personal characteristics (e.g., color, gender, etc.) are examples of individual discrimination. Intentional or inadvertent injury can result from both individual and systemic prejudice, regardless of how the individual perceives it (Luo et al., 2012). As a social stressor, prejudice can have detrimental effects on people's health in the short and long term, and it can even cause physical symptoms like anxiety, heartburn, or irregular pulse (Pascoe & Richman, 2009).

With 31% of American adults reporting at least one significant instance of discrimination in their lives and 63% reporting daily encounters, prejudice is a very typical occurrence (Luo et al., 2012). Although just 8% of American teenagers say they have encountered racial or ethnic prejudice, there is a notable difference between White youths

(2 percent), non-Hispanic Black youths (17.1 percent), and Hispanic kids (11.0 percent) (Sykes et al., 2017). Smoking and other health behaviors that are clearly linked to certain disease outcomes may be associated with discrimination (Corral & Landrine, 2012) or alcohol abuse(Martin et al., 2003). It might also be connected to not engaging in health-promoting activities like using condoms, managing diabetes, and getting screened for cancer (Luo et al., 2012).

Different demographic groupings, especially specific racial/ethnic groups, are impacted by various forms of prejudice Shavers et al. (2012), women Fazeli Dehkordy et al. (2016), lesbian, gay, bisexual, transgender, and queer (LGBTQ) individuals Mays & Cochran (2001), people with disabilities Kirschner et al. (2007), and older adults (Luo et al., 2012).

2.4 Challenges in Non-Clinical Patient Needs

There has been a meteoric rise in the movement to address social issues in hospital settings. Some programs are already in place and widely used, such as those that assess a patient's social needs and put them in touch with the right providers. The possibility for new multi-sector partnerships to tackle social determinants of health and specialized health needs is highlighted by these and other substantial investments from the healthcare business.

Kreuter et al., (2020) explores the expanding corpus of research outlining the connections between social needs and health, as well as the effects of social needs therapies on health usage, costs, and improvement. They also draw attention to information gaps and the need to address implementation issues. They conclude that efforts to modernize social services, improve social safety net policies, and reallocate resources to address social determinants of health can be further advanced through complementary collaborations across healthcare, public health, and social services.

A study Jack et al., (2022) sought to find data on (i) the type and frequency of social and financial loss that patients encountered as a result of postponed surgery, and (ii) any patient evaluation instruments that would be able to gauge or forecast the degree of such harm. The JBI methodological criteria were followed to conduct a rapid scoping review. Medical professionals searched several databases for information pertaining to October 2020, including Medline, psych info, cochrane, the jbi, amed, bni, cinahl, embase, emcare, hmic, and embase. There was a total of twenty-one publications that were considered. Work, leisure and social function, money, patient waiting experiences, and evaluation tools that could direct choices were the five categories into which the data were subdivided.

According to the research, waiting for surgery might cause serious emotional, financial, and social problems for certain individuals. Validated assessment tools are scarce. More studies on patients' experiences with surgical delays are desperately needed to guide a more comprehensive approach to allocating surgical waiting list priority during the COVID-19 pandemic recovery phases.

According to Berry et al. (2018) Clinical support staff and nonclinical staff provide direct patient care. It is the patients' service encounters, particularly the important exchanges with professional staff, that greatly influence their overall perception of their care experience. Ignoring this fact means missing out on opportunities to set patients up for success at every appointment.

Focusing on clinical-support and nonclinical services can help medical practices enhance the overall care they offer in five important ways: (1) Highlighting the significance of exceptional front desk service to establish a positive first impression at every visit; (2) Carefully screening candidates to ensure their values and character match those of the organization; (3) Encouraging all employees, regardless of rank, to consistently learn and grow in order to provide exceptional service; (4) Reducing the frequency of service delivery delays that can greatly impact patients and their families; and (5) Giving priority to the services that patients value most. It is in cancer care that these generalizable medical principles are most clearly demonstrated, as we demonstrate. Professionals in clinical support and nonclinical roles who care for patients at every touchpoint are essential for the optimal performance of any clinical service.

In Manning & Islam (2023) examines a component of routine clinical practice that is rarely formally taken into account yet is crucial to almost every clinical consultation. A large number of non-clinical factors affect clinical decision-making, which in turn impacts medical decisions. Numerous factors can influence healthcare. Some of these include patient-related aspects, such as socioeconomic status, quality of life, expectations, and desires; physician-related aspects, such as personality traits and interactions with the professional community; clinical practice features, such as private versus public practice; and local management policies. Factors that do not directly affect clinical decision-making are compiled from the many fields of expertise in this overview. The most significant obstacle to the actual deployment of evidence-based medicine may lie in this decision-making process. It is necessary to comprehend evidence-based medicine to develop therapeutic strategies that will support its use.

According to several research, public hospitals have a complicated system that makes it difficult to have an efficient patient flow Kriegel et al. (2015); Kriegel et al. (2016) talked about a complicated system that is marked by a lot of parties, a lot of division of labor, and a variety of skilled health professionals along a patient-related performance and value-creation process. The report goes on to identify this intricate patient care framework by Bean et al. (2017) Previous research that isolated the linear flow mechanism of a single patient channel, according to the researchers, did not do justice to the intricate topology of a real-life hospital where hundreds of patients' time-space excursions overlap.

Everyday fluctuation adds another layer of difficulty for the public system. Winasti et al. (2018) and Qin et al. (2017) Talk about how variations in patient flow lead to additional complexity. The intricacy of the hospital organization makes interventions complicated by factors like workloads and resource distribution throughout the entire facility. Kreindler (2017) investigated this idea further and came to the conclusion that difficulties were created by a lack of a logical system-level approach. A decentralized system is reflected in interventions to enhance patient flow that concentrate on specific system components and the patient trip.

Kreindler (2017a) in their explanatory study, they documented three syntheses: (1) efforts have made some progress but have failed to address the system's most pressing problems; (2) with well-defined objectives, regional programs and sites could spearhead reform; and (3) the stateless patient phenomena exemplify the incoherent architecture of the system. The services offered by a well-designed system are suitable for each patient population's demands.

Both covered strategies for enhancing this theme Kreindler (2017) and Kriegel et al. (2015) published two research apiece that expanded on related topics. The majority of studies discovered the need for a whole system approach to address the complexity present in public health systems Kreindler (2017) identified three paradoxes pertaining to patient flow using a thorough interviewing technique. Initiators need a systemic approach that incorporates a patient-centered strategy, a comprehensive evaluation of the patient's treatment trajectory, and the establishment of repeatable procedures if they are to effectively meet the needs of the population they are serving. Kriegel et al. (2015) discovered four controls that can be used to affect the system-level flow of patients. Among these were the following: central patient reception, communication, and hospital case management. Determine the needs for system-level activities by Bean et al.

(2017) employed a data-driven methodology to identify problem locations in the patient flow network. They postulated that actions aimed at improving patient flow would be more effective if they were aware of the systemic areas that needed to be changed. However, this was only a pilot study, and more research is needed.

2.5 Benefits of Uncovering Hidden SDOH Elements in Healthcare Data

"Social determinants of health" (SDoH) refers to the policies, programs, and circumstances that influence an individual's surroundings from the moment of their birth all the way into old age. Systemic variables, such as economic policies and social norms, impact people's environments and behaviours within them; for instance, racism and climate change impact people's health and quality of life. The root causes and variables that contribute to health inequalities must be addressed in order to reduce them. It is possible that SDoH has a more significant effect on health outcomes and quality of life than healthcare expenditure and lifestyle choices alone Amaro (2014); Bradley et al. (2016) In order to monitor, analyze, and execute projects on SDoH, including greenspace renovations, analytical models must be informed. These models are used to evaluate the influence of exposures or treatments on health outcomes (South et al., 2023).

In He et al., (2023) details cutting-edge approaches to improving health equality in populations, communities, and the general public by combining real-world data from sources like electronic health records (EHRs) with social determinants of health (SDoH). Additionally, they noted challenges, successes, and possible solutions. Clinical and public health applications that can benefit from using SDoH in conjunction with real-world data include improving risk stratification, designing public health treatments, and predicting unmet social needs. We did this by looking at data sources and modern informatics methods. The conceptual framework that underpins this opinion review is grounded in the social-ecological model. Not only did we summarize the sources of the data, but we also discovered security holes in the manner in which existing EHR systems collect SDoH data. We also identified opportunities to apply informatics approaches to extract SDoH data from various types of EHR data, including structured and unstructured data, as well as data from public and environmental surveys.

Methods that leverage SDoH for developing individualized treatments, foretelling public health crises, and categorizing illness risk were also discussed, as were newly developed ontologies for standardizing SDH data. Using real-world data with SDoH in public health and clinical applications is essential for the success of both non-technical solutions, such as supportive policies, incentives, and training, and technical solutions, such as innovative social risk management tools integrated into clinical workflow. There is a chance to enhance population health, decrease inequalities, and establish health equality by using SDoH in social risk management, illness risk prediction, and the creation of SDoH-tailored treatments for illness prevention and management.

People reporting food insecurity were 2.4x more likely to report numerous ER trips and 2x more IP visits during a 12-month period, according to a McKinsey survey from 2019. A similar pattern emerged over the course of a year: those whose transportation needs went unfulfilled were twice as likely to report an emergency room visit and 2.2 times as likely to document an IP visit. Using weather data and other community-level SDOH data, we can identify high-risk areas where non-clinical variables might significantly impact health outcomes.

Pollution levels, for instance, have been linked to health issues including respiratory illnesses and even chronic kidney disease in nations like India. It has been established that individuals with asthma, lung conditions, or anemia are more prone to adverse health effects in local environments where suspended particulate matter levels are above a particular threshold. By using this information, at-risk populations can be identified and their disease development can be managed, prevented, and delayed. Healthcare organizations can use social determinants of health to improve preventive care in this way, which would guarantee better delivery of preventive therapies and enhance population health overall. Data on social factors can help population health management programs make more informed operational and financial choices. Designing successful interventions that take into account the resources and needs of the neighborhood can be aided by the insights gathered through SDOH.

Access to compiled, cleaned, synergized, integrated, and easily accessible data for analysis and decision-making is necessary to reap the benefits of SDOH. If we are able to get beyond these obstacles, this will be feasible.

Data availability: A major obstacle to utilizing SDOH to create preventive health models is the availability of thorough, granulated data at the micro level. Compared to behavioral or psychological elements, information about sociodemographic parameters such as income, education, employment, etc. is easier to find. Surveys of direct contacts are the only trustworthy way to get this data, but scaling this approach is difficult.

Data accuracy: Although sociodemographic data is available, it is challenging to demonstrate the correctness of the data. It is crucial that the SDOH data be updated in accordance with a person's personal development. Ensuring data accuracy at all times is essential if health care organisations want to create genuinely effective and individualised care plans for their members.

Data interoperability: It is frequently discovered that many businesses collect pertinent data using various mathematical models or algorithms and screening instruments. This makes it more challenging to build a trustworthy, comprehensive

SDOH database of people and makes it challenging to find significant connections and correlations between social determinants and real health outcomes.

2.6 Role of Non-Clinical Factors in Healthcare Decision-Making

According to Lavelle et al., (2019), The usage of clinical simulation has grown over time as a means of training non-healthcare personnel who work in healthcare settings (such as hospital managers) or regularly interact with clinical populations in the course of their employment (such as police officers). This acknowledges the significant impact these professionals have on patients' experiences receiving healthcare, sometimes serving as their initial point of contact with medical services.

There is no proven tool to assess non-clinical staff members' human factors learning, despite the training's goal of improving the team's communication and coordination skills. We set out to create, deliver, and assess a non-clinical practitioner-specific version of the Human Factors Skills for Healthcare Instrument. This 18-item test takes into account the following human aspects competencies: situational awareness, decision-making, communication, leadership, teamwork, compassion, caring, stress and fatigue management, and situational awareness. 188 individuals who were not employed in healthcare settings participated in the eleven-month (June 2017–April 2018) instrument pre- and post-training pilots while they were trained using mental health simulations.

There were a variety of professions represented among the trainees, including social work (n = 10, 5%), probation (n = 13, 7%), law enforcement (n = 112, 59%), and primary and hospital care administration (n = 53, 28%). The bulk of the subjects were White (n = 144, 77% of the total), and women (n = 110, 59%) were the most numerous. The reliability of one item was low, five items were found to be insensitive to changes (d <.3), and six things were removed from consideration. With a Cronbach's alpha of.93, the remaining 12 items were examined. The results of an exploratory factor analysis showed that 58.3% of the variance could be explained by a single factor. Substantial effect sizes (d >.7) and sensitivity to change (p <.0001) were observed in the final 12-item test following training. According to cluster analysis, those who had lower scores before training showed the most improvement. The Human Factors Skills for Healthcare Instrument Auxiliary version (HuFSHI-A) after training non-clinical populations functioning in healthcare settings is a viable and reliable instrument for evaluating the acquisition of human factors skills. Although the HuFSHI-A was developed and tested in mental health training programs, it can be applied to any training program that considers the importance of collaboration and coordination between clinical and non-clinical staff.

Some studies reported on practice-related NCFs, Linden et al. (1999); Bennett et al. (2010); Linden (1998); Zemanovich et al. (2006); Sharpe et al. (2007); Lee et al. (2009); Betof et al. (1985) which encompassed: 1) practice size; Zemanovich et al., (2006); Betof et al. (1985) 2) percentage of patients with high and low socioeconomic status (SES); Bennett et al. (2010); Lee et al. (2009) 3) percentage of patients with insurance; Bennett et al. (2010); Lee et al. (2009) 4) number of hygienists used; Linden et al. (1999); Zemanovich et al. (2006) and 5) geographic location of or distance to referring periodontist (Betof et al., (1985); Linden, (1998); Linden et al., (1999); Sharpe et al., (2007); Zemanovich et al., (2006)).

Two American studies Betof et al., (1985); Zemanovich et al. (2006) found that compared to GDs working alone or in groups, those working with another GD were more likely to refer patients. However, research examining populations in England and Northern Ireland failed to find a link between the number of referrals and the size of the respective practices (Linden, 1998; Linden et al., 1999). Referrals were lower for practices with fewer insured patients than for those with more insured patients. Patients' SES and insurance status, however, should be taken cautiously as they might have been subjectively judged by GDs.

GDs surveyed by Zemanovich et al. (2006) Compared to practices with one or no hygienist, those with two or more hygienists had a higher referral rate. In a research that contrasted GDs from Northern Ireland and North West England (NWE), Linden et al. (1999) Among NWEs, referral rates were similar to those of the general population; however, NI practices without a hygienist on staff were more likely to send patients elsewhere. How hygienists in practice might influence referrals is not well studied. Using semi-structured interviews, just one study investigated whether hygienists had an effect on referral rates: Sharpe, Durham and Preshaw (2007) An opinion from the hygienist was found to be a commonly related factor for GDs who did not recommend, albeit the specific impact on referral was not immediately apparent.

Zemanovich et al., (2006) found that GD clinics five miles or more away from the closest periodontist had a referral rate about 2.5 times higher than those closer by Linden (1998) and Linden et al. (1999) discovered that fewer patients were referred by practices more than 25 miles from the closest periodontist. According to reports, a practice's likelihood of referring patients decreases with its distance from the periodontist (Sharpe et al., 2007).

2.7 Social Determinants and Health Plan Strategies

Public and corporate institutions now prioritize addressing SDOH. State Medicaid programs and CHIP have implemented care models that engage patients in improving their own well-being, and private health plans are also making efforts to address environmental factors that impact an individual's health. Insurance providers are in a special position to encourage healthier lives for populations of all sizes. Health plans have been tackling SDOH by organizing housing, jobs, education, and food services in addition to providing typical health care services. They have also been supporting additional needs including childcare.

Insurers are recognizing that substantial advantages that enhance individual access and results while reducing overall costs can be obtained by attempting to lessen the adverse effects of SDOH. According to the largest membership organization for health providers, America's Health Insurance Plans (AHIP), many health plans are mapping and documenting current community resources, identifying at-risk populations, and creating programs based on member requirements. For example, one major Medicaid provider found that people with jobs who had previously served time in prison reduced their healthcare costs by half. Certain insurance firms are identifying locations with the worst health outcomes in the US and are developing strategies to help such communities. SDOH data is constantly changing. As a result, AHIP clarified that health plans do not use SDOH data to determine premiums and rates, affect market participation, or refuse care. Two effective programs that addressed the stress of food security and loneliness in the elderly were described by AHIP.

Research has connected stress to an increased risk of cardiovascular disease, obesity, diabetes, depression, cognitive impairment, autoimmune illnesses, inflammatory diseases, and impaired physical and mental mobility with age, making it one of the most detrimental effects of SDOH. A higher risk of poor birth outcomes has been associated with stress, and there have been changes in symptoms of depression, asthma, hypertension, substance addiction, diabetes, and obesity in children. These two instances are only a few of the creative initiatives that health plans are starting to put into place.

The therapeutic effects of Geisinger's Fresh Food "Farmacy" in the treatment of diabetes have been far better than those of pharmaceuticals, which require billions of dollars to create, but at a significantly cheaper cost. Individuals who are "food insecure" are more likely to develop diabetes, be obese, and have worse health because they are unable to consistently obtain wholesome food. Food insecurity affects 14% of the general population and 23% of children in the counties Geisinger services, compared to 12.7% of the US population and 18% of children. Diabetes affects one in eight of these food deprived individuals. In 2016, Geisinger launched Farmacy by searching for its electronic health record (EHR) database for adult patients in certain zip codes with a type 2 diabetes diagnosis and haemoglobin A1c (HbA1c) levels higher than eight, indicating uncontrolled diabetes. Geisinger then used an EHR-connected tool to screen for food insecurity by asking two questions: (1) "In the last 12 months, I/we were concerned that our food would run out before we had the money to buy more," in addition to "In the last 12 months, the food I/we purchased simply didn't last, and we didn't have the money to get more." People who could agree with either of these statements were considered to be food insecure.

Patients who meet these requirements and show interest in the program are directed to an enrolment session where they meet their care team and are given a "prescription" for a healthful, diabetes-appropriate diet. In order to allow patients to receive care and nourishment in one place, Geisinger constructed a food pantry inside one of its clinical centres. The pantry offers the goods, plans, and recipes that patients and their families need to cook two healthful, fresh meals five days a week. Patients received normal diabetic medical care and attended group lessons on diabetes self-management for 15 hours. In addition to continuing case management and health coaching, they also received follow-up with a trained nutritionist and direct medication-management support from a chemist. The medical home model is used to deliver this care, ensuring that participants receive dependable, patient-centered, interdisciplinary, and cooperative treatment (Social Determinants of Health, 2019).

The Importance of Social Determinants of Health Data

Understanding the Domains of SDOH

The first domain, economic stability, directly correlates with health through factors such as income level, employment status, and financial security. People who are well-off monetarily and have steady jobs are more likely to take care of their health and avoid unhealthy habits. On the flip side, financial instability can cause stress, which in turn can limit access to important health resources, which can contribute to worse health outcomes.

The second domain, education, also has profound implications for health. Access to quality education not only equips individuals with knowledge about health practices but also improves employment prospects, thereby enhancing economic stability. Higher levels of education are linked to healthier lifestyle choices and better control of health issues, according to research.

Healthcare access and quality is the third domain, emphasizing the importance of availability and accessibility of medical services. Barriers to healthcare, such as lack of insurance, transportation issues, and inadequate facilities, exacerbate health disparities. This domain underscores the necessity for equitable healthcare systems that ensure all individuals can receive timely and appropriate care.

The fourth domain, neighborhood and built environment, examines how physical surroundings influence health. Safe neighborhoods, access to recreational spaces, and availability of healthy food options contribute to overall well-being. In contrast, neighborhoods plagued by violence or food deserts can lead to heightened stress and poor health outcomes.

Lastly, the social and community context social networks and support systems play a significant role in this domain. While being socially isolated can have negative impacts on one's mental and physical health, having strong community ties and being actively involved in social activities can boost resilience and encourage better behaviours.

The Need for Integrated Solutions

Given the intricate relationship between these domains, there is a desperate need for comprehensive solutions that address SDOH to achieve better health outcomes and reduce healthcare costs. The literature indicates that despite the growing recognition of SDOH, processes associated with these factors often operate in silos or through a fragmented approach. This lack of integration hampers effective interventions and limits the potential for widespread improvements in health equity.

Moreover, the review establishes that the landscape of SDOH is evolving, with increased emphasis from federal initiatives mandating the reporting of specific SDOH elements. This push not only enhances accountability but also opens revenue opportunities for healthcare providers through incentives tied to improved health outcomes. Providers who actively engage in addressing SDOH can benefit financially while simultaneously contributing to the health of their communities.

Toward Health Equity and Value-Based Payment Programs

The proposed solutions to address SDOH represent a promising avenue for achieving health equity, particularly within the framework of value-based payment programs. By aligning financial incentives with health outcomes, these programs encourage providers to adopt a holistic approach to patient care that includes addressing the social determinants influencing health. This not only fosters better individual health outcomes but also promotes a more sustainable healthcare system by reducing unnecessary costs associated with untreated health disparities.

In conclusion, the review of literature reveals a critical need to address the elements of SDOH comprehensively. By recognizing the interconnectedness of the five domains and implementing integrated solutions, we can pave the way for improved health outcomes and greater equity within healthcare systems. The evolving landscape of SDOH, combined with federal initiatives and financial incentives for providers, offers a unique opportunity to transform how health is approached and delivered in our communities.

2.8 Summary

A comprehensive literature review found that social determinants of health (SDOH) significantly impact individual health outcomes. A wide range of factors, including socioeconomic status, educational opportunity, healthcare accessibility, and the physical and local environments, contribute to SDOH. When it comes to people's and communities' health and happiness, all five of these factors matter greatly. Furthermore, the interconnectedness of these domains highlights the complexity of health outcomes and the necessity of addressing them collectively rather than in isolation. The next chapter addresses the research methodology adopted for the study.

CHAPTER III:

RESEARCH METHODOLOGY

3.1 Overview of the Research Problem

Traditional healthcare delivery focuses mainly on clinical treatment without regarding important social determinants of health (SDOH) that strongly affect patient results. Medical organizations increasingly understand how economic conditions alongside social aspects including property accommodation and food access and transportation should be handled in healthcare delivery systems, yet these interventions remain difficult to implement in clinical operations. The Value-Based Payment (VBP) models might solve this issue by encouraging both whole-person and patient-oriented medical services though their execution remains unreliable. The research study investigates the effectiveness of VBP models in solving SDOH to generate better health results and establish health equality.

3.2 Operationalization of Theoretical Construct

The research uses operationalization to transform sociodemographic conceptual factors into analytical variables which allow standardized analysis of health outcomes. The pre-processing steps for USA Social Determinants of Health (SDOH) dataset indicators include mode imputation combined with duplicate removal and label encoding techniques for data points such as age, gender, education level and employment status, income, insurance status and healthcare access and health conditions.

Synthetic Minority Oversampling Technique (SMOTE) performs class balance to improve model accuracy through the normalization techniques implemented with MinMaxScaler. Predictive modeling benefits from Mutual Information which determines the most helpful attributes. The provided dataset underwent an 80/20 split for training and testing purposes, and researchers utilized Decision Tree and XGBoost classifiers because of their predictive accuracy and interpretability abilities. Model tuning reduces overfitting and decision models are evaluated through accuracy values together with precision, recall, specificity, sensitivity and F1-score. This method provides a solid identification process for selecting the best classifier to forecast healthcare results through socio-demographic information that supports data-centric healthcare policy development and intervention design

3.3 Research Design

This study employs a quantitative research method using statistical and machine learning techniques to analyze socio-demographic data for healthcare outcome prediction. The research method for this work is a sequential procedure of data collection, data cleaning, model training, and model evaluation. For social media determinants, obtain the USA Social Determinants of Health (SDOH) dataset which covers information about 20,000 people and their main socio-demographic characteristics. Once the data has been collected proceed to the pre-processing stage.

Imputation methods in this data pre-processing were mode imputation for the categorical data, deletion of duplicated records, and label encoding for the categorical data. SMOTE was applied for data balancing, and feature normalization was performed using MinMaxScaler to standardize data for improved model accuracy. Mutual Information was employed to identify key features most relevant to the target variable. The dataset was then split into an 80/20 ratio for training and testing. For model selection, Decision Tree and XGBoost classifiers were trained, utilizing their respective strengths in interpretability and boosting for enhanced predictive accuracy. Both classifiers were trained and tuned to minimize loss and prevent overfitting through regularization. Model evaluation was based on accuracy, precision, recall, specificity, sensitivity, and F1-score to ensure robust comparative analysis of classifier performance. Through this methodology, aimed to determine the most effective classifier for predicting healthcare outcomes based on socio-demographic factors. The following research design steps and phases are shown in figure

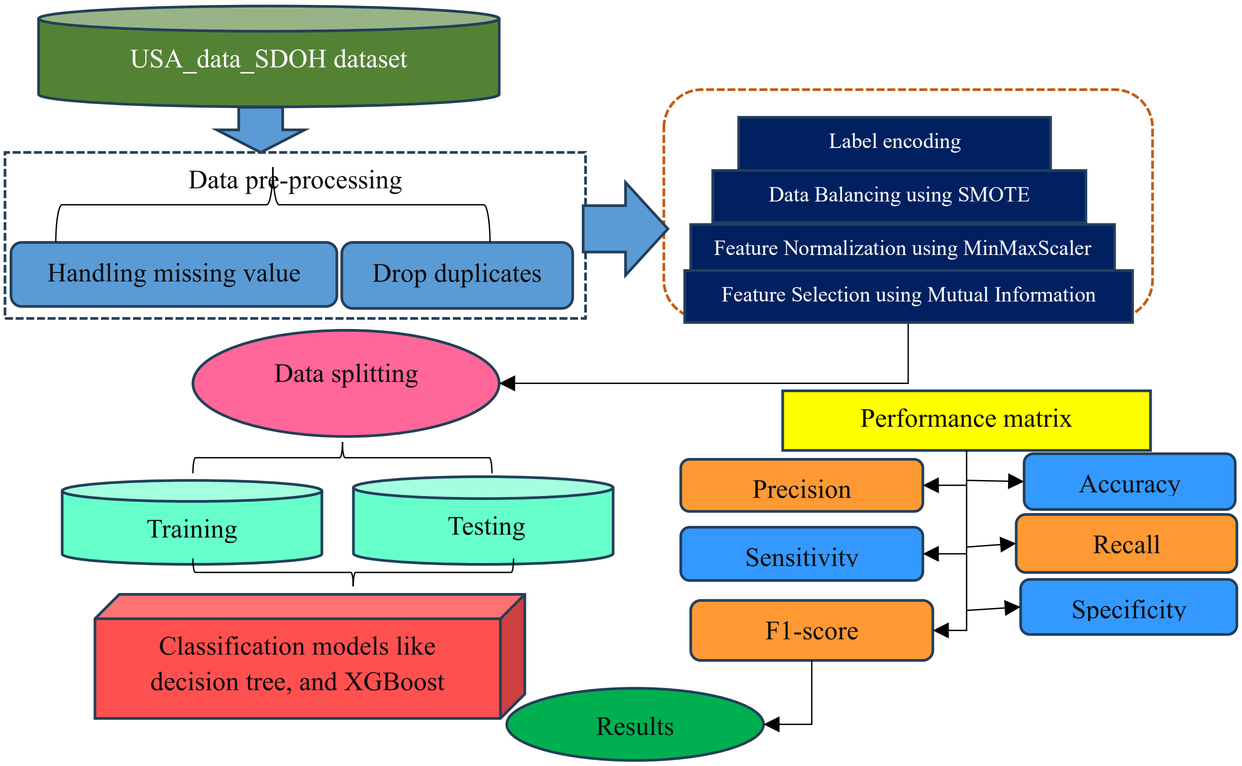

Figure 3.1: Proposed flowchart for social determinants of health in value-based care

3.4 Data collection

The USA Social Determinants of Health (SDOH) Dataset a comprehensive collection of socio-demographic information relevant to healthcare studies and outcomes. It includes 20,000 entries with attributes such as state, city, age, gender, race, healthcare access, education level, annual income, employment status, and housing quality. This dataset captures diverse categories, providing a granular view of the factors affecting health and well-being across the United States.

Data pre-processing

ML relies heavily on data preparation, the significance of which cannot be emphasized enough. To ensure data consistency, first handle missing values by identifying any columns with null entries. For categorical variables, missing values are imputed using the mode, or most frequent value, which preserves the existing distribution without introducing bias. Additionally, duplicate rows are identified and removed, as they can skew the analysis and lead to unreliable outcomes. The encoding technique is used to transform categorical feature values into numerical representations, which is essential for the computer to learn from the data and create an appropriate model. Below is a list of the pre-processing steps:

Label encoding

Label encoding is a method for making numerical values out of category data so that ML algorithms may use them. It is necessary to convert categorical values to numerical representations in order to train a ML model. This will help with model construction. This is accomplished by substituting integers between 0 and (n-1) for categorical values, where 'n' stands for the total number of distinct classes (Talukder et al., 2024).

Data Balancing using SMOTE

A small dataset is ideal for SMOTE's performance. In contrast, SMOTE's efficiency plummets as the dataset size increases since it takes time for SMOTE to generate false data points. In addition, SMOTE has a significant probability of overlapping data points for the minority class while making fake data points.

$$x_{new,attr} = x_{i,attr} + rand(o, 1)\ x\ (x_{ij,attr} - x_{i,attr} \dots \dots \dots (3.1)$$

Feature Normalization using MinMaxScaler

The procedure of bringing the distribution of the independent variables into a more consistent shape is called feature normalization or feature scaling. The data obtained disturbs the mean and the variance, causing a negative effect on the ML model's performance. In order to standardize the gathered data, this research used min-max scaling or normalization, which adds a new dimension to the feature's scale, either between [0, 1] or [1, −1]. From equation 1, the min-max scaler is defined.

$$x' = \frac{X - \min(x)}{\max(x) - \min(x)} \dots \dots \dots \dots (3.2)$$

where:

- x’= scaled sample point
- x = sample point

Feature Selection using Mutual Information

The process of feature selection forms a vital component of solving problems in ML to make a reliable model (Franklin, 2005). Feature selection in this research is based on mutual information. It is also possible to quantify how much information about the dependent variable is described by a given characteristic, often through using mutual information. The use of mutual information allows us to determine which traits are most useful by looking at how dependent they are on the target variable(Barraza et al., 2019). It measures the dependency between two stochastic variables, such as X and Y, which give information about each other. Here is its definition:

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$
$$= H(X) + H(Y) - H(X,Y) \dots\dots\dots\dots\dots (3.3)$$

Here, H (X, Y) means the combined entropy of X and Y, H(X|Y) is the conditional entropy of X given Y, and H (Y|X) means the conditional entropy of Y given X.

Data splitting

The information pre-processing leads to the generation of a training set and a testing set. For the training of the model, the training data constitutes 80 % of the entire data while the testing data constitutes 20% of the entire data.

3.4 Proposed Models

This study explored a variety of ML classifiers to suggest a most promising predictor. Some classifiers, such as the decision tree (DTC), and XGBoost classifier that are explained below:

1) **Decision Tree Classifier**

Predictions are made using a hierarchical structure called a decision tree, which employs a series of feature checks (Song & Lu, 2015). The input characteristics will be represented by X, the decision tree by DT, and the target variable by Y. The goal of the decision tree's recursive feature test dataset splitting is to maximize class separation or minimize impurity. One way to describe the decision tree's forecast is as follows:

$$DT(X) = \sum_{i=1}^{L} y_i . I(X \in R_i) \dots \dots \dots (3.4)$$

Here, L stands for the quantity of decision tree leaf nodes, yi stands for the label given to the i-th leaf node in terms of class, and Ri stands for the area or subset of cases that were allocated to the i-th leaf node according to the feature tests. ($X \in Ri$) is a function that indicates if the input instance X is in the area Ri by returning 1 if it is and 0 otherwise. The instance is assigned the corresponding class label yi at the leaf node of the decision tree, traversing up from a node depending on results of the features tests.

2) **XGBoost Classifier**

The XGBoost is an example of an ensemble tree method where a learner is run repeatedly to mix errors of the algorithms (Chen et al., 2019). The method decreases the residual size through boosting techniques and then fits many trees to the pseudo residuals which are the actual less projected value. This results in enhanced evaluation of classification models and reduced risks such as overfitting. Ada Boost is an iterative process of training a weak learner to focus building stronger classifiers from a set of weak learners, in the direction of the gradient of the loss function of weak learner. The model then uses the fitted trees to provide expected values by:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i) \dots \dots (3.5)$$

The notations Xi is the feature vector for the ith data point while fk is the classification tree. Log Loss is included in its framework specifically for binary classification method. To avoid high model complexity, and thereby over-fitting use of the model, a regularization term is applied. This regularization term is used by the XGBoost method:

$$\Omega = \gamma L + \frac{1}{2}\lambda \sum_{j=1}^{L} w_j^2 \dots \dots \dots \dots .. (3.6)$$

The number of leaves is denoted by L, the degree of regularization by γ and λ, and the score, which may be transformed into probabilities on the jth leaf employing the sigmoid function, is represented by wj 2. A model's objective function is the sum of its loss function and regularization function:

$$obj^{(t)} = \sum_{i=1}^{n} L\left(y_i, \hat{y}_1{}^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \dots .. (3.7)$$

The regularization term is called Ω, and the loss function is called L. The obtained model entails the use of gradient descent in order to reach the maximum of the model’s objective function. It says that gradient descent tries successively to adapt an algorithm in an attempt to get the best minimum of a differentiable function. The XGBoost method has outstanding promise, as stated in the literature, and it is a novel algorithm (Alzubi et al., 2018). In addition to being computationally efficient, the approach works well with heterogeneous data types, as this research demonstrates.

3.5 Performance Matrix

Numerous detection algorithms were applied to the dataset, and their outcomes were evaluated for accuracy and other statistical parameters in order to identify the top-performing method. The measures developed for evaluating the performance of the pattern recognition algorithms include accuracy, precision, recall, specificity, sensitivity and F1.

3.6 Proposed Algorithm

This section gives the suggested algorithm of social media determinants with the assistance of machine learning. All steps of system implementation are displayed in the proposed algorithm.

Proposed Algorithm social media determinants in of health in value-based care

Step 1: Data Collection

- Load the USA Social Determinants of Health (SDOH) dataset.

Step 2: Pre-process the data for data cleaning and make it more efficient.

Step 3: Convert categorical attributes into numerical values through label encoding.

Step 4: To even out the distribution of classes, use SMOTE to generate synthetic samples from the minority group.

Step 4: Normalize data using MinMaxScaler to scale each feature within the range [0, 1], reducing disparities in feature magnitude.

Step 5: Use Mutual Information to select only the most informative features to improve model efficiency and accuracy.

Step 6: Split the pre-processed dataset into training (80%) and testing (20%) subsets.

Step 7: Classification model like decision tree, and XGBoost.

Step 8: To assess the quality of each model's categorisation, compute performance measures including recall, accuracy, specificity, sensitivity, and F1-score.

Step 9: Compare the proposed models with existing models based on evaluation metrics.

3.7 Data Analysis

For the purpose of understanding the data, analysis through SPSS (Statistical Packages for Social Sciences) was conducted. Many types of researchers use this program to analyses complex statistical data. The SPSS software suite was developed for social science data management and statistical analysis. Data miners, government agencies, industries, marketing organizations, survey firms, market researchers, health researchers, and others use it as well. Frequency and correlation analyses were performed using this software. A descriptive statistical technique used in this case is frequency analysis, which displays the quantity of instances of each response selected by the respondents. In contrast, correlation analysis is a statistical technique used in research to calculate the link between two variables and quantify the strength of their linear relationship. In a nutshell, correlation analysis determines how one variable changes as a result of changes in the other one.

3.8 Research Design Limitations

The study adopts a structured methodology but requires several important points to be acknowledged. The USA Social Determinants of Health (SDOH) dataset contains significant information which nonetheless falls short of presenting every element impacting healthcare performance. Data quality improvement occurred through the integration of mode imputation and SMOTE techniques, which achieved a balance between data representation and effective bias management. The research utilizes interpretable and efficient Decision Tree and XGBoost classifiers, but future work will focus on additional deep learning models to reach maximum optimization. The chosen evaluation metrics excel at assessing model performance yet prove insufficient when dealing with real-life healthcare conditions which exceed basic assessment measures. The study presents a solid starting point to study general healthcare practices, yet investigators should verify the results through diverse population-size analyses across various healthcare institutions.

3.9 Conclusion

The study investigates how well Value-Based Payment (VBP) models succeed at managing Social Determinants of Health (SDOH) to enhance healthcare results. Decision Tree and XGBoost classifiers as machine learning approaches provide the study with capacity to analyze socio-demographic data for developing actionable insights. Improved predictive capability stems from mode imputation alongside SMOTE class balancing and Mutual Information-based feature selection which allow the models to detect how SDOH elements affect patient health status.

The study introduces essential database boundaries alongside forecasting model applicability boundaries for making data-based healthcare decisions. The study demonstrates that value-based programs improve patient care and policy development when they adopt socio-demographic information as an integral component. Future investigation should concentrate on expanding the relevant datasets while employing deep learning models and implementing real-time administrative data analysis techniques for wide-ranging healthcare usage.

CHAPTER IV:

RESULTS

4.1 Types of machine learning

1. Supervised learning

Classification and prediction algorithms may be trained using supervised learning, a kind that relies on past instances or outputs. The characteristics and associated predictions or outcomes make up the training set, which is a key differentiator for this learning approach. Supervised learning, in its simplest version, involves taking feature information from a training set and using it to build a model that accurately predicts outcomes in a training set. Then, on a testing set, a model is used to generate predictions using other characteristics.

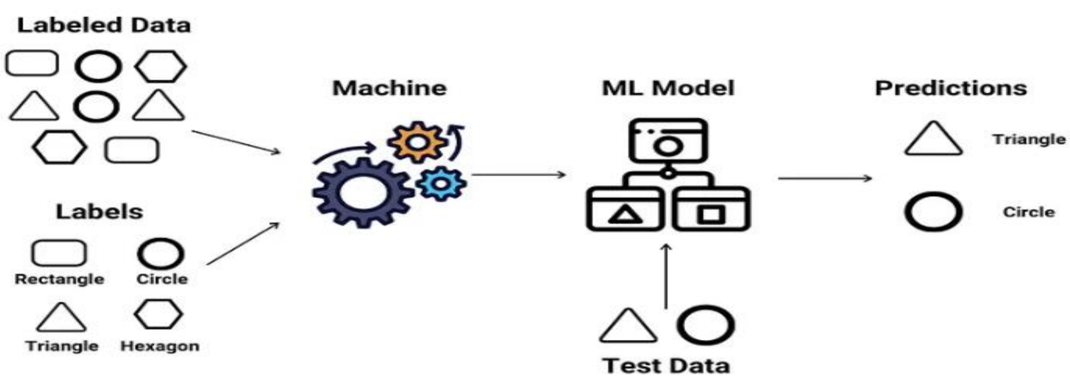

Figure 4.1: Workflow of Supervised Learning

ML algorithms that use supervised learning techniques include DT, RF, SVM, and ANN. A decision-support tool known as a decision tree algorithm takes a single node as its starting point and finds all the potential outcomes of that choice. Once the tree reaches a final product, it stops making choices and starts processing the products of those decisions. SVMs are a kind of supervised learning classification technique that can find features in two categories: problems with data organization and problems with dividing data along the largest margin hyperplane. Each ANN consists of three layers: the input layer, the hidden layer (or layers), and the output layer. In an ANN, each neurone in the layer before and after is linked to every functional unit or neurone in the layer below. To mention a few, supervised ML techniques are often used in the healthcare industry for image detection, hospital outcome identification, and illness prediction.

2. Unsupervised Learning

Unsupervised ML is more often used for data reduction, stratification, and analysis than for prediction. Unsupervised clustering methods, in their simplest form, use algorithms to cluster unlabeled or unclassified data sets autonomously. This technique goes above and beyond the typical ML practice of pre-processing and extracting features from raw data before input. Specifically, it uses feature extraction to find possible data clusters and identifies underlying connections and characteristics (Miotto et al., 2017).

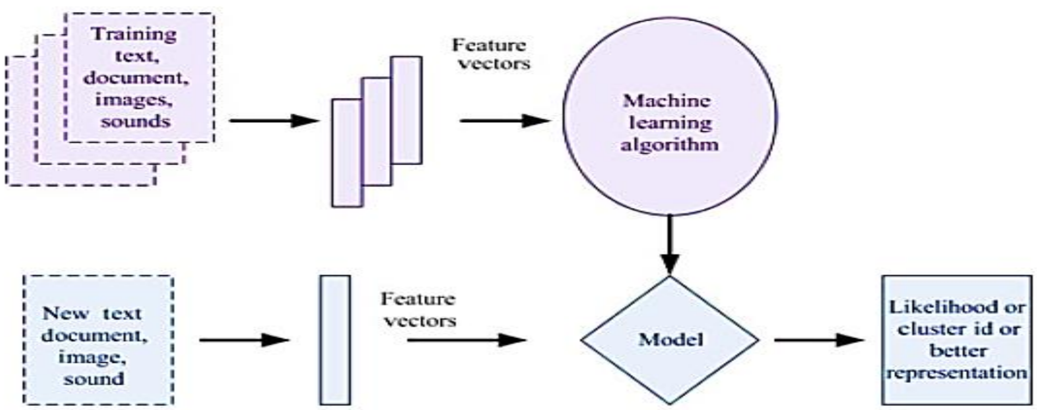

Figure 4.2: Unsupervised learning

Unsupervised learning methods include CNNs, DBNs, and the k-Means algorithm. Among unsupervised learning techniques, the k-Means algorithm stands out as the most common. As a clustering method, it finds the average across groups in unlabeled datasets and uses that average to form new groups. Feature detection and correlation identification are the responsibilities of the many hidden layers that make up a DBN. It usually employs unsupervised learning and consists of intra-level connections that are helpful for data retrieval. CNNs rely on feature recognition and identification to perform functions such as anomaly detection, image recognition, and identification. (Ravi et al., 2017). Despite the speed and efficacy of unsupervised approaches, they are only partly popular in the healthcare business. This is because clustering generally uses algorithms that do not have predefined outputs and deal with data that is not homogeneous.

3. Reinforcement Learning

Reinforcement learning is another strategy for learning that occupies a midway ground. Forming a strategy for operation in a certain issue area, this learning is dependent on reward sequences, like the psychological principles of conditioning. Reinforcement learning techniques are designed to maximize the error criterion, have the ability to affect their surroundings, and have been characterized as the most similar kind of learning seen in both humans and animals (Sutton & Barto, 2005). Given the variety of learning techniques, choosing a learning method is often determined by the implementation goal and is comparatively simpler than choosing an algorithm.

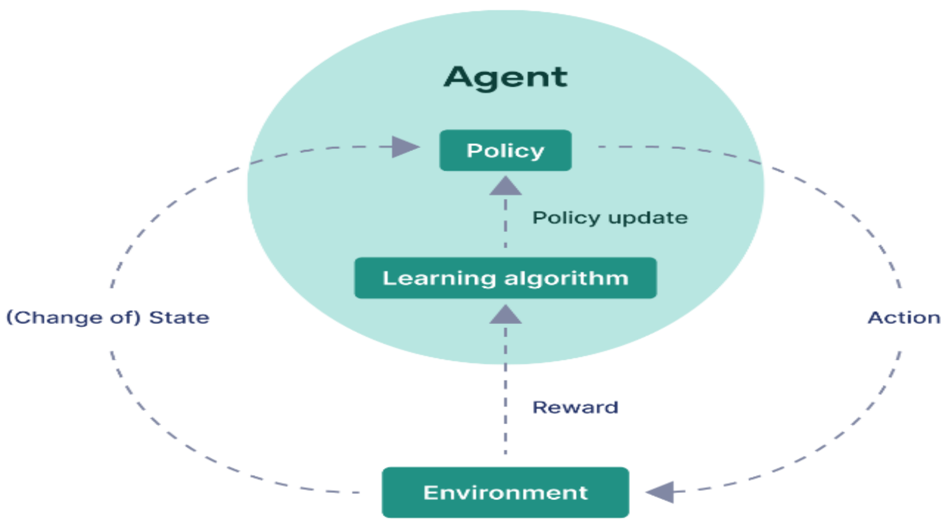

Figure 4.3: Reinforcement learning

The RNN is a popular neural network that employs reinforcement learning. An RNN is one of the NN that has all of the artificial neurons linked to it. The artificial neurons may reuse outputs from earlier stages as input for subsequent steps and can accept inputs with temporal delays. Rhythm learning, music creation, translation, voice recognition, and time series prediction are some of its many applications. Reinforcement learning currently has few healthcare applications because of its requirements for structure, diverse data, reward definition and implementation, and large computer resources. Nevertheless, it still has the potential to make substantial advancements in healthcare.

Finding and using a strategy that is well-suited to the healthcare application is of the utmost importance, considering the many forms of DL and ML techniques. Considerations like as feature count, sample size, and data distributions (Gu et al., 2014), are important because they influence learning and prediction.

4.2 Dataset Description

The dataset "sdoh_usa_20000_rows.csv" contains 20,000 rows of data related to Social Determinants of Health (SDOH) across various states and cities in the United States. The dataset includes several demographic and socioeconomic factors such as age, gender, race, healthcare access, education level, annual income, employment status, and housing quality. The dataset includes key columns such as State and City/Town, which indicate the individual's geographical location; Age, segmented into various age ranges; Gender, including male, female, non-binary, and others; and Race, covering categories like Native American, Asian, Black or African American, among others. It also contains Healthcare Access levels, Education Level ranging from primary to postgraduate, Annual Income grouped into different brackets, Employment Status with categories like employed, self-employed, retired, and unemployed, and Housing Quality classified from very poor to very good.

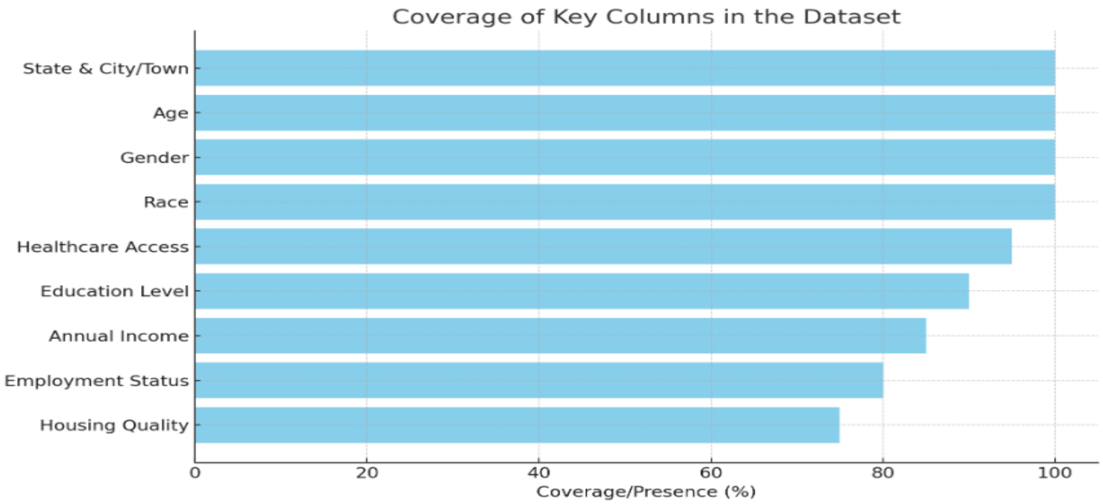

Figure 4.4: Bar chart of key columns in dataset

4.3 Evaluation parameter

Figure 4.5 shows a 2x2 confusion matrix that is used for binary classification. In machine learning, initially, training data was used to train the model, and then we tested its ability to generalize. To put it simply, look at the model's performance when evaluated on unseen data. Based on the kind of problem (regression or classification), we employ assessment metrics to assess the model's performance. The assessment criteria for the categorization issue are addressed here.

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

Figure 4.5: Representation of confusion matrix

- **True Positive (TP):** A TP situation exists when both the actual and expected values are positive.

- **False Positive (FP)**: In an FP scenario, the anticipated value is positive while the actual value is negative.
- **True Negative (TN):** In TN, both the actual and anticipated values are negative.
- **False Negative (FN):** FN occurs when the projected value is negative, but the actual value is positive.

1) Accuracy

The greatest way to compare the outcomes of a model simulation is by looking at their accuracy. The ratio of accurate forecasts to total predictions is the measure of this metric. The following formula is (4.1)

$$Accuracy = \frac{\mathrm{TN} + \mathrm{TP}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \ldots (4.1)$$

2) Recall

A measure of sensitivity or recall, it is illustrated in equation 4.2 as the ratio of real positive cases to the number of true positives. Keep in mind that recall is just the number of true positives discovered (recalled) relative to the total number of true positives.

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \ldots. (4.2)$$

3) Precision

It is the proportion of correct results divided by the sum of correct and false positives, as stated in equation 4.3. A basic definition of precision would be the proportion of detected cases that were actually positive.

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \ldots. (4.3)$$

4) F1 Score

A harmonic means of recall and precision is known as an F1 score, F-measure, or simply F. On a scale from 0 to 1, with 0 being the worst and 1 the best, its value falls somewhere in the middle. The formula for it is (4.4).

$$F1 = \frac{2 * (\text{precision} * \text{recall})}{\text{precision} + \text{recall}} \ldots . (4.4)$$

5) Specificity

A metric that is more relevant in the context of this project is specificity. It is defined as (4.5):

$$specificity = \frac{TN}{FP + TN} \ldots \ldots \ldots . (4.5)$$

6) ROC

Receiver Operating Area In the field of predictive analysis, characteristic curve is among the most popular metrics for evaluation. It shows us how well a model works at various probability levels. The "True Positive Rate" (TPR), also known as sensitivity, and "False Positive Rate" (FPR) are plotted on this graph. The formula to determine FPR is (1-Specificity).

4.4 Description of Demographic Details of Respondents

Table 4.1: Demographic Details

		Frequency	**Percent**
Age	18-24 Years	2294	11.5
	25-34 Years	3211	16.1
	35-44 Years	3184	15.9
	45-54 Years	3206	16
	55-64 Years	3288	16.4
	65 more than Years	4817	24.1
Gender	Male	5045	25.2
	Female	4994	25
	Non-Binary	5077	25.4
	Others	4884	24.4
Race	Asian	3977	19.9
	Black or African American	3975	19.9
	Native American	4085	20.4
	White	3889	19.4
	Others	4074	20.4

Healthcare Access	Very Poor	4041	20.2
	Poor	4039	20.2
	Moderate	4046	20.2
	Good	3979	19.9
	Very Good	3895	19.5
Education Level	Primary	3289	16.4
	Secondary	3393	17
	High School	3377	16.9
	Undergraduate	3361	16.8
	Graduate	3305	16.5
Annual Income (USD)	Below $50,000	4281	21.4
	$50,001-$80,000	3060	15.3
	$80,001-$100,000	2116	10.6
	$100,001-$150,000	5215	26.1
	$150,001-$200,000	5328	26.6
Employment Status	Part-time	3220	16.1
	Employed	3331	16.7
	Self-employed	3407	17
	Unemployed	3329	16.6
	Student	3291	16.5
	Retired	3422	17.1

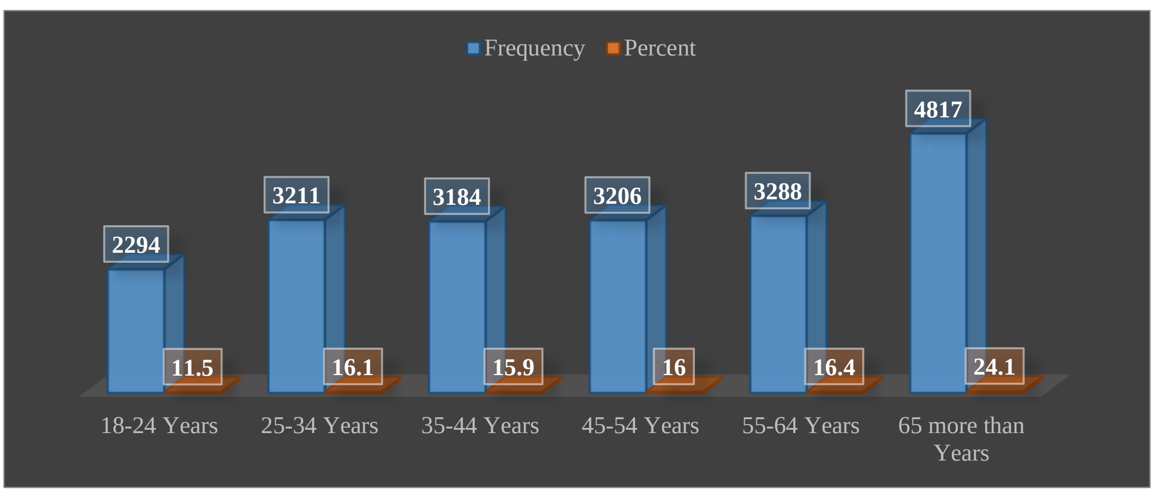

Figure 4.6: Age

Figure 4.6 shows the age distribution of a sample population. Most of the samples, 24.1%, are 65 years or older. For the 55-64 age group, 16.4% suggests a large share nearing retirement. Middle-aged adults are evenly represented by the 45-54, 35-44, and 25-34 age groups, which make up 16%, 15.9%, and 16.1%, respectively. Young adults aged 18-24 make up 11.5% of the population, fewer than older age groups.

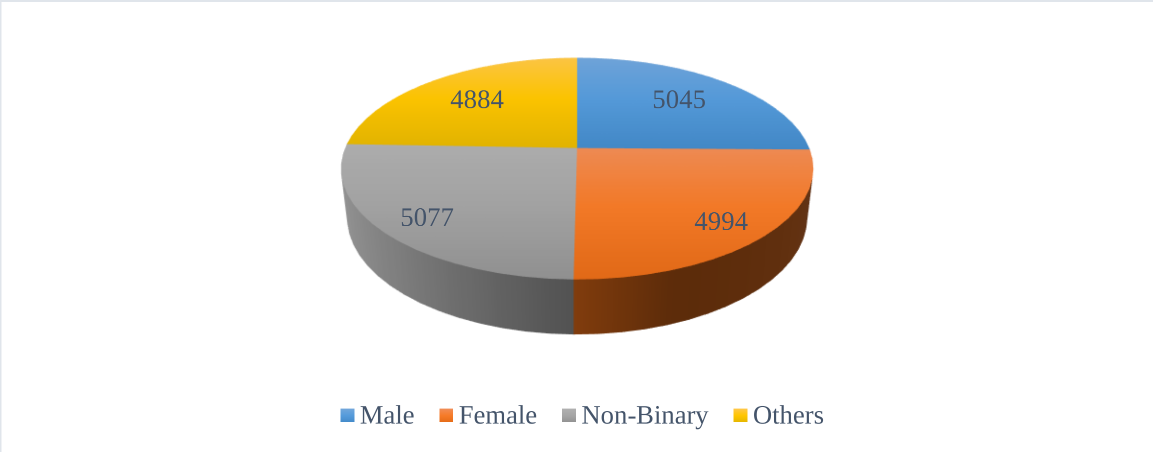

Figure 4.7: Gender

Figure 4.7 indicates a roughly equal gender distribution in the sample population. The majority of the population is non-binary (25.4%), followed by male (25.2%) and female (25%). This gender split shows a fair representation of gender identities, with each group contributing almost equally to the sample.

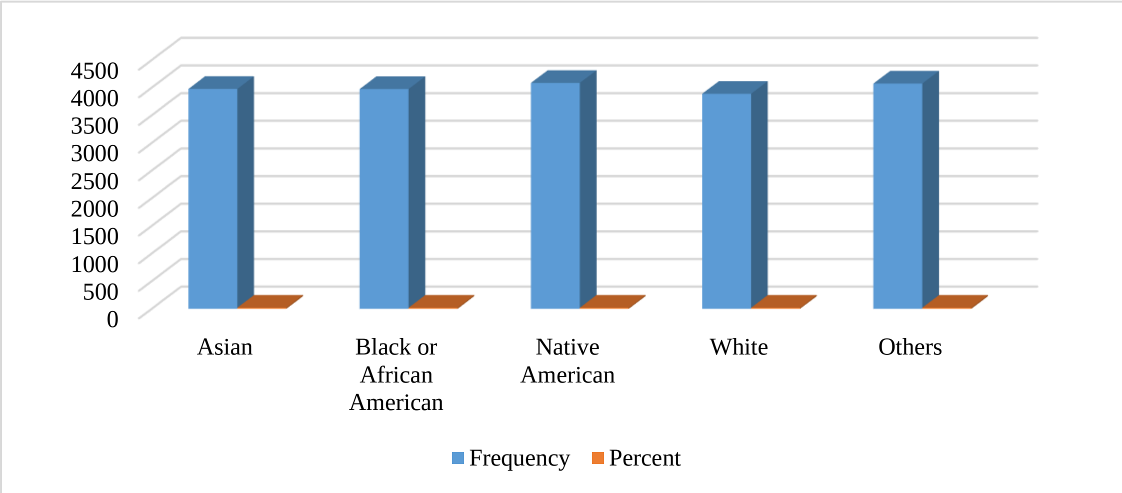

Figure 4.8: Race

Figure 4.8 shows a balanced racial distribution in the sample, with similar frequencies for each race. Largest groups are Native Americans and "Others," each 20.4% of the population with frequencies of 4,085 and 4,074. At 3,977 and 3,975, Asians and Blacks made up 19.9%. White people follow with 19.4% of the sample and 3,889. This distribution reflects a balanced racial sample with no dominant group.

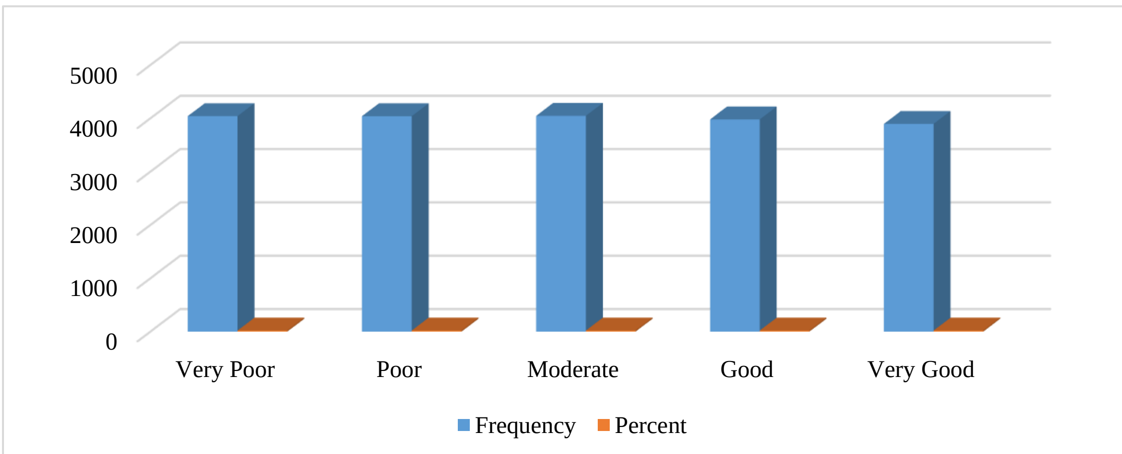

Figure 4.9: Healthcare Access

Figure 4.9 shows that population healthcare access ratings are uniformly distributed among levels. At 20.2% (4,046 people), "Moderate" is the most represented, followed by "Very Poor" and "Poor" at 4.04 and 4,039. At 19.9%, 3,979 people report "Good" healthcare access, while 19.5%, 3,895, have "Very Good" access. This distribution implies that healthcare access is almost equal, with slightly more people reporting lower access than higher access.

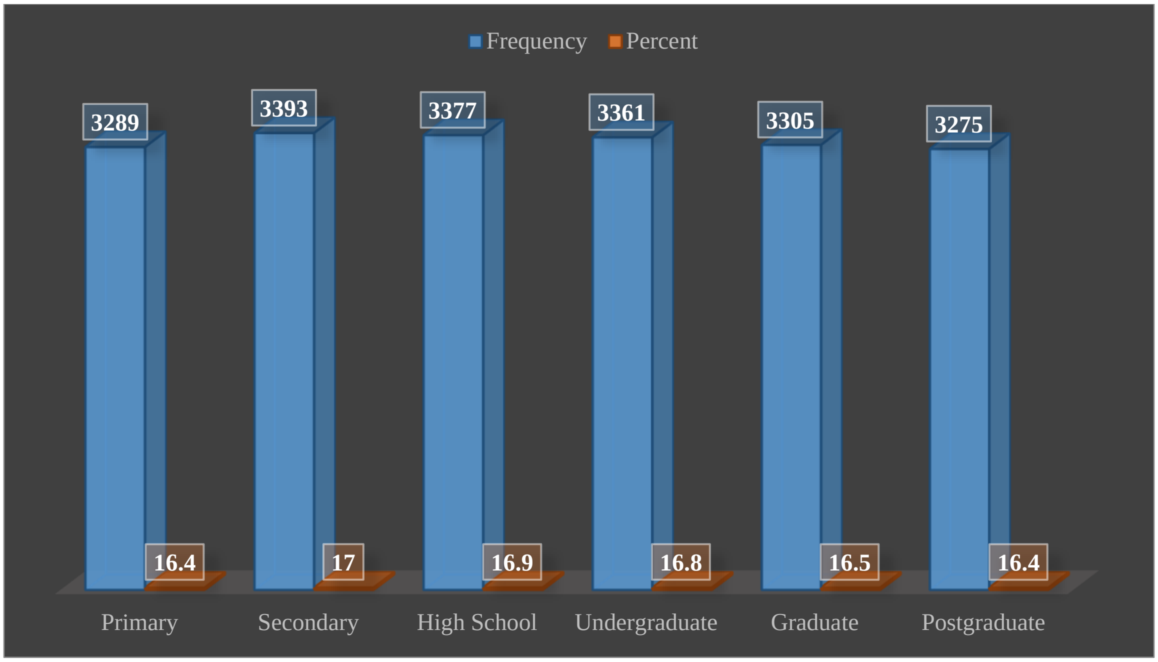

Figure 4.10: Education Level

Figure 4.10 illustrates an equitable distribution of population education levels. The highest presence is 17% (3,393) in secondary education. High school is just behind at 16.9% (3,377 people) and undergraduate at 16.8% (3,361). Graduate education makes up 16.5% of the sample (3,305 people), and primary education is 16.4% (3,289 people). This balanced distribution shows that the sample contains people from a wide range of educational levels, with almost equal representation at each level.

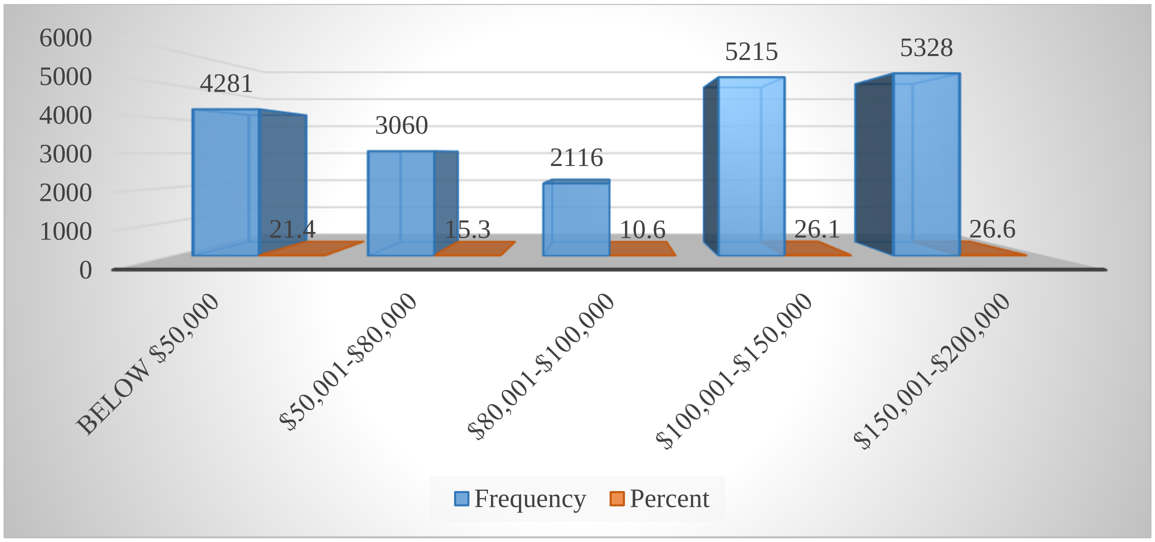

Figure 4.11: Annual Income (USD)

See Figure 4.11 for a diverse population income distribution, with upper-income categories dominating. Earning between $150,001 and $200,000 is the largest category at 26.6% (5,328), followed by $100,001 to $150,000 at 26.1% (5,215). Below $50,000, 21.4% (4,281) are low-income earners. Those earning $50,001 to $80,000 make up 15.3% (3,060), and those earning $80,001 to $100,000 make up 10.6% (2,116). This distribution indicates a high-income concentration and low middle-income percentage.

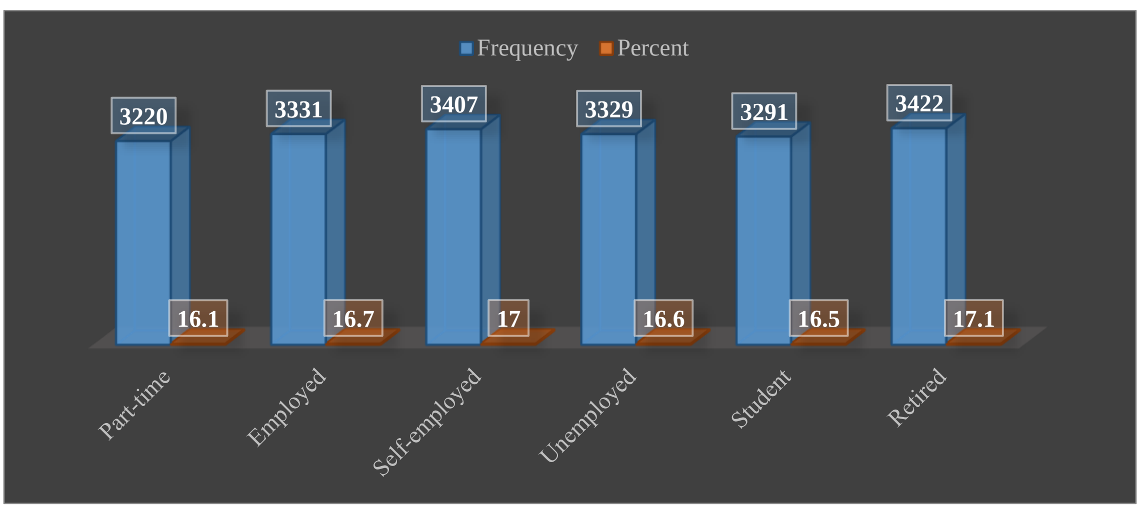

Figure 4.12: Employment Status

Population employment status is quite equitable, as shown in Figure 4.12. The largest group is retired at 17.1% (3,4H22), followed by the self-employed at 17% (3,407). 16.7% (3,331) are full-time workers, whereas 16.6% (3,329) are unemployed. Students make up 16.5% of the population (3,291), and part-time workers make up 16.1% (3,220). This balanced distribution across employment categories reflects a diversified sample, with each group contributing roughly equally to the population.

Table 4.2: Housing Quality

	Frequency	**Percent**
Very Poor	3959	19.8
Poor	4021	20.1
Moderate	4031	20.2
Good	3977	19.9
Very Good	4012	20.1
Total	20000	100.0

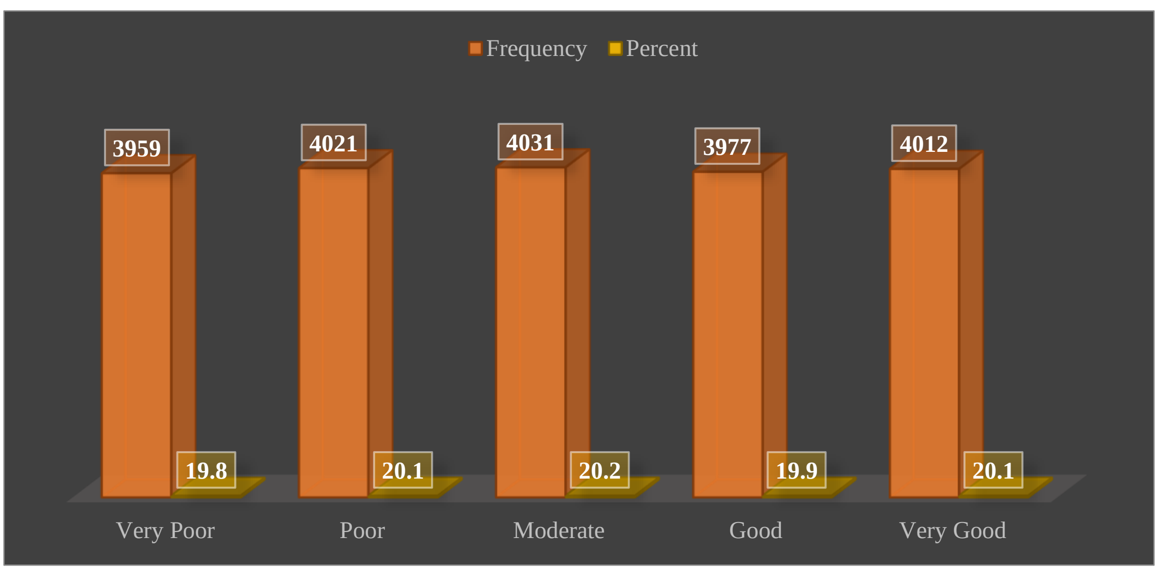

Figure 4.13: Housing Quality

Figure 4.13 shows a rather uniform distribution of dwelling grade categories. "Moderate" dwelling quality is most common at 4,031, 20.2% of the population. Following closely is "Poor" housing quality at 4,021 (20.1%) and "Very Good" housing at 4,012. "Good" dwelling quality is 3,977 (19.9%) and "Very Poor" is 3,959 (19.8%). It appears that the sample population is uniformly distributed throughout all housing quality levels, with no severe bias towards high or low-quality housing.

Table 4.3: Pollution Level

	Frequency	**Percent**
Very Low	3951	19.8
Low	3949	19.7
Moderate	3974	19.9
High	4092	20.5
Very High	4034	20.2
Total	20000	100.0

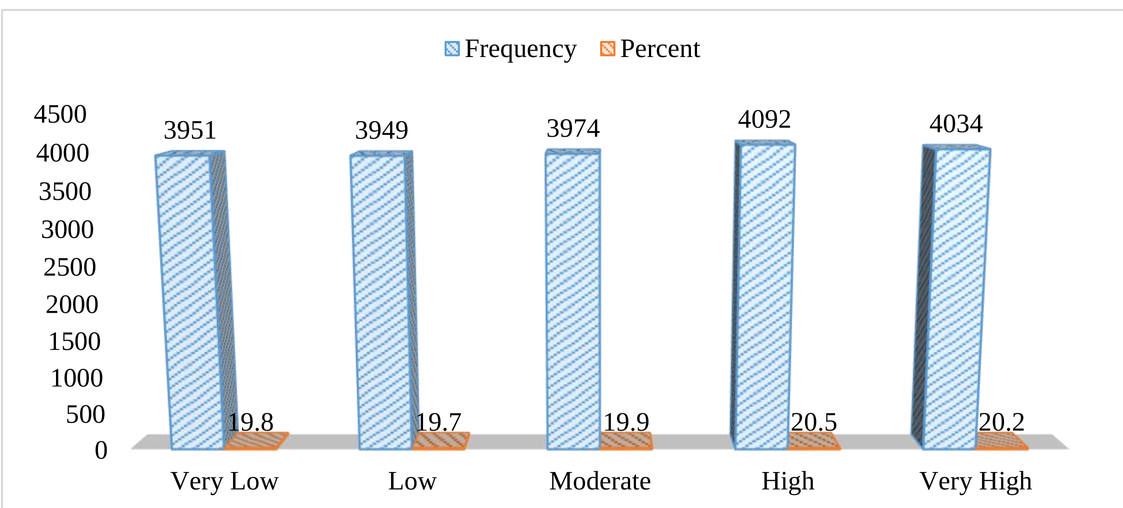

Figure 4.14: Pollution Level

Pollution levels are evenly distributed throughout categories in the figure of 4.14, with each level representing a considerable share of the population. "High" pollution is most common, with 4,092 people (20.5%), followed by "Very High" at 4,034 (20.2%). The population reports 1.9% "Moderate" pollution, 19.8% (3,951) "Very Low" pollution, and 19.7% (3,949) "Low" pollution. This shows that pollution levels are uniformly distributed across the sample, with no category dominating.

Table 4.4: Crime Rate

	Frequency	Percent
Very Low	3966	19.8
Low	4088	20.4
Moderate	4005	20.0
High	3905	19.5
Very High	4036	20.2
Total	20000	100.0

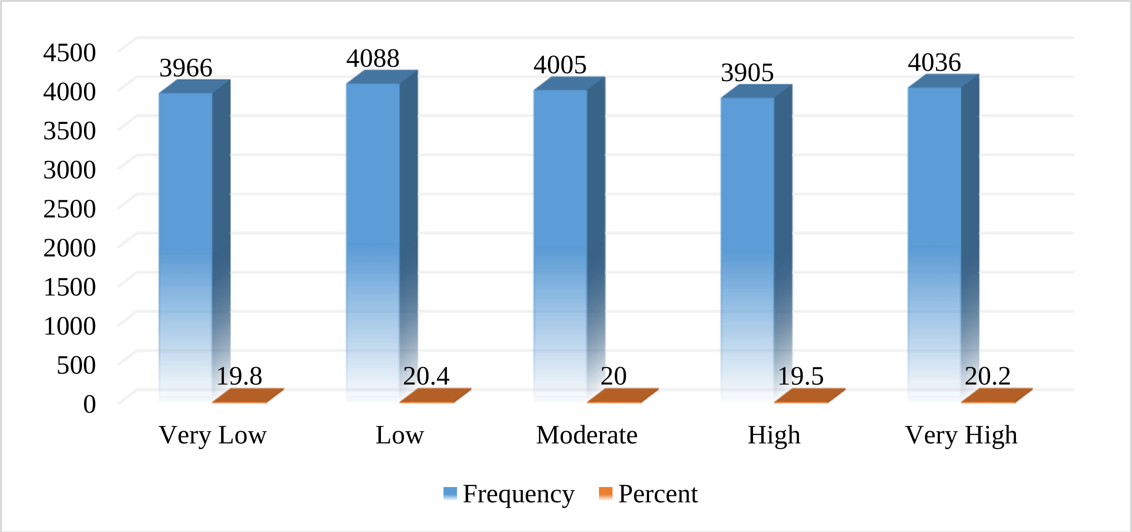

Figure 4.15: Crime Rate

Crime rate figure 4.15 shows a very evenly distributed population throughout crime levels. The highest frequency of "Low" crime is 4,088 (20.4%), followed by "Very High" at 4,036 (20.2%). "Moderate" crime rates are 4,005 (20%), "Very Low" is 3,966 (19.8%), and "High" is 3,905 (19.5%). No group has a disproportionately high crime rate in this population.

Table 4.5: Nutrition Access

	Frequency	Percent
Very Poor	3912	19.6
Poor	3928	19.6
Moderate	4124	20.6
Good	4011	20.1
Very Good	4025	20.1
Total	20000	100.0

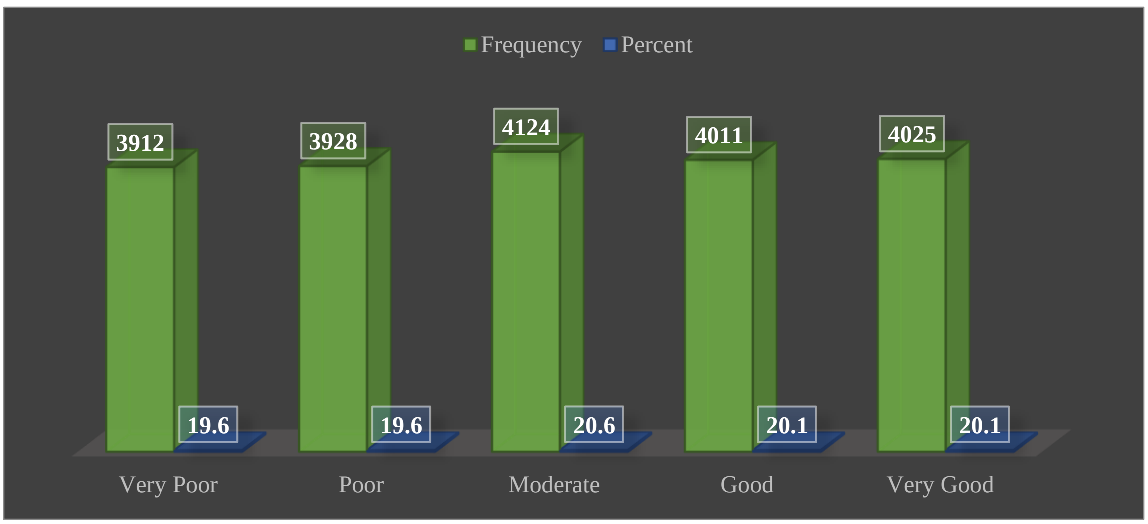

Figure 4.16: Nutrition Access

See Figure 4.16 for a relatively even distribution of nutrition availability levels. The greatest percentage is "Moderate" (4,124 people, 20.6%), followed by "Very Good" (4,025 people) and "Good" (4,011 people). At 19.6%, 3,912 people have "Very Poor" access and 3,928 have "Poor" access. Access to nutrition is balanced, with a slight plurality in "Moderate" and "Very Good" categories.

Table 4.6: Physical Activity Level

	Frequency	Percent
Very Low	3937	19.7
Low	4062	20.3
Moderate	4002	20.0
High	3960	19.8
Very High	4039	20.2
Total	20000	100.0

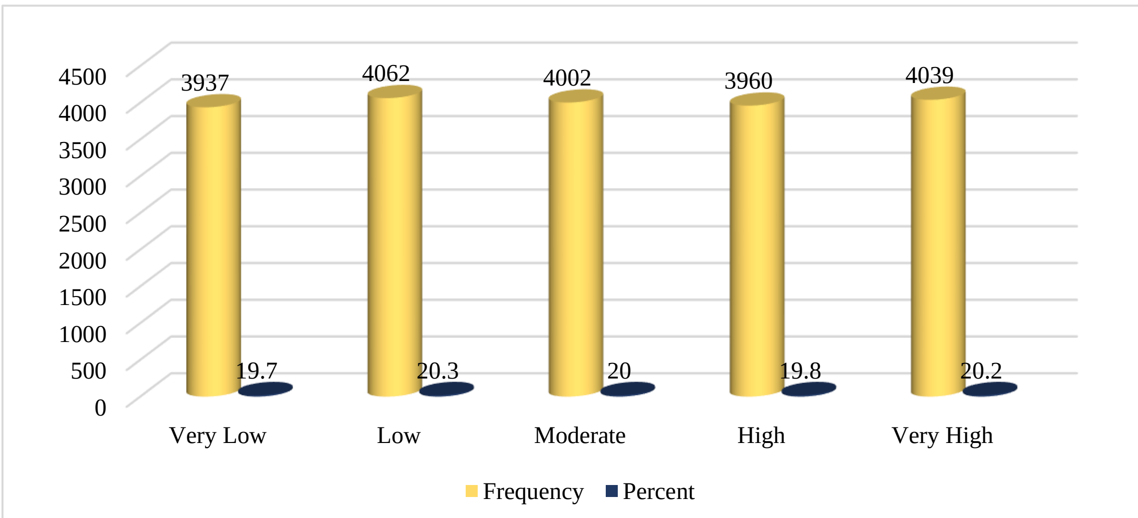

Figure 4.17: Physical Activity Level

The physical activity levels in Figure 4.17 are very evenly distributed across activity groups. "Low" activity is most common with 4,062 (20.3%), followed by "Very High" at 4,039 (20.2%). "Moderate" activity is 4,002 (20%), "Very Low" is 3,937 (19.7%), and "High" is 3,960 (19.8%). This shows that population physical activity levels are balanced, with no major disparities.

Table 4.7: Health Status

	Frequency	Percent
Very Unhealthy	4032	20.2
Unhealthy	4059	20.3

Moderate	3892	19.5
Healthy	4012	20.1
Very Healthy	4005	20.0
Total	20000	100.0

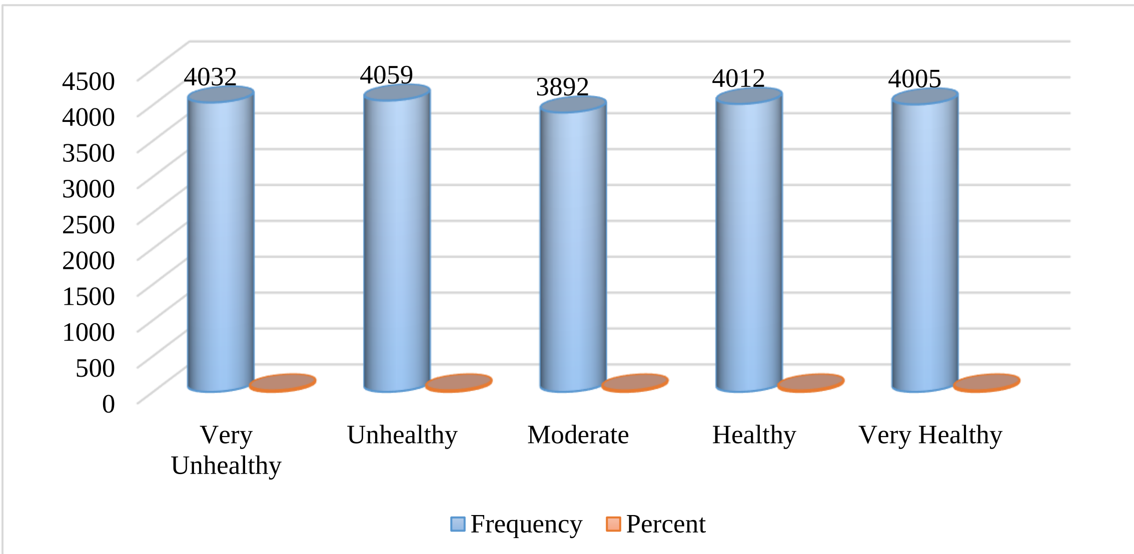

Figure 4.18: Health Status

Health status distribution is quite even in Figure 4.18. There are 4,059 "Unhealthy" people (20.3%) and 4,032 "Very Unhealthy" people (20.2%). Twenty per cent are "Healthy" (4,012), twenty per cent are "Very Healthy" (4,005), and twenty per cent are "Moderate" (3,892). With no category outnumbering the others, this distribution reflects a balanced health status representation.

Table 4.8: Diet

	Frequency	Percent
Very Unhealthy	3936	19.7
Unhealthy	3946	19.7
Moderate	4107	20.5
Healthy	4032	20.2

Very Healthy	3979	19.9
Total	20000	100.0

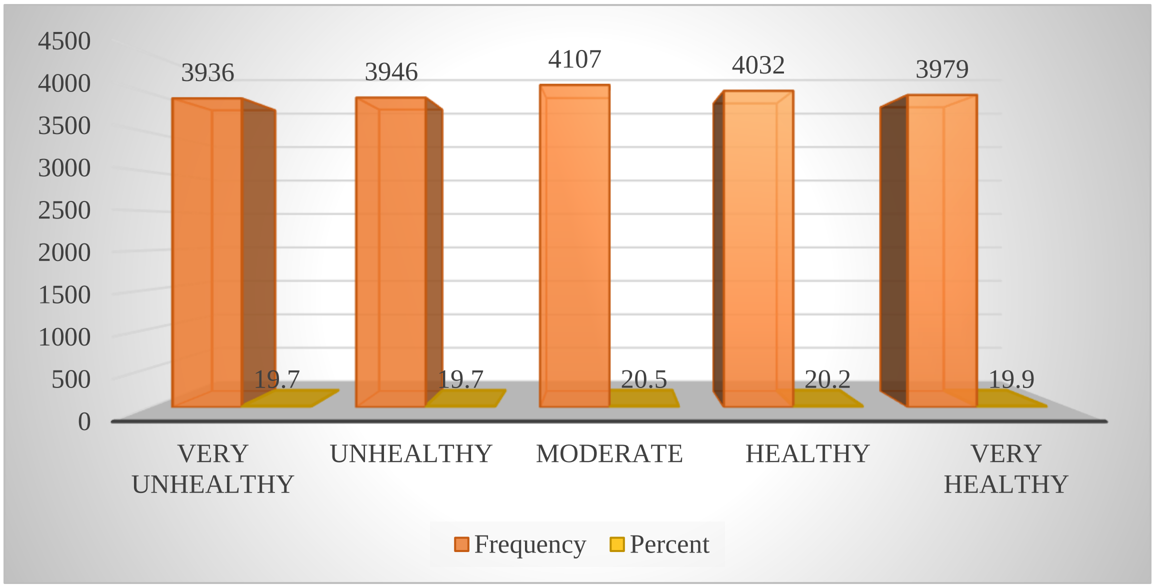

Figure 4.19: Diet

Figure 4.19 on diet quality demonstrates an equitable distribution across categories. The most common diet group is "Moderate" with 4,107 (20.5%), followed by "Healthy" at 4,032 (20.2%). 19.7% of people have "Very Unhealthy" or "Unhealthy" diets, with 3,936 and 3,946 frequencies, respectively. 19.9% (3,979) eat "Very Healthy". This distribution implies that the population's diet is balanced, with a minor edge in "Moderate" and "Healthy".

Table 4.9: Exercise Level

	Frequency	Percent
Frequent	3909	19.5
Occasional	4087	20.4
Rare	4034	20.2
Regular	3926	19.6
None	4044	20.2

Total	20000	100.0

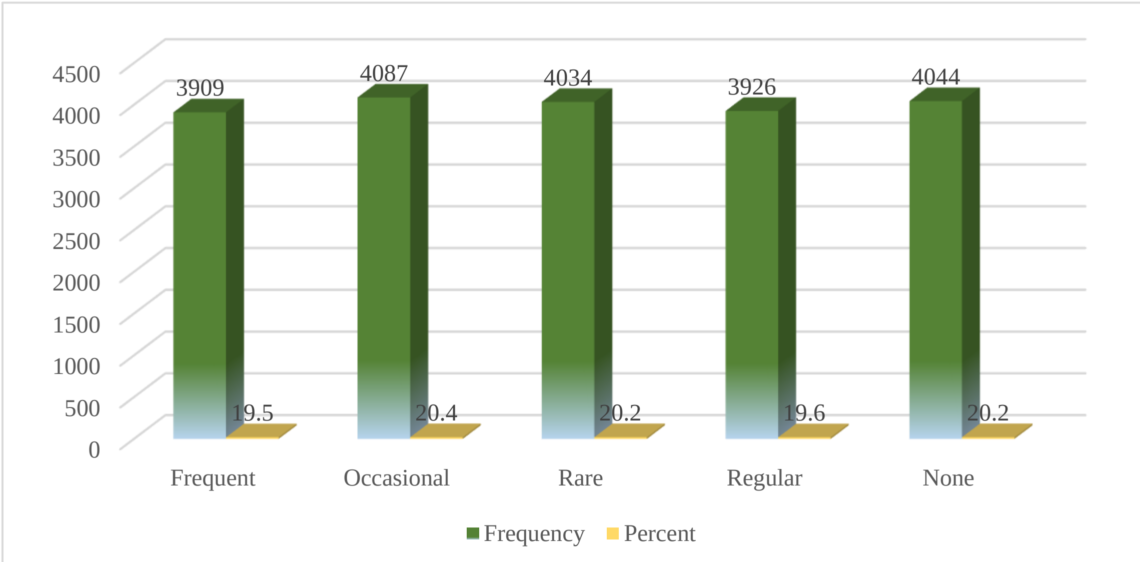

Figure 4.20: Exercise Level

Figure 4.20 on exercise levels demonstrates a balanced distribution across categories. The most frequent activity level is "Occasional" with 4,087 (20.4%), followed by "None" at 4,044 (20.2%). "Rare" and "Regular" exercise make up 20.2% (4,034) and 19.6% (3,926). The "Frequent" exercise category includes 19.5% of people (3,909). This distribution shows that no category dominates the sample's exercise habits.

Table 4.10: Smoking Habit

	Frequency	Percent
Yes	5018	25.1
No	4989	24.9
Frequent	4967	24.8
Occasional	5026	25.1
Total	20000	100.0

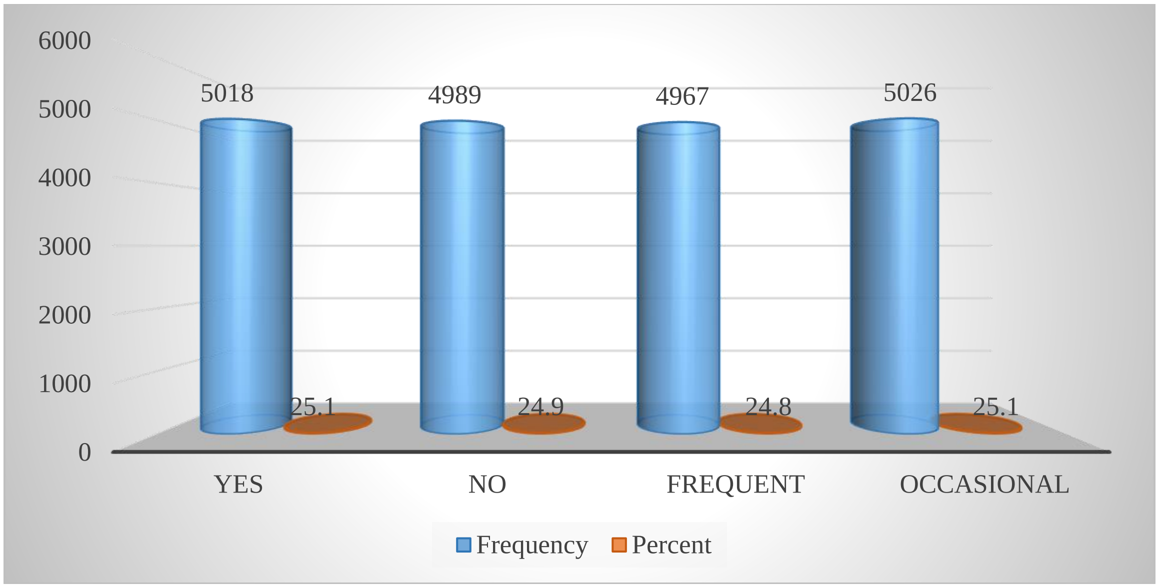

Figure 4.21: Smoking Habit

Figure 4.21 demonstrates a relatively even distribution of smoking tendencies. The "Occasional" smoking category had the highest frequency at 5,026 (25.1%), followed by "Yes" at 5,018 (25.1%). "Frequent" smokers are 4,967 (24.8%) and "No" smokers are 4,989 (24.9%). This distribution shows that smoking and non-smoking are evenly distributed throughout the population.

Table 4.11: Alcohol Consumption

	Frequency	Percent
Yes	5005	25.0
No	4963	24.8
Frequent	4981	24.9
Occasional	5051	25.3
Total	20000	100.0

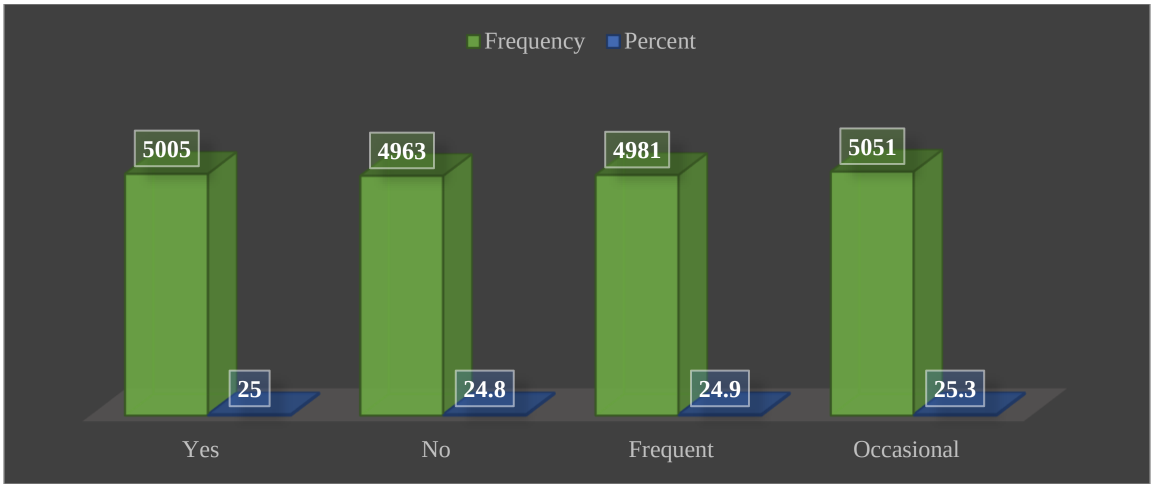

Figure 4.22: Alcohol Consumption

In Figure 4.22, alcohol consumption is distributed nearly equally across groups. "Occasional" drinkers make up 5,051 (25.3%), followed by "Yes" drinkers, 5,005 (25%). 4,981 people (24.9%) are "frequent" drinkers, while 4,963 are "No" drinkers (24.8%). This implies that alcohol consumption is evenly distributed among the population, with similar proportions of consumers and non-consumers.

Table 4.12: Stress Level

	Frequency	Percent
Very Low	3972	19.9
Low	4090	20.5
Moderate	4002	20.0
High	3968	19.8
Very High	3968	19.8
Total	20000	100.0

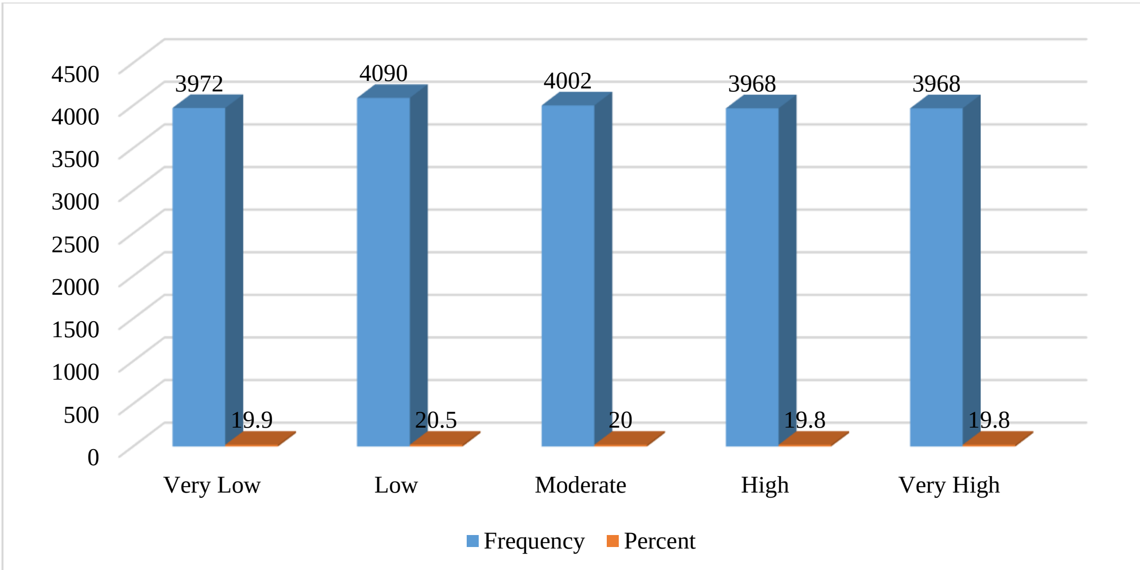

Figure 4.23: Stress Level

According to Figure 4.23, stress levels are evenly distributed throughout categories. The "Low" stress group comprises 4,090 people (20.5%), followed by the "Moderate" stress at 4,002 people (20%). The population has 19.9% (3,972) "Very Low" stress and 19.8% (3,968) "High" stress. About 19.8% (3,968) are "Very High" stressed. It appears that stress levels are evenly distributed across the sample, with no category dominating.

Table 4.13: Socioeconomic Status

	Frequency	Percent
Low	5068	25.3
Middle	4941	24.7
Upper Middle	5068	25.3
High	4923	24.6
Total	20000	100.0

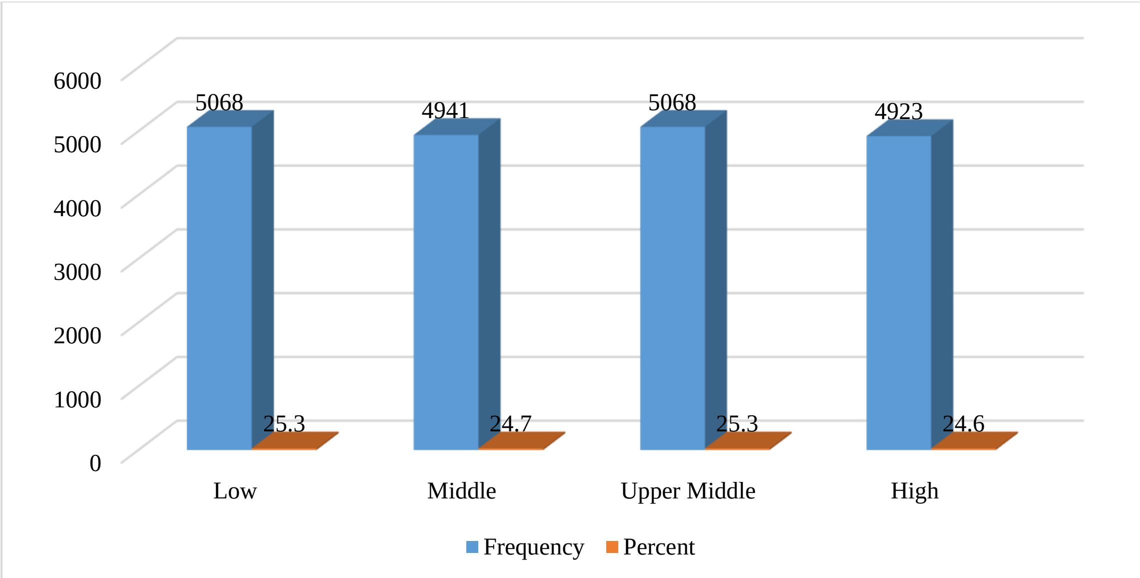

Figure 4.24: Socioeconomic Status

In Figure 4.24, socioeconomic status is distributed evenly across categories. Each of the "Low" and "Upper Middle" groupings has 5,068 people or 25.3% of the population. Our "Middle" class has 24.7% (4,941 people) and our "High" class has 24.6% (4,923 people). There is no dominant socioeconomic group in the population, according to this.

Table 4.14: Conditions

	Frequency	Percent
Very Poor	4064	20.3
Poor	3971	19.9
Moderate	4004	20.0
Good	4065	20.3
Very Good	3896	19.5
Total	20000	100.0

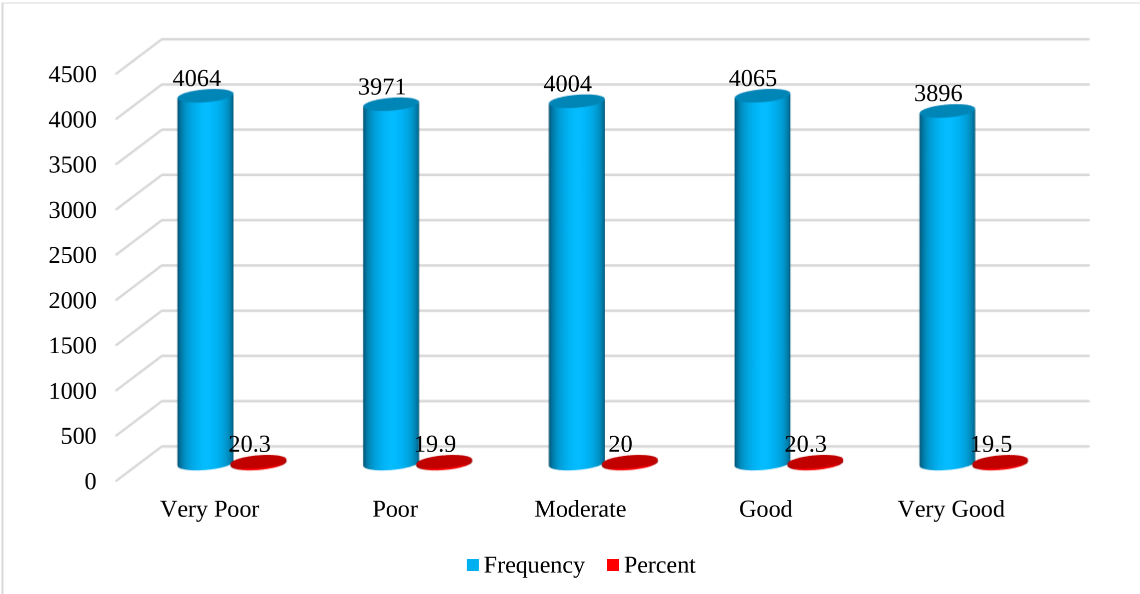

Figure 4.25: Living Conditions

Living circumstances Figure 4.25 indicates a balanced distribution across categories. Twenty.3% of the population, 4,064 and 4,065, are "Very Poor" or "Good" living conditions. "Moderate" living conditions make up 4,004 (20%), and "Poor" is 19.9% (3,971). Of the population, 19.5% (3,896) live in "Very Good" conditions. It appears that no category dominates the sample's living conditions.

Table 4.15: Sanitation

	Frequency	Percent
Very Poor	3923	19.6
Poor	4127	20.6
Moderate	4038	20.2
Good	3998	20.0
Very Good	3914	19.6
Total	20000	100.0

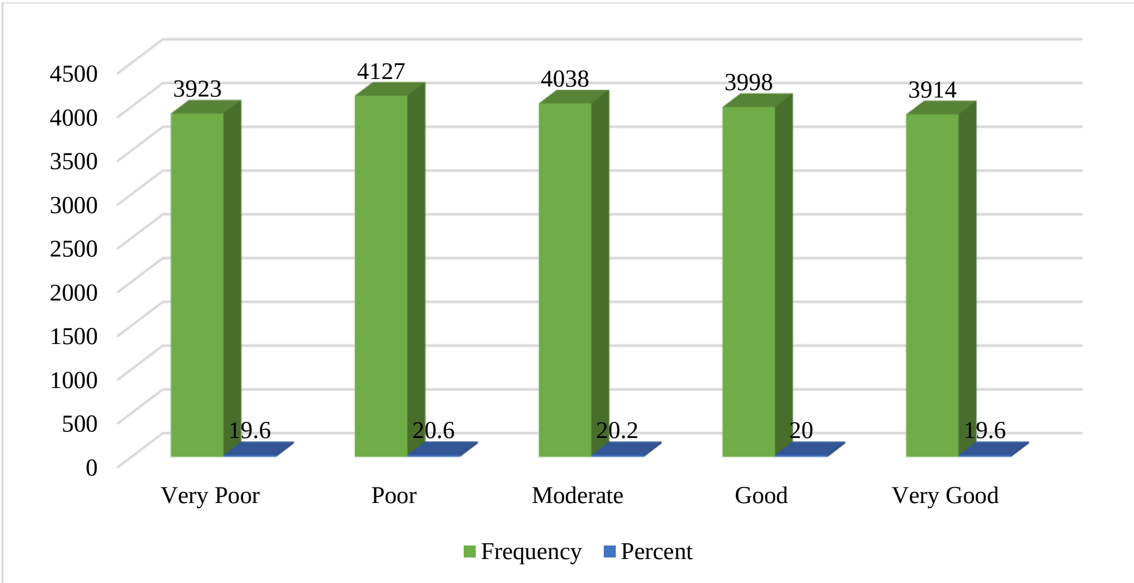

Figure 4.26: Sanitation

Figure 4.26 shows sanitation conditions somewhat evenly across categories. The largest group is "Poor" sanitation with 4,127 (20.6%), followed by "Moderate" at 4,038 (20.2%). "Good" sanitation is 20% of 4,000 people, while "Very Poor" and "Very Good" are 19.6% (3,923 and 3,914). This shows that sanitation conditions are fairly distributed across the population.

Table 4.16: Environmental Hazard Exposure

	Frequency	Percent
Very Low	4102	20.5
Low	4087	20.4
Moderate	4028	20.1
High	3892	19.5
Very High	3891	19.5
Total	20000	100.0

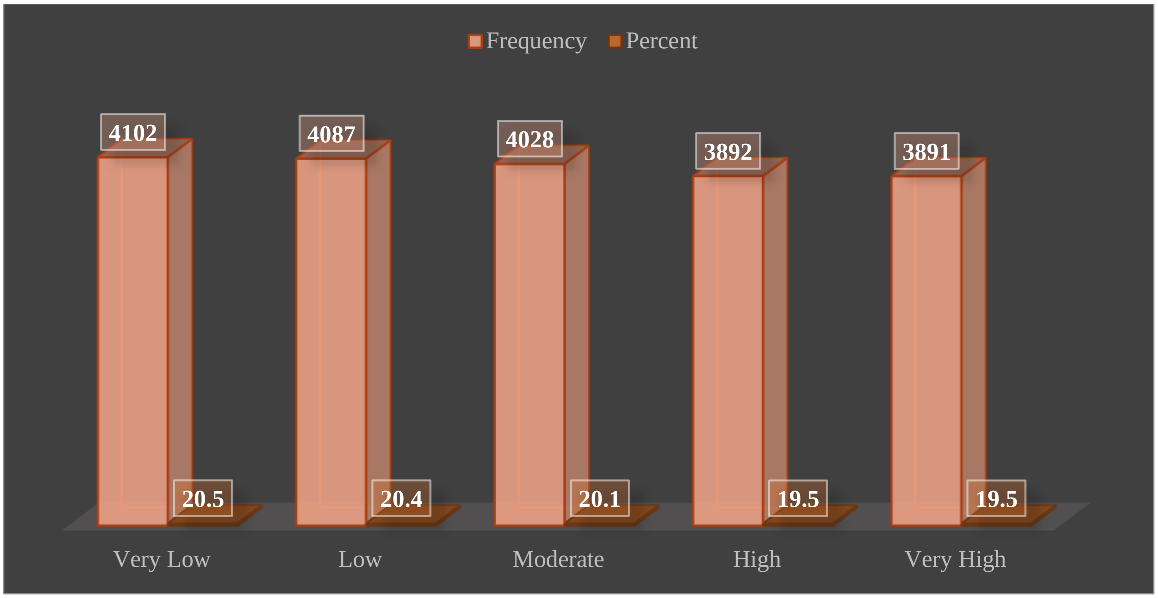

Figure 4.27: Environmental Hazard Exposure Environmental Hazard Exposure

The distribution of environmental hazard exposure levels is approximately even in figure 4.27. The highest frequency is "Very Low" (4),102 (20.5%), followed by "Low" (4),087 (20.4%). "Moderate" exposure totals 4,028 (20.1%), while "High" and "Very High" exposure total 19.5%, with 3,892 and 3,891 persons, respectively. This shows that environmental danger exposure is evenly dispersed, with no major concentrations.

Table 4.17: Access to Healthcare

	Frequency	Percent
Very Difficult	3993	20.0
Difficult	4111	20.6
Moderate	3953	19.8
Easy	3991	20.0
Very Easy	3952	19.8
Total	20000	100.0

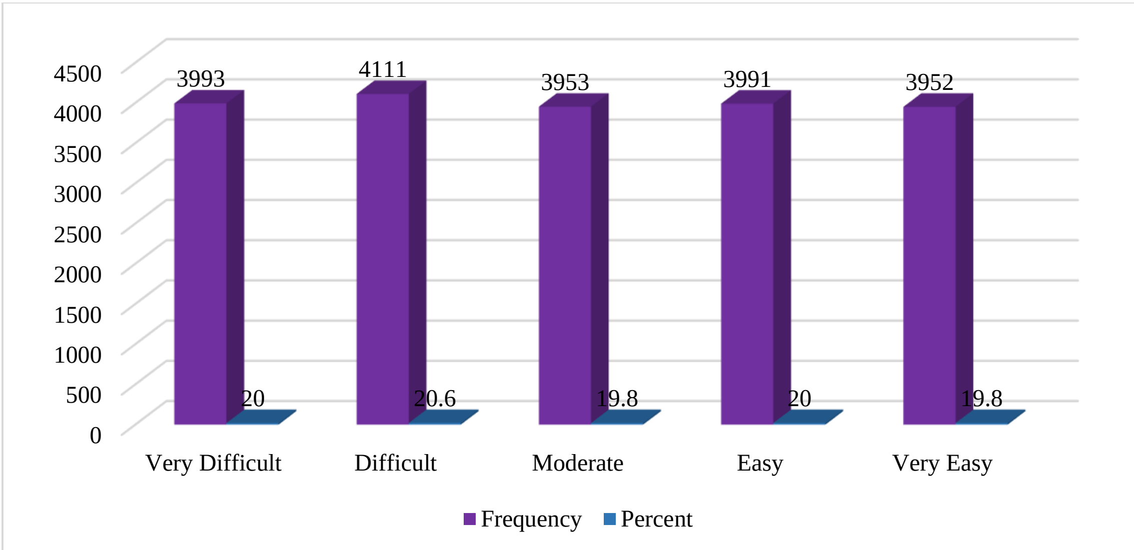

Figure 4.28: Access to Healthcare

Figure 4.28 presents a very even range of healthcare accessibility levels. "Difficult" is the most common category at 4,111 (20.6%), followed by "Very Difficult" at 3,993 (20.0%). 3,991 people (20%) have "Easy" healthcare access, while 3,953 and 3,952 have "Moderate" and "Very Easy" access, respectively. No category dominates healthcare access in the sample.

Table 4.18: Quality of Care

	Frequency	Percent
Very Poor	3946	19.7
Poor	4012	20.1
Moderate	4027	20.1
Good	4006	20.0
Very Good	4009	20.0
Total	20000	100.0

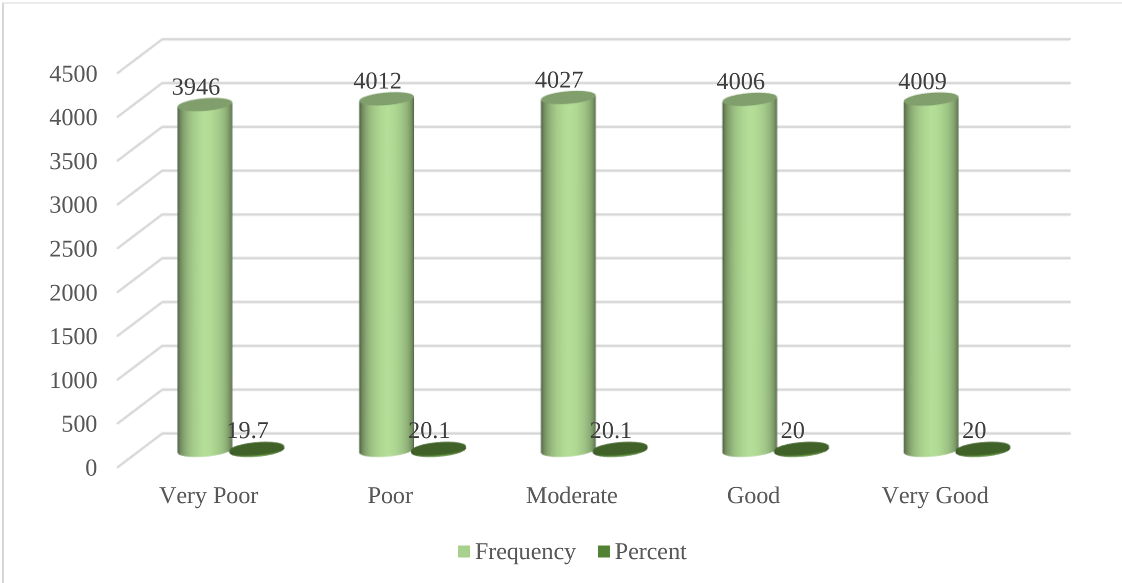

Figure 4.29: Quality of Care

Above figure 4.29, care quality is evenly spread throughout tiers. After "Poor" quality care, "Moderate" and "Very Good" quality care follow with 20.1% (4,027) and 20% (4,009), respectively. "Very Poor" care covers 3,946 people (19.7%) and "Good" care 4,006 (20%). This shows that no category dominates the population in terms of care quality.

Table 4.19: Cost of Care

Cost of Care		
	Frequency	Percent
Very Affordable	4113	20.6
Affordable	3906	19.5
Moderate	3941	19.7
Expensive	4065	20.3
Very Expensive	3975	19.9
Total	20000	100.0

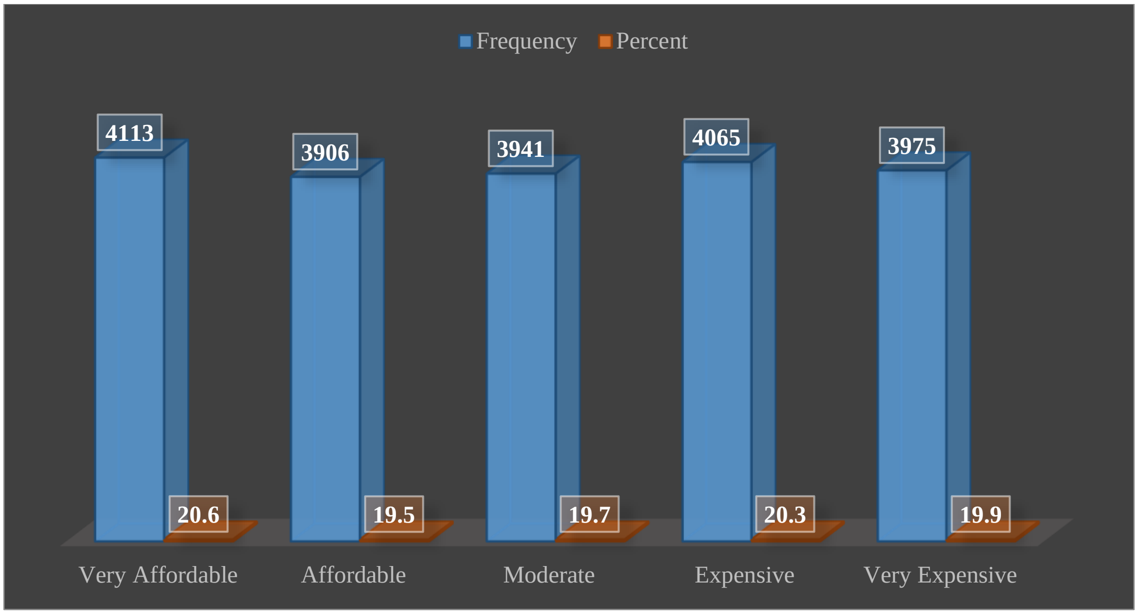

Figure 4.30: Cost of Care

Figure 4.30 indicates a balanced distribution of care costs across affordability groups. The most common category is "Very Affordable" at 4,113 (20.6%), followed by "Expensive" at 4,065 (20.3%). The population is 19.5% "Affordable" (3,906), 19.9% "Very Expensive" (3,975), and 19.7% "Moderate" (3,941). This shows that the population's expense of care is evenly distributed, with no group heavily outnumbered.

Table 4.20: Health Policies Awareness

	Frequency	Percent
Unaware	5047	25.2
Aware	5019	25.1
Partially Aware	4929	24.6
Very Aware	5005	25.0
Total	20000	100.0

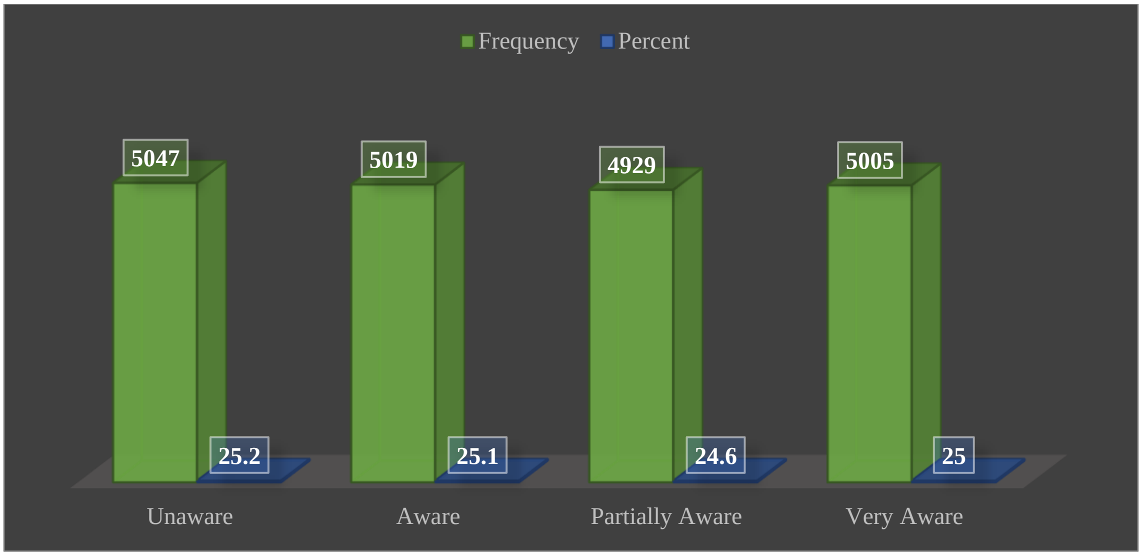

Figure 4.31: Cost of Care

Figure 4.31 shows health policy awareness is evenly distributed across knowledge levels. The largest group is "Unaware" (5,047 people, 25.2%), followed by "Very Aware" (5,005 people, 25%). 25.1% (5,019) are "Aware" and 24.6% (4,929) are "Partially Aware". Health policy awareness is evenly dispersed, with no group notably larger than the others.

Table 4.21: Descriptive Statistics

	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Age	3.82	1.712	-.175	.017	-1.270	.035
Gender	2.49	1.115	.008	.017	-1.353	.035
Race	3.01	1.415	.001	.017	-1.297	.035
Healthcare Access	2.98	1.410	.016	.017	-1.292	.035
Education Level	3.49	1.700	.010	.017	-1.258	.035
Annual Income (USD)	3.21	1.513	-.272	.017	-1.422	.035
Employment Status	3.52	1.704	-.005	.017	-1.262	.035
Housing Quality	3.00	1.412	.000	.017	-1.296	.035
Pollution Level	3.02	1.414	-.019	.017	-1.300	.035
Crime Rate	3.00	1.414	.011	.017	-1.300	.035
Nutrition Access	3.02	1.409	-.015	.017	-1.286	.035

Physical Activity Level	3.01	1.413	.002	.017	-1.298	.035
Health Status	2.99	1.418	.006	.017	-1.310	.035
Diet	3.01	1.408	-.011	.017	-1.286	.035
Exercise Level	3.01	1.411	.005	.017	-1.294	.035
Smoking Habit	2.50	1.120	.001	.017	-1.365	.035
Alcohol Consumption	2.50	1.121	-.005	.017	-1.366	.035
Stress Level	2.99	1.411	.011	.017	-1.296	.035
Socioeconomic Status	2.49	1.118	.004	.017	-1.359	.035
Living Conditions	2.99	1.412	.004	.017	-1.297	.035
Sanitation	2.99	1.405	.012	.017	-1.286	.035
Environmental Hazard Exposure	2.97	1.413	.032	.017	-1.296	.035
Access to Healthcare	2.99	1.412	.014	.017	-1.299	.035
Quality of Care	3.01	1.411	-.004	.017	-1.295	.035
Cost of Care	2.99	1.420	-.004	.017	-1.310	.035
Health Policies Awareness	2.49	1.120	.009	.017	-1.365	.035
Valid N (listwise)						

Table 4.21 shows variable statistics. Each category's mean score is close to 3, with tiny differences depending on the variable. "Age" has a mean of 3.82, indicating a significantly older population, whereas "Gender" and "Health Policies Awareness" both have 2.49, indicating a near-equal gender and awareness distribution. Moderate data variability is indicated by standard deviations of 1.115 to 1.712. Skewness values like -.175 for "Age" and -.272 for "Annual Income," indicate slight negative skewness for most variables, implying the data leans somewhat lower. The Kurtosis values, largely negative (e.g., -1.270 for "Age" and -1.422 for "Annual Income"), imply flatter distributions than normal distributions. The data shows a very even distribution across categories, with some small underperformance in gender distribution and socioeconomic class.

4.5 Nonparametric Correlations

Table 4.22: Correlations

			Healthcare Access	Pollution Level	Nutrition Access	Physical Activity Level	Health Status	Smoking Habit	Alcohol Consumption	Socioeconomic Status	Environmental Hazard Exposure	Access to Healthcare	Health Policies Awareness
Spearman's rho	Healthcare Access	Correlation Coefficient	1	0.005	0.001	0.005	0.005	-0.001	0.006	-0.001	-0.002	0.008	-0.01
		Sig. (2-tailed)	.	0.5	0.861	0.452	0.482	0.899	0.425	0.868	0.734	0.236	0.179
	Pollution Level	Correlation Coefficient	0.005	1	0.001	0.002	0.009	0.004	-0.004	-0.011	0.007	-0.001	-0.006
		Sig. (2-tailed)	0.5	.	0.896	0.745	0.215	0.542	0.598	0.105	0.342	0.92	0.376
	Nutrition Access	Correlation Coefficient	0.001	0.001	1	.019**	-0.001	0.009	-0.01	-0.008	0.004	-0.004	0.003
		Sig. (2-tailed)	0.861	0.896	.	0.007	0.889	0.224	0.165	0.278	0.574	0.58	0.63
	Physical Activity Level	Correlation Coefficient	0.005	0.002	.019**	1	0.007	-0.001	0.002	0.001	-0.006	-0.013	.021**
		Sig. (2-tailed)	0.452	0.745	0.007	.	0.315	0.838	0.798	0.835	0.407	0.061	0.003
	Health Status	Correlation Coefficient	0.005	0.009	-0.001	0.007	1	-0.01	-0.001	-0.003	-.014*	-0.004	-0.006
		Sig. (2-tailed)	0.482	0.215	0.889	0.315	.	0.178	0.853	0.645	0.043	0.54	0.389
	Smoking Habit	Correlation Coefficient	-0.001	0.004	0.009	-0.001	-0.01	1	.014*	-0.001	0.011	0.001	0.001
		Sig. (2-tailed)	0.899	0.542	0.224	0.838	0.178	.	0.046	0.93	0.108	0.886	0.839

	Alcohol Consumption	Correlation Coefficient	0.006	-0.004	-0.01	0.002	-0.001	.014*	1	0.007	-0.002	-0.003	0.008
		Sig. (2-tailed)	0.425	0.598	0.165	0.798	0.853	0.046	.	0.302	0.82	0.655	0.25
	Socioeconomic Status	Correlation Coefficient	-0.001	-0.011	-0.008	0.001	-0.003	-0.001	0.007	1	-0.003	-0.003	-0.006
		Sig. (2-tailed)	0.868	0.105	0.278	0.835	0.645	0.93	0.302	.	0.672	0.629	0.384
	Environmental Hazard Exposure	Correlation Coefficient	-0.002	0.007	0.004	-0.006	-.014*	0.011	-0.002	-0.003	1	0.004	-.015*
		Sig. (2-tailed)	0.734	0.342	0.574	0.407	0.043	0.108	0.82	0.672	.	0.609	0.029
	Access to Health care	Correlation Coefficient	0.008	-0.001	-0.004	-0.013	-0.004	0.001	-0.003	-0.003	0.004	1	-0.007
		Sig. (2-tailed)	0.236	0.92	0.58	0.061	0.54	0.886	0.655	0.629	0.609	.	0.357
	Health Policies Awareness	Correlation Coefficient	-0.01	-0.006	0.003	.021**	-0.006	0.001	0.008	-0.006	-.015*	-0.007	1
		Sig. (2-tailed)	0.179	0.376	0.63	0.003	0.389	0.839	0.25	0.384	0.029	0.357	.

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

In table 4.22, most correlations between variables are weak or negligible. Notably, "Physical Activity Level" is positively correlated with "Nutrition Access" ($r = 0.019$, $p = 0.007$) and "Health Policies Awareness" ($r = 0.021$, $p = 0.003$), suggesting that more active people may have better nutrition access and health policy awareness. "Health Status" has a significant negative connection with "Environmental Hazard Exposure" ($r = -0.014$, $p =$

0.043), suggesting that increased exposure may worsen health. "Smoking Habit" has a slight positive link with "Alcohol Consumption" ($r = 0.014$, $p = 0.046$), suggesting smokers may drink more. "Socioeconomic Status" usually has no significant associations with "Healthcare Access," "Pollution Level," or "Nutrition Access". Similarly, "Healthcare Access" and "Pollution Level" ($r = 0.005$, $p = 0.500$) and "Healthcare Access" and "Socioeconomic Status" ($r = -0.001$, $p = 0.868$) exhibit no statistically significant connections in this dataset.

4.6 Experimental Result of Decision Tree Classifier

An integral aspect of developing a model is putting it into action. This aids in determining which model best represents the data and how well that model will perform going forward. Here, the following outcome is presented graphically. Visualisation of data with a classification system using graphs. This section provides the decision tree classifier implementation outcome with a performance matrix. The following table 4.23 shows the Proposed decision tree classifier training and testing performance in terms of performance measures.

Table 4.23: Decision Tree Classifier performance on training and testing dataset

Performance matrix	Decision Tree Classifier	
	Training	Testing
Accuracy	64.83	56.32
Precision	64.87	56.32
Recall	64.83	56.32
F1-score	64.81	56.30
Specificity	67.08	58.35
Sensitivity	62.60	54.22

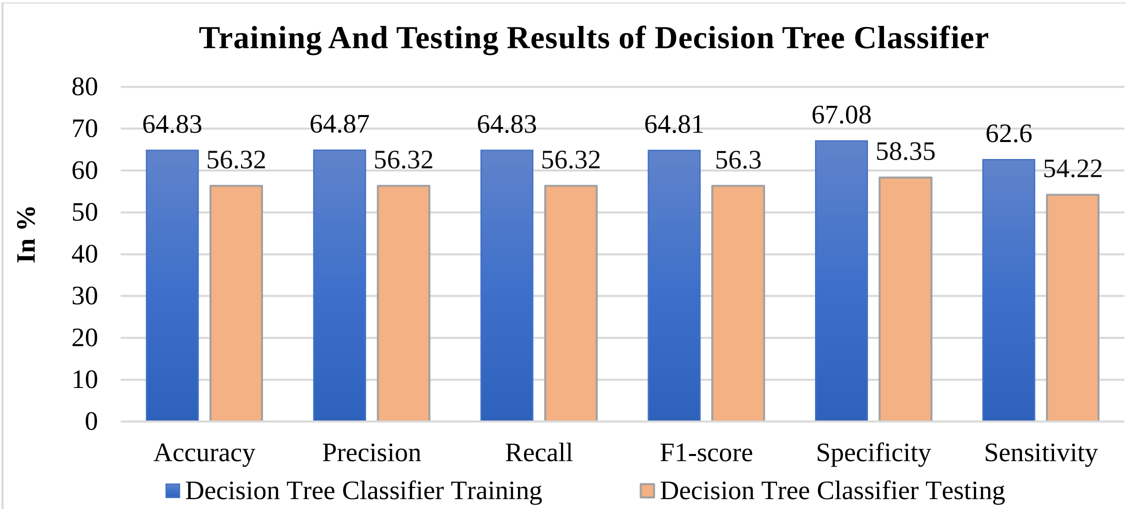

Figure 4.32: Performance of DT classifier on USA_data_SDOH

The above table 4.23 and figure 4.32 show the Performance of DT classifier on USA_data_SDOH. On training dataset, classifier achieved an accuracy of 64.83%, precision of 64.87%, recall of 64.83%, F1-score of 64.81%, specificity of 67.08%, and sensitivity of 62.60%. For the testing dataset, performance declined slightly, with accuracy, precision, and recall of 56.32%, F1-score of 56.30%, specificity of 58.35%, and sensitivity of 54.22%. This variation between training and testing performance indicates some generalization challenges, suggesting that while the model captures patterns well on the training data, it may require further tuning or regularization to improve performance on unseen data.

```
Classification Report:
              precision    recall  f1-score   support

           0       0.64      0.67      0.66      9491
           1       0.66      0.63      0.64      9563

    accuracy                           0.65     19054
   macro avg       0.65      0.65      0.65     19054

weighted avg       0.65      0.65      0.65     19054
```

Figure 4.33: Training Classification report for decision tree

The training Classification report for decision tree is shown in Figure 4.33. A general accuracy of 0.65 for the DT model means that it gets 65% of the training set samples right. With a precision of 0.64, recall of 0.67, and an F1-score of 0.66 for Class 0, the model achieves balanced performance, successfully predicting Class 0 samples 64% of the time and correctly identifying 67% of real Class 0 cases. Class 1 shows a quite low recall but respectable precision with a precision of 0.66, recall of 0.63, and an F1-score of 0.64. With similar class support, the accuracy, recall, and F1-score macro and weighted averages are 0.65, indicating equal performance across classes.

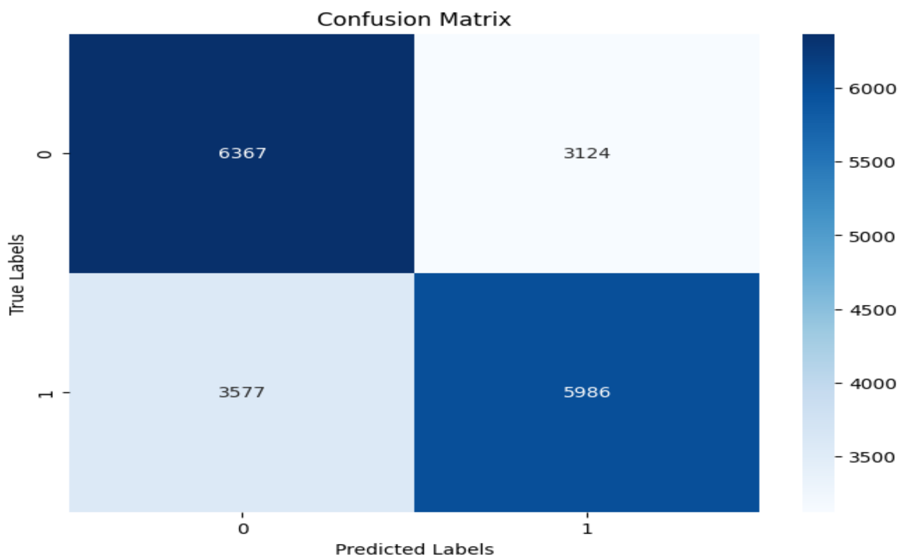

Figure 4.34: Training Confusion matrix for decision tree

Figure 4.34 confusion matrix compares the model's predicted classifications to the actual labels, giving a clear picture of the model's performance. Here, 5,986 cases were appropriately classified as class 1 (TP) and 6,367 as class 0 (TN) by the model. But it also incorrectly labelled 3,577 instances as class 1 (FN) and 3,124 as class 0 (FP) when in fact they were class 0. These values highlight the areas where the model made errors, with FP and FN indicating the instances where the model's predictions diverged from the true classifications.

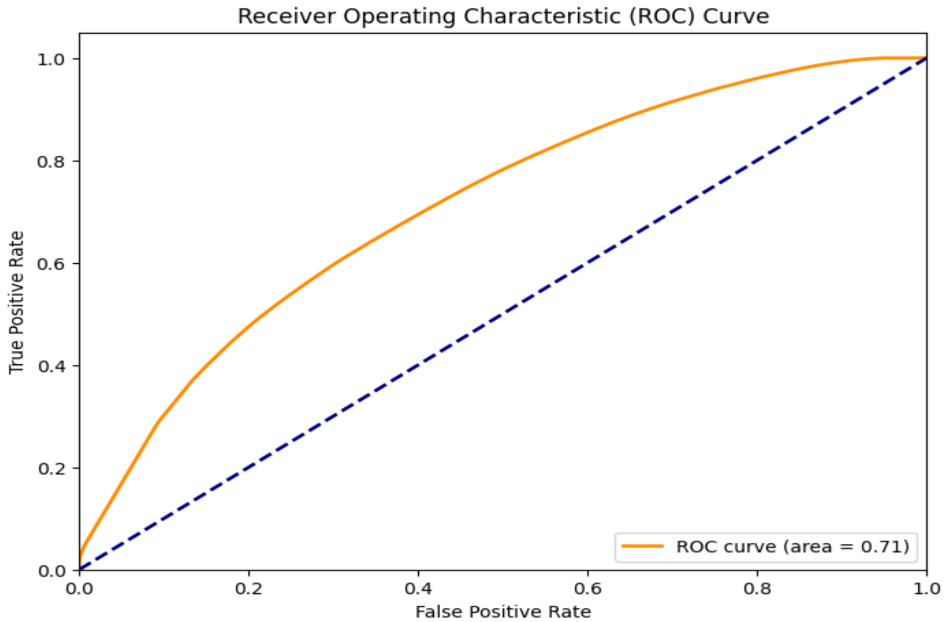

Figure 4.35: Training ROC curve for decision tree

Figure 4.35 displays the decision tree's training ROC curve. As the area under the curve (AUC) for the model is 0.71, it appears to perform adequately, as it outperforms a random classifier (AUC = 0.5). While the model correctly identifies positive cases (high TPR) and minimizes false positives (low FPR), there is still room for improvement in its classification ability. The AUC value indicates a trade-off between TP and FP rates, and further tuning could enhance performance.

Classification Report:

	precision	recall	f1-score	support
0	0.57	0.58	0.58	2418
1	0.56	0.54	0.55	2346
accuracy			0.56	4764
macro avg	0.56	0.56	0.56	4764
weighted avg	0.56	0.56	0.56	4764

Figure 4.36: Testing Classification report for decision tree

Classification reports for decision tree classifiers are shown in Figure 4.36 above. The model successfully predicted 56% of samples, with an accuracy of 0.56, as shown in the image. A precision of 0.57, recall of 0.54, and F1-score of 0.58 are achieved for class 0, whereas a precision of 0.56, recall of 0.54, and F1-score of 0.55 are achieved for class 1. With a weighted average of 0.56 for precision, recall, and F1-score, we can see that performance across classes is balanced yet moderate, with room for improvement.

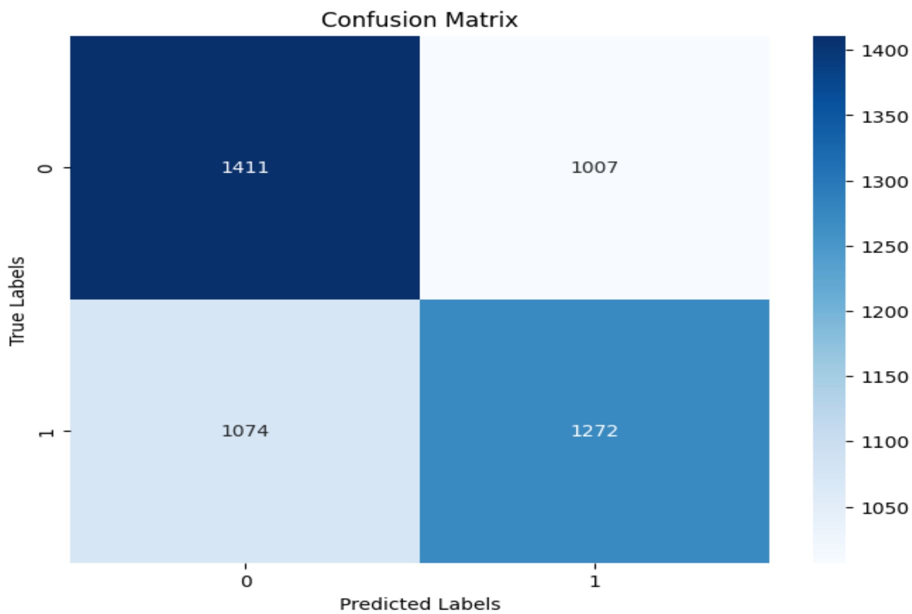

Figure 4.37: Testing Confusion matrix for decision tree

Figure 4.37 shows the DT model's performance in the confusion matrix, which shows that 1411 cases were correctly classified as class 0 (TP), while 1007 instances were wrongly projected as class 1 when they were properly classed as class 0 (FP). Alternatively, 1272 occurrences were spot-on classed as class 1 (TN), while 1074 were incorrectly labelled as class 0 (FN). Further evaluation utilising performance metrics including F1-score, recall, and precision reveals the model's accuracy in categorising both classes and places where misclassifications occurred.

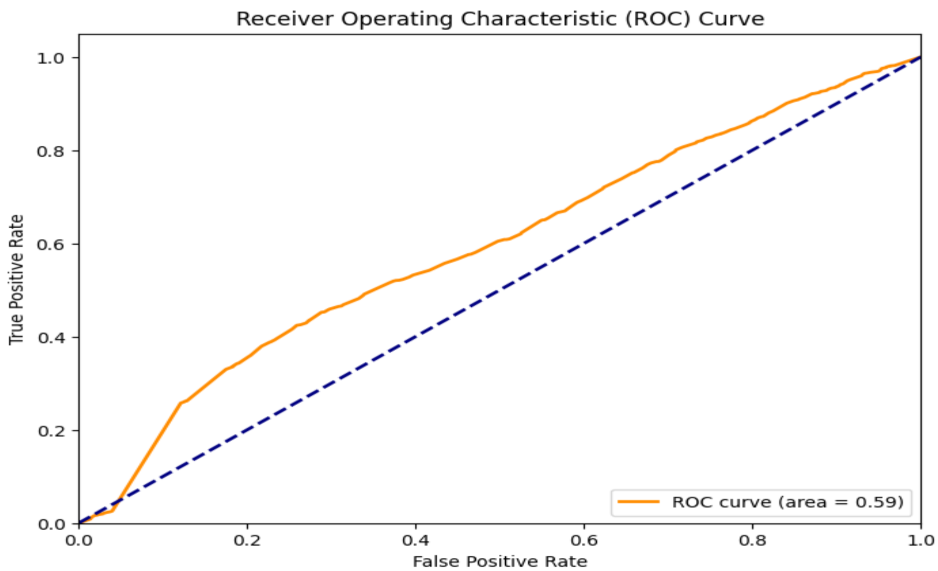

Figure 4.38: Testing ROC curve for decision tree

The ROC curve assesses a classifier's performance by plotting true positive rate (sensitivity) versus the false positive rate shown in Figure 4.38. An AUC of 0.59 here shows moderate model performance, with a steeper curve shape indicating stronger class separation. The diagonal represents random guessing; a curve below it would suggest performance worse than random.

4.7 Experimental Result of XGBoost Classifier

This section provides the XGBoost model outcome with performance measurement. The proposed are implemented on the USA_data_SDOH dataset. The following table 4.24 shows the XGBoost classifier results on the training and testing dataset.

Table 4.24: XGBoost Classifier performance on training and testing dataset

Performance matrix	**XGBoost Classifier**	
	Training	**Testing**
Accuracy	99.64	62.28
Precision	99.64	62.57
Recall	99.64	62.28

F1-score	99.64	61.94
Specificity	99.28	71.46
Sensitivity	99.28	52.81

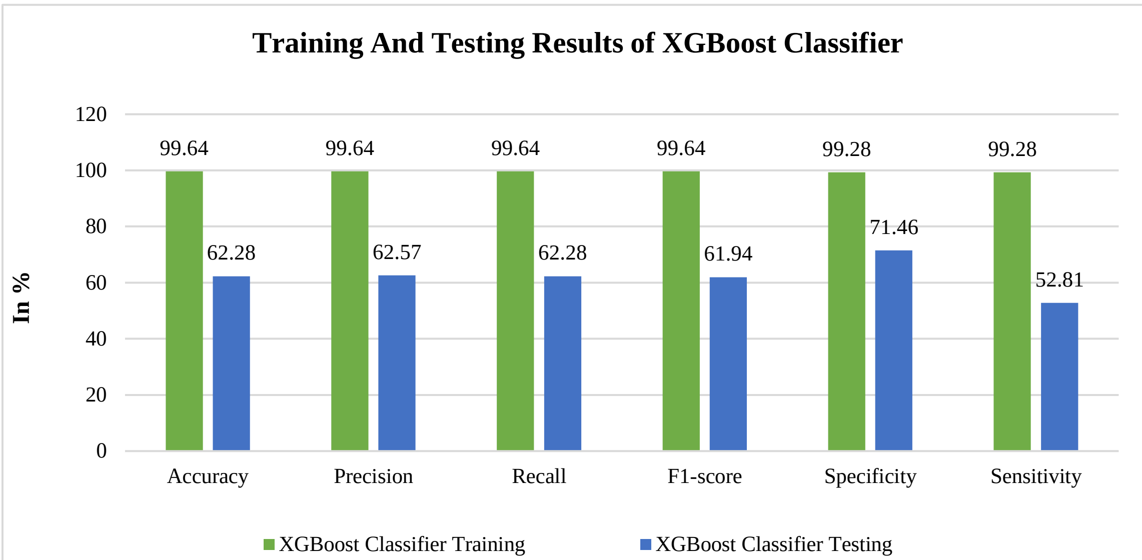

Figure 4.39: Performance of XGBoost classifier on USA_data_SDOH

The performance of the XGBoost classifier shown in Table 4.24 and Figure 4.39 reveals a significant drop in performance from training to testing data. While the model achieves high accuracy, recall, precision, and F1-score on the training set (around 99.64%), these metrics decrease substantially on the testing set, with accuracy, recall, precision, and F1-score dropping to 62.28%, 62.57%, 62.28%, and 61.94%, respectively. The specificity increases on the testing set to 71.46%, indicating a better ability to correctly identify negatives, but sensitivity decreases to 52.81%, suggesting the model struggles to identify positives effectively in the test data. The model's strong performance on the training data belies its inability to generalise to new data, a phenomenon known as overfitting.

Classification Report:

	precision	recall	f1-score	support
0	1.00	0.99	1.00	9491
1	0.99	1.00	1.00	9563
accuracy			1.00	19054
macro avg	1.00	1.00	1.00	19054
weighted avg	1.00	1.00	1.00	19054

Figure 4.40: Training Classification report for XGBoost

The classification report in figure 4.40 reveals exceptional performance of the XGBoost, achieving near-perfect results across all metrics. Class 0 has an F1-score of 1.00, indicating that the model achieves a balance between recall and precision, with 99% recall and 100% precision. A flawless F1 score of 1.00, 100% recall, and 99% precision are also achieved by the model for Class 1. In total, the model says it can predict every single one of them with a perfect score of 100%. A perfect score of 1.00 on the F1-score, recall, and precision/weighted averages shows that the performance is consistent and exceptional in both classes, even with potential class imbalances.

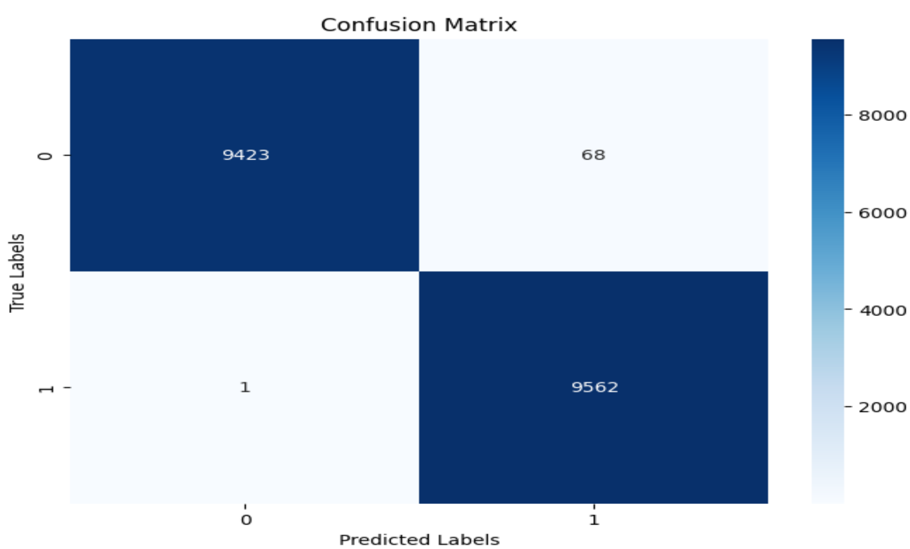

Figure 4.41: Training Confusion matrix for XGBoost

Figure 4.41 shows the Training Confusion matrix for XGBoost. It shows that the model correctly predicted 9423 instances as class 0 (True Positives, TP) and 9562 instances as class 1 (TN). There were 68 FP, where class 0 instances were misclassified as class 1, and only 1 FN, where a class 1 instance was incorrectly predicted as class 0. This indicates that the model performs exceptionally well with minimal misclassifications, particularly in identifying class 1 instances.

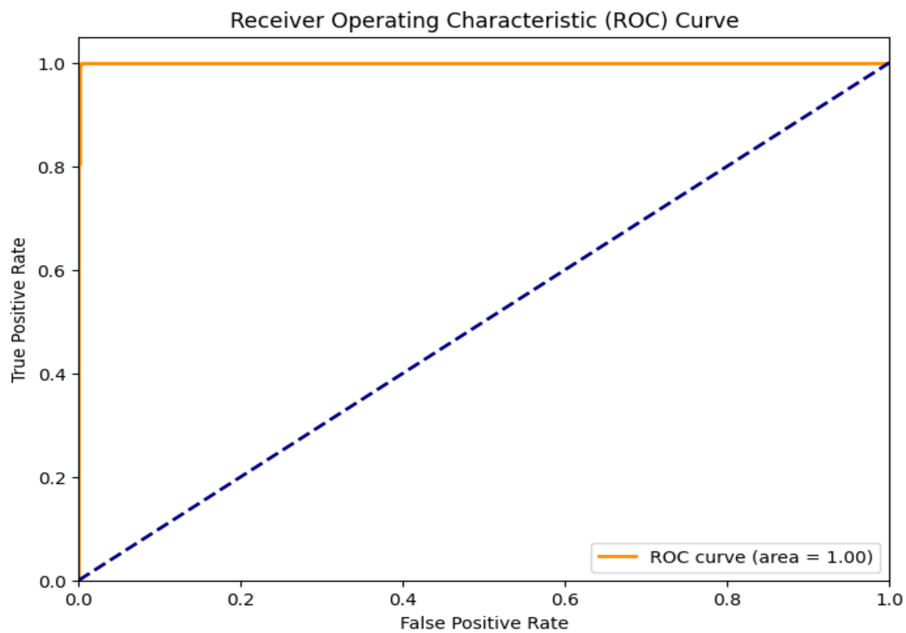

Figure 4.42: Training ROC curve for XGBoost

Figure 4.42 displays the XGBoost training ROC curve. An AUC of 1.00 on the provided ROC curve indicates that the model is performing as expected. This proves that the model achieves its optimal performance by correctly classifying all occurrences without any false positives or negatives. A higher AUC value generally reflects better model performance, and an AUC of 1.00 represents flawless classification.

Classification Report:

	precision	recall	f1-score	support
0	0.61	0.71	0.66	2418
1	0.64	0.53	0.58	2346
accuracy			0.62	4764
macro avg	0.63	0.62	0.62	4764
weighted avg	0.63	0.62	0.62	4764

Figure 4.43: Testing Classification report for XGBoost

Figure 4.43 shows the XGBoost classification report, which includes an F1-score, recall, and accuracy breakdown for each class. The model's accuracy for class 0 is 0.61, recall is 0.71, and F1-score is 0.66. The F1-score is 0.58, recall is 0.53, and precision is 0.64 for class 1. With a weighted average of 0.62 for precision, recall, and F1-score and a macro average of 0.62 for overall accuracy, the model shows moderate performance and might be improved when it comes to detecting class 1 occurrences.

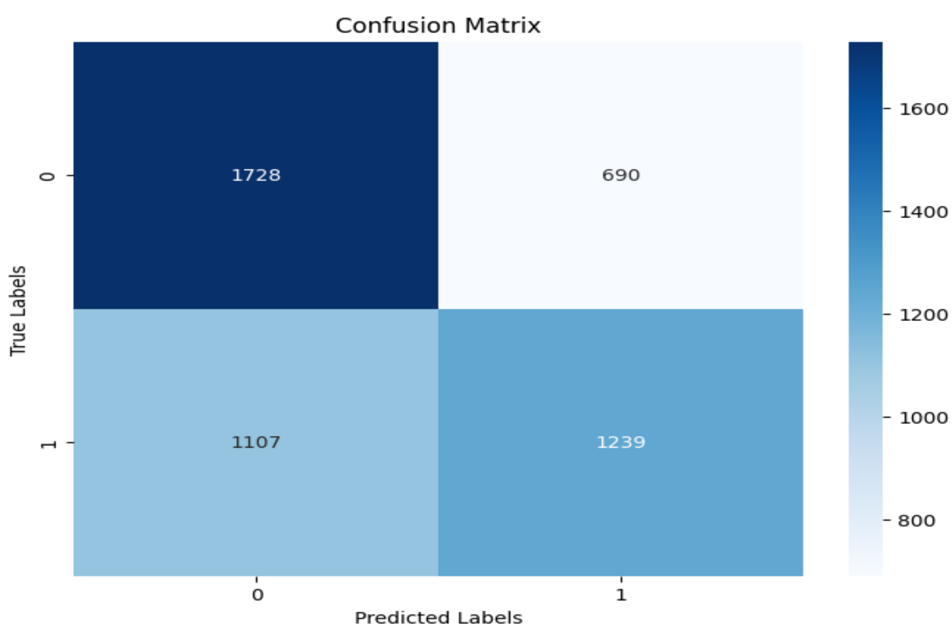

Figure 4.44: Testing Confusion matrix for XGBoost

Figure 4.44 confusion matrix shows how well a classification model with two classes (0 and 1) performs. There are 1728 TP, where instances of class 0 were correctly predicted as class 0, and 690 FP, where class 0 was incorrectly predicted as class 1. Similarly, 1239 TN correctly predicted instances of class 1 as class 1, while 1107 FN incorrectly predicted class 1 instances as class 0. This matrix provides a detailed overview of the model's prediction accuracy, highlighting areas of correct and incorrect classifications for both classes.

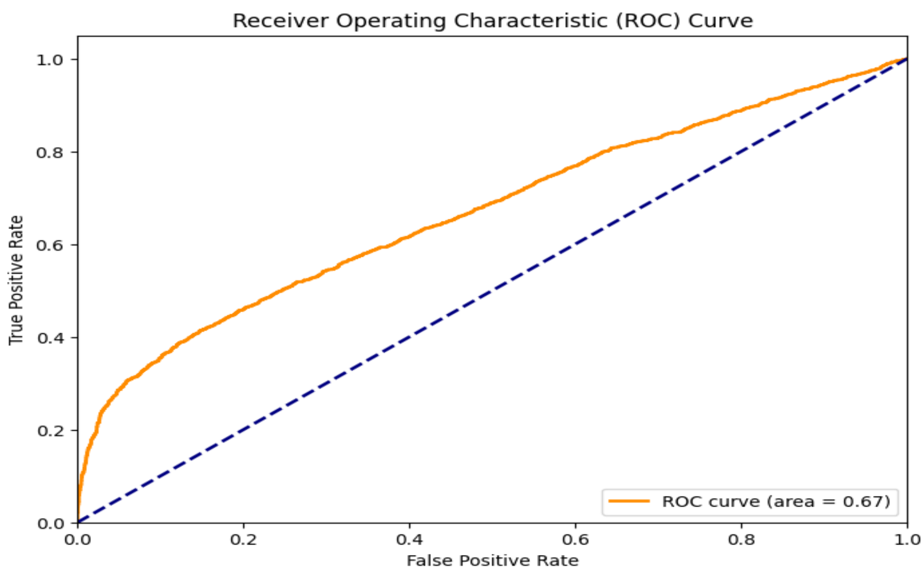

Figure 4.45: Testing ROC curve for XGBoost

The ROC curve, shown in figure 4.45, illustrates the efficiency of a binary classification model by comparing the true positive rate (TPR) with the false positive rate (FPR) at various thresholds. In the provided ROC curve, the Area Under the Curve (AUC) is 0.67, suggesting that the model performs moderately well, being better than a random classifier (AUC = 0.5), but still has room for improvement. A higher AUC would indicate better overall model performance, with a good classifier aiming for a high TPR and low FPR.

4.8 Comparison of Training and Testing of Proposed Model Performance

The comparison of the training and testing performance of proposed model highlights the effectiveness of the model across training and testing datasets and evaluation metrics such as precision, accuracy, sensitivity, recall, specificity, and f1-score. The following table 4.25 shows the proposed model comparison between training and testing datasets.

Table 4.25: Comparison between Training and Testing proposed models on the dataset

Performance matrix	**Decision Tree Classifier**		**XGBoost Classifier**	
	Training	**Testing**	**Training**	**Testing**
Accuracy	64.83	56.32	99.64	62.28
Precision	64.87	56.32	99.64	62.57
Recall	64.83	56.32	99.64	62.28
F1-score	64.81	56.30	99.64	61.94
Specificity	67.08	58.35	99.28	71.46
Sensitivity	62.60	54.22	99.28	52.81

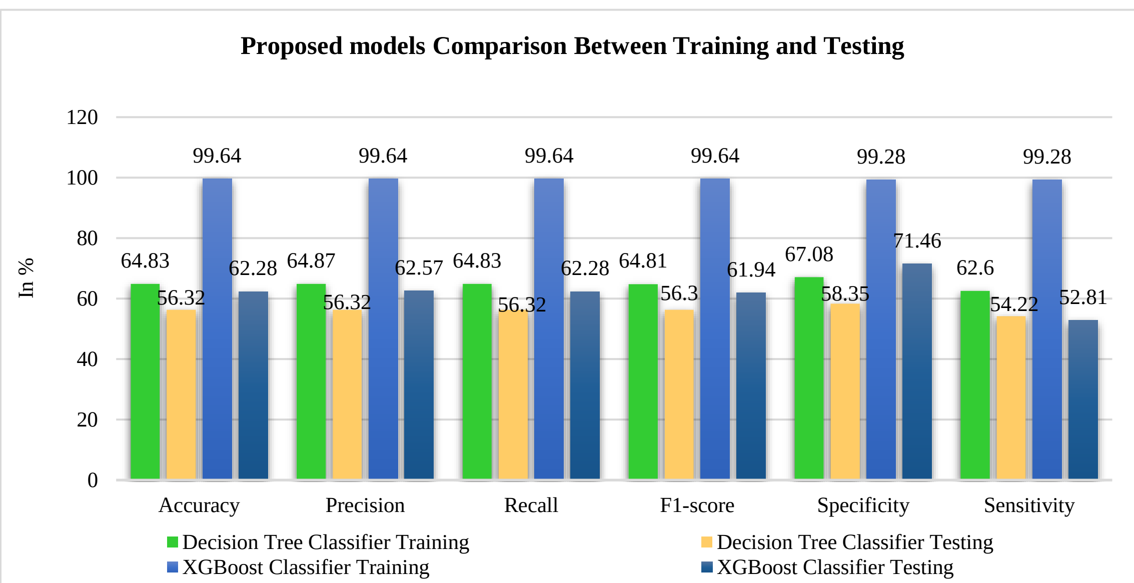

Figure 4.46: Comparison Between Training and Testing of proposed models

The comparison between the DT Classifier and XGBoost Classifier reveals significant performance differences shown in figure 4.46. DT shows moderate performance with accuracy, recall, precision, and F1-score around 64% on the training set, dropping to approximately 56% on the testing set. In contrast, XGBoost exhibits exceptional training performance with accuracy, recall, precision, and F1-score near 100%, but it experiences a noticeable drop on accuracy, testing set, precision, recall, and F1-score around 62%. Specificity is higher for XGBoost, especially on the test set (71.46%), compared to the Decision Tree (58.35%), while sensitivity is lower for XGBoost (52.81%) compared to the DT 54.22%, indicating that XGBoost struggles more with identifying certain positive instances. Overall, XGBoost performs better in training but faces a drop in generalizability, while DT shows more consistent but lower performance across both training and testing.

RQ1: What gaps exist in current data collection efforts related to Social Determinants of Health (SDOH), and is there a need for additional data capture?

The available database needs more data. The available dataset helps us understand basic SDOH aspects but fails to include detailed information that would strengthen our understanding of these social determinants. The dataset lacks complete information about neighbourhood conditions and social support networks plus specific details of financial and housing stability factors. The measured correlation showed significant results but weak values suggesting that the available data cannot explain all the health determinants and health outcomes relationships. Adding personal health assessments along with nutrition standards and safety perceptions makes health data better for aiding healthcare choices.

RQ2: In what ways can existing SDOH data be optimally leveraged to enhance value-based care programs?

Health providers use existing data to help patients most at risk by developing specific treatment plans. SDOH data helps determine patients at increased medical risk because of their weak financial situation and limited healthcare options through risk adjustment tools. This SDOH information directs healthcare resources toward providing community health worker and telehealth services. Our community's nutrition needs drive us to establish health programs that bridge food access problems through nutrition support. SDOH information becomes more useful for individual care when combined with medical records which helps both patient health outcomes and better use of resources. Health providers can update their services right away based on data information to generate better healthcare results.

RQ3: Which open-source tools and datasets are available to support the generation of actionable insights on SDOH?

Healthcare providers can use current patient information to make better health outcomes by planning services for specific patient groups at risk. Healthcare providers use risk adjustments of SDOH data to find patients who face health risks because of their social conditions and help these patients through community health worker visits or telehealth services. SDOH statistics inform preventive health service creation to help people who lack proper nourishment through nutrition plans. SDOH insights combined with medical information helps healthcare teams design complete personalized care plans that help their patients receive better care without wasting resources. Regular updates of SDOH data help healthcare staff and health plans to quickly fine-tune their programs for better outcome results.

RQ4: Beyond SDOH-specific factors, what additional insights can be derived from integrated clinical, claims, and social data sources?

The research provides several main areas that enable better understanding beyond Social Determinants of Health (SDOH). Research on healthcare utilization patterns comprising data of emergency department visits and hospital admissions along with outpatient services helps determine improved healthcare service distribution. Effective quality improvement strategies emerge from patient feedback mechanisms which allow the identification of areas that require improvement in order to provide better patient-centered care. Economic assessment, which includes cost-effectiveness analysis together with financial barrier evaluation, leads to healthcare delivery systems being more equitable. Modern healthcare receives valuable prospects for improved patient care through integrating digital health together with telehealth solutions into the existing healthcare system.

RQ5: What are the most effective strategies for communicating SDOH-related insights into health plans and providers, and would a separate, dedicated reporting mechanism improve decision-making?

The report shows insights work best when organizations combine different ways of sharing data to satisfy all their audience groups. Different groups need special content and visual displays like charts or graphics to learn better from information directed towards each of their needs. Regular online sessions help people understand data more effectively. To replace static reports with dynamic dashboards, help users receive updated data and practical insights in real time. The need for extra reports depends on how difficult the findings are to understand and match reporting needs from stakeholders; however, putting analysis directly into communications saves time.

4.9 Summary

The SDOH dataset includes 20,000 rows which collect diverse data points from demographic statistics to socioeconomic elements and environmental variables as well as behavioural measurements. The collected data contains variables such as housing quality together with pollution levels, nutrition access, and health policy awareness that demonstrate uniform distribution throughout the complete sample. Most of the statistical variables focus on average values while showing average dispersion and exhibiting a negative tendency. The key factors in this study display minimal connection patterns except for the correlations observed between physical activity and nutrition access and the negative effects between environmental hazard exposure and health status. Two classifier models were used for this analysis: Decision Tree and XGBoost. During training, the Decision Tree generated moderate accuracy at 64–65%, yet its test performance dropped to about 56%. The XGBoost model achieved almost perfect results on training data (approximately 99.6% across all key metrics) before experiencing a significant drop, indicating overfitting during testing, which resulted in 62% performance metrics.

4.10 Conclusion

The balanced and inclusive design approach in the SDOH dataset demonstrates its substantial value in public health studies about social determinants and health results. The evaluative results demonstrate that Decision Tree maintains consistent though average results between its training and testing performance, yet XGBoost achieves outstanding training accuracy but exhibits underperformance when assessing unseen data due to overfitting. Additional model optimization techniques, such as cross-validation and regularization, together with feature selection methods, need implementation to enhance the generalizability of complex models. The SDOH dataset proves valuable for directing public health decisions through its research but requires precise model optimization for strong predictive applications. The next chapter discusses the results of the study.

CHAPTER V:

DISCUSSION

5.1 Discussion of Results

Consequently, the results of this study enlighten the understanding of the associations among multiple SDOH, healthcare utility, and discrete individual behaviors, especially in an ethnically diverse population. From these analyses, evidence provided by low but significant coefficients means that though such factors have correlations, the correlation coefficients are, nevertheless, small when read in isolation. This goes a long way in supporting the contention that the SDOH has a chain-like influence on health than the individual determinants.

This work also offered a qualitative analysis of Decision Tree Classifier and XGBoost Classifier based on the USA SDOH dataset. The performance measures adopted while training and testing the models were accuracy, precision, recall, F1-score, specificity, and sensitivity. The applied models showed poor performance while tested on unseen data, where Decision Tree Classifier brought moderate accuracy, positive as well as negative predictive values and recall on both the training as well as testing sets, where the model seems to have a generalization problem.

By the same token, the XGBoost model sacrifices high accuracy on the testing data for that of high precision reflected in high variances where the model performs a spectacular performance on the training dataset but a very poor performance on the testing set due to overfitting. The comparison of these two models showed that although XGBoost achieved very high accuracy on the training dataset, it was not as accurate in detecting the testing dataset, except for improved specificity, it had lower sensitivity. These findings suggest that higher model tuning, adding algorithms like regularization, and maybe feature engineering are required to improve the model's generalization and its performance on validation datasets.

5.2 Discussion of Research Question One

As it is revealed, a primary source of information on various SDOH aspects is existing data, which has its strengths and weaknesses, including the lack of detailed and extensive information. The current data include such categories as healthcare, the level of income, the presence of pollution, and a sedentary lifestyle. It does not include finer distinctions such as neighbourhood environmental characteristics, targeted social support system, and precise financial/housing stability parameters. These elements are essential in order to refer to the global perspective on the given people's health.

Although the correlation coefficients that were computed for this study analysis were statistically significant, they are rather low, implying that the current data may not be sufficient in capturing the interconnection between health determinants and health outcomes. Furthermore, finer grained data might help the healthcare providers and health plans to realize that the factors constituting the SDOH are significantly intertwined with each other. For instance, incorporating self-assessed mental health care, indexes of specific nutritional quality, and perceived safety might increase the effectiveness of information on SDOH.

As such, there is an urgent need for more extensive data collection regarding SDOH that comes geographically comprehensive and quantitatively illustrative and can add depth to value-based care practices by offering a more accurate picture of how these factors affect health.

5.3 Discussion of Research Question Two

Existing data, though often broad, provides valuable insights into key areas such as healthcare access, socioeconomic status, and lifestyle behaviours. By strategically applying this data, health plans and provider organizations can tailor interventions to target specific at-risk groups, thus improving clinical outcomes and enhancing the effectiveness of value-based care initiatives.

A viable strategy involves using SDOH information to risk-adjust the patients to those who are more susceptible to experiencing health issues from social or environmental determinants. Still, for example, for people with lower income and access to healthcare, school Social Behavioral Health Analysts should provide additional services – community health worker visits and telehealth services. Moreover, SDOH data can be helpful for prophylactic efforts. For instance, information regarding lack of access to adequate nutrition can inform a feeding scheme/project or nutritionist presently/ in future possibly lowering incidences of diseases caused by poor nutrition.

It is also essential to involve SDOH data in care management processes as well due to the reasons that will be explained below. Current SDOH knowledge in combination with clinical data allows the development of individual comprehensive care plans that will improve the quality of treatment. Additionally, using existing data to continuously track program performance enables HLPs and health plans to adapt in real-time to the needs of the population while enhancing performing and effectiveness of envisioned value-based programs.

5.4 Discussion of Research Question Three

The study acknowledges that other freely accessible resources exist that health plans and provider organizations can use to better navigate and utilize SDOH data. Secondary data sources, which include repositories and tools from the Government, academic institutions and public health organizations, provide useful information that can be collected in addition to internal data to enrich the insight-generation process.

For example, the U.S. Census Bureau and the American Community Survey (ACS) provide demographic, socioeconomic, and housing data at both national and local levels, which are useful for identifying population-level trends related to SDOH. Furthermore, all the counties have organizations and agencies such as the Centers for Disease Control and Prevention (CDC) offer resources like the Social Vulnerability Index (SVI) that identify populations that may experience additional challenges during public health emergencies depending on the aspects like poverty, housing, language among others.

In addition, through this project, health organizations can use online data visualization tools, including Google Data Studio and Tableau Public, to facilitate data analysis and presentation of SDOH data to enhance understanding of the data gathered. These open source resources can be deliberately incorporated into existing data to enhance the understanding of need within specific populations, monitor SDOH effects on health status, and design interventions within the value-based care models.

5.5 Discussion of Research Question Four

While SDOH are vital for understanding health outcomes, additional dimensions of data can further enhance healthcare quality and effectiveness. Key areas for exploration include:

- **Utilization Patterns:** Access and utilization information highlighted includes emergency visits, hospitalizations and outpatient services where understanding trends and hurdles essential in decision-making and coordination of care can be achieved. Urgent care facilities do well in controlling non-emergent conditions, which, in return, leads to lower co-payments and shorter waiting times as compared to emergency departments.

- **Patient Experience and Satisfaction:** Their views are crucial in establishing areas of inefficiency in the treatment of patients hence the importance of the feedback. Such findings, hence, provide the basis for understanding the patients' preferences when determining the rights of patient-centred care programs to implement to increase general satisfaction.
- **Economic Factors:** The assessment of the economic effectiveness of healthcare interventions is useful in making determinations of cost-effectiveness that give information on financial concerns of health inequality. Insurance trends are equity indicators, which can help us track how payment approaches distort patient access.
- **Technological Advancements:** Knowledge of health information technology is beneficial in strengthening the flow and openness of information between caregivers. Also, the examination of the pattern makes it possible to suggest specific approaches to the enhancement of telehealth services.

In conclusion, the analysis of clinical data, patient experience, economic aspects, and the use of technologies, as well as broadened focus beyond SDOH, offers the opportunity for further comprehensiveness of the health care needs and results. It goes without saying that this concept is quite helpful for the creation of effective as well as fair healthcare policies.

5.6 Discussion of Research Question Five

Creating accurate messages that help others better understand findings from SDOH and other data is critical to support their adoption and inform their practice in decision-making and value-based care settings. The absence or inadequate communication of those insights is a major factor that prevents health plans and providers from making better decisions on SDOH and the overall effects on health outcomes. Here are several strategies for effectively communicating these insights, along with a discussion on the necessity of a separate report:

1. **Tailored Communication Strategies:**
   - **Audience-Specific Messaging:** The horizon issue depends on the interests and knowledge of the health plans, the providers, and other administrative organizations. May also focus the communication to fit the redundancy and knowledge level of each audience to improve reception. Thus, the idea can entail more, for example clinical case presentations for the care givers, or data analysis and financial aspects for administration teams.
   - **Use of Visual Aids:** Any kind of graphic images such as graphs, charts and information graphics are useful in Order Paper since they expound the data in an easily understandable manner. Charts, graphs, tables and other forms of graphic displays can help underscore and queue important observations, trends and implications in relation to datasets on SDOH.
2. **Regular Briefings and Workshops**
   - **Interactive Sessions:** The stakeholders could be brought together by a constant hosting of briefings, workshops or webinars. They offer a chance to report findings, show implications and to receive feedback. In particular, people use the interactive forms to pose questions and to discuss and amplify the knowledge and its practical implication.
   - **Collaborative Learning:** Interprofessional learning opportunities among health plans and providers enable them to learn from each other and develop new, effective solutions to managing SDOH. This collective approach can result in better spread in terms of insights that can be used in practical contexts.
3. **Integrated Communication Channels**

- **Utilizing Existing Platforms:** If possible, insights can be presented in newsletters, intranet or other specialized websites instead of preparing set apart reports. This must be done in aperiodic updates and summaries to the various stakeholders through the mentioned channels.
- **Digital Dashboards:** The reason is the ability of digital dashboards to offer the actual data on SDOH and make them work with the information as with a problem-solving tool. Dashboards can be more flexible where views can be created and delivered to allow users to view the components appropriate to their responsibilities.

4. **Actionable Recommendations**

- **Focused Insights:** While sharing findings it is crucial to offer practical suggestions based on the findings obtained. To be actionable, health plans and providers are more inclined to utilize insights that present processes for implementation. For example, stating a particular SDOH and recommending ways to work on them increases the chances of adopting the interventions from stakeholders.
- **Highlighting Impact:** Sharing the possibility of demonstrating that improvements in SDOH lead to improvements in health status and decrease in cost may help to encourage health plans and providers to interact with the insights. Results from case studies or pilot programs can be the best form of evidence that can be presented.

5. **Feedback Mechanisms**

- **Gathering Input:** Engaging in feedback mechanisms permits the stakeholders to provide narrations of their feelings as well as experiences with reference to the divulged wisdom. They can be used to refer back to in future communication plans so that relevant data is continuously to be provided.

Based on the results of the study, the message and process for sharing insights about SDOH to HPS and providers should be framed in a method that is engaging, participatory, and using multiple forms of communication. Adding more features like offering serviceable suggestions or including the possibility of how such findings can be valuable will make the engagement even more appealing. Whether a separate report is necessary depends on the complexity of the insights, stakeholder preferences, and available resources. Ultimately, the goal should be to ensure that the communication of insights is clear, actionable, and conducive to improving health outcomes.

5.7 Discussion of Research Question Six

The study underscores the importance of aligning the goals, metrics, and strategies across different value-based care initiatives to enhance the overall effectiveness of care delivery. Achieving synergy involves ensuring that various programs complement each other, share insights, and work collaboratively to improve both patient outcomes and cost-efficiency.

Noteworthy is the fact that there should be cooperation between different value-based programs. Most health plans and provider organizations have more than one value-based program (such as ACO, PCM, bundled payment, etc.). It is not good for these programs to operate in silos for synergy to be realized in the implementation of the programs. Coordination of care is implemented through cross-program teams of program providers, care coordinators, and program administrators to promote information and resource sharing as well as compliance with best practices across all programs. It fosters a well-orchestrated workflow for sharing ideas on patients' management, SDOH information, and treatment plans.

Synergy also depends on data integration which helps in achieving the goal. There is the possibility of sharing data between value-based programs and enhancing a 360-degree perspective of the patient encompassing clinical, behavioral, as well, as SDOH data. The pursuit of effectiveness can be achieved by combining these databases into a single data system that will be used in healthcare facilities to develop a coherent strategy for change implementation where all activities, including individual programs, are aligned towards a specific set of objectives.

Finally, the culture of developing partnerships within healthcare organizations should also be developed. In addition to effective communication, health plans can also use encouragement of cross-relationships and identified goals that support and encourage multi-program efforts among the providers.

Thus, it can be concluded that integration of value-based programs implies both flexible cooperation and integration of data, metrication, and general approach to providing care. Together, they help to guarantee that several programs run not only effectively but also coordinated, which ultimately creates benefits for patient care and helps to decrease the costs of healthcare.

CHAPTER VI:

SUMMARY, IMPLICATIONS, AND RECOMMENDATIONS

6.1 Summary

This study investigated the relationships among social determinants of health (SDOH), healthcare access, lifestyle factors, and health outcomes, and assessed the effectiveness of leveraging SDOH data within value-based care programs. Findings from this study advance the literature on the complexity of SDOH's effects on health outcomes and the possibility of more effective interventions when SDOH data is incorporated with patient clinical data.

The study also addressed potential tasks and limitations of existing data and the need for gathering other SDOH data. Contemporary datasets contain valuable knowledge but do not have fine distinctions in matters such as the communal support system, the neighborhood context, and housing resilience relevant to value-based programs for stratifying risks. Consequently, improved collection of information could allow performing risk evaluations and focus on people's needs more effectively.

With reference to the utilization of existing data, the study further found that SDOH data can be used in care management, risk adjustment, and prevention under value-based programs. Policies, guidelines, and legislative instruments the SDOH data make it possible to develop even better individual preliminary treatment plans because there is real-time tracking of program results as well as leading real-time involving interventions.

The study also found some open-source resources that can potentially complement SDOH findings: federal and state agency data, which are available in public domains, and public health data like the US Census and CDC. Some of the tools that were useful in analyzing SDOH data included Google Data Studio and Tableau Public, which were also mentioned as useful in visualizing SDOH data.

Apart from the SDOH findings, the study identified combining information about healthcare consumption, patient satisfaction, economic indicators, and technological innovations as beneficial to enhance the understanding and enhance healthcare delivery. To support DM and clinical practice change, key communication strategies included: messaging that was adapted to each target group; graphic illustrations aimed at raising awareness about SDOH; brief, interactive sessions to help providers and health plans apply SDOH findings; and digital dashboards for sharing insight.

This research work finalized that; indeed, one can effectively develop options towards realizing the first of the six areas of synergy among the value-based programs and hence improve the health of people by merely implementing coordinated care strategies and integration of the data and metrics in accordance with Universal Health Care. These results can be used to guide decision-making and partnership to reduce health disparities in practical settings by applying a systems-oriented approach among healthcare constituents.

6.2 Implications

Theoretical Implications: This research contributes to the existing literature on SDOH by identifying the interactions between the factors and, therefore, provides a better understanding of how SDOH impacts health. The low but statistically significant coefficients compel the need to evaluate SDOH not as individual factors but as an interrelated complex with impacts on health outcomes. This insight forms the basis of practical theoretical frameworks rather than adding to the theoretical constructs that support a multilevel perspective of health determinants, including evolving facets such as healthcare accessibility, SES (socioeconomic status), and ecological conditions, among others. Such a perspective might lead the future research estimating the advanced multivariate models, such as sophisticated predictive analytics to describe overall interdependence between various SDOH and health outcomes. Furthermore, using the Decision Tree and XGBoost Classifier in machine learning models for the SDOH data analysis, the study reveals a theoretical limitation of generalization in predictive healthcare analytics. The results of the analysis of the model performance imply future research directions for the areas of tuning hyperparameters of the model, controlling for overfitting of the model and improving variable selection. These research contributions advance the field of healthcare informatics and imply that with any practical application of data-driven tools in health, fundamental model performance and generalization capabilities cannot be understated.

Managerial Implications: From a managerial standpoint, the findings underscore the necessity for healthcare organizations, payers, and policymakers to prioritize the integration of SDOH data into value-based care initiatives. Managers should promote improved data capture processes that would allow more detail of SDOH factors, like neighbourhood attributes, financial status, and available resources, to improve the identification of risk for targeting.

The inclusion of other SDOH data enables the development of more comprehensive risk predictive risk profiles which may positively impact on patient outcomes and resource utilization in the high-risk groups. Besides, the study findings about the current use of data provided important prescriptions to health managers who intend to use SDOH in ordering the delivery and coordination of healthcare services. In its effort to make SDOH information accessible to providers and relevant stakeholders, managers might want to look into open-source tools like data visualization platforms like Google Data Studio and Tableau Public. The use of this approach could greatly assist in making sense of extensive SDOH data as well as trigger timely initiatives by the public. When it comes to communicating the SDOH insights to the health givers, the study finds that even though traditional reports remain useful, there is a need for developing relevant reporting formats, conducting engaging workshops and incorporating the dashboards. Managers are encouraged to adopt a multifaceted communication approach, ensuring insights are accessible, actionable, and relevant to diverse audiences within healthcare organizations.

6.3 Recommendations for Future Research

This study's findings suggest several directions for future research to advance the understanding and application of SDOH in healthcare.

- **Enhanced Data Collection on SDOH:** Further research should involve the collection of additional SDOH beyond binaries, with greater temporal samples, including housing and living conditions, living context, and mental wellbeing. With these subtleties, it will be possible to develop more accurate patient pictures and approaches to handling potential SDOH-related interventions, which will in and of itself be informative of the SDOH-health outcomes mediation.
- **Advanced Modeling and Analysis:** In the future studies based on machine learning strategies, deep learning and ensemble learning are suggested to make the models more accurate and implementation oriented. Other practices like cross-validation, feature engineering, feature selection, regularization, use of ensembles and architecture search can avoid overfitting, offering models that can handle new varieties of data properly.
- **Integration of SDOH with Clinical Data:** Integrating SDOH with other clinical information such as the Electronic Health Record or patient history and vital signs will offer a total patient picture. Further research should examine the interplay and impact of these data kinds when it comes to risk assessment in patients with chronic diseases toward enhancement of patient care.

- **Frameworks for Synergy Among Value-Based Programs:** Further research should interrogate for best practice models for integration and concurrency of a range of value-based care programs, SDOH measurement and data sharing and discussion among programs. It could bring the idea of having an alienated approach for creating a unified strategy for offering fuller patient-centred care based on addressing SDOH.
- **Open-Source Tools for SDOH Analysis:** Subsequent qualitative research inquiries should investigate other tools SDOH data extracted from EHRs can be similarly analyzed and visualized using open sources appropriate for the healthcare context. The use of SDOH data in isolation can be problematic, so we need to adopt a set of tools that enable us to compare them and identify the best platforms that can convert this information into useable insights.

In sum, future research should focus on advanced data methods, comprehensive integration with clinical insights, and practical evaluation of SDOH-informed interventions. These efforts will strengthen healthcare strategies and improve patient outcomes by fully incorporating the influence of SDOH factors.

6.4 Conclusion

This study has explored the critical role that SDOH plays in shaping healthcare outcomes and the potential for leveraging these insights to improve care delivery, especially in value-based programs. The study underscores how multifaceted SDOH interactions are because the status factors, healthcare utilization, and behaviours have a summative impact on health. They also assessed machine learning models and found that there are difficulties in conducting cross-study comparisons and noted the importance of developing and selecting appropriate data and features.

Such data is necessary to augment the current limited and unrepresentative data and encourage a more widespread and comprehensive collection of geographically distributed SDOH. With that much richer data, healthcare providers and health plans could tailor more effective and precise interventions and enhance value-based care programs. Additionally, the study emphasizes the importance of adopting SDOH as part of clinical information and the identification of sources of open data that can complement the results of data analysis.

Further, the study outlines major approaches notable in the enhancement of the communication of SDOH insights such as contextualizes communication and technology-enabled communication. Another area that needs attention is the opportunity for integration between value-based programs, whereby approaches such as care coordination and integrated information technologies provide a ground for potentially enhancing patient outcomes while reducing costs significantly.

Consequently, the study offers an extensive review of the issues and prospects of SDOH in healthcare pointing to the areas of significant improvement in the research and practice for effective patient care delivery.

APPENDIX A:

INFORMED CONSENT

IMPACT OF SDOH ELEMENTS IN VALUE BASED CARE MODEL TO DRIVE BETTER CLINICAL OUTCOME FOR US HEALTH PLANS

I, **……………………………………………………………** agree to be interviewed for the research which will be conducted by **…………………………………………………………...**a doctorate student at the Swiss School of Business and Management, Geneva, Switzerland.

I certify that I have been told of the confidentiality of information collected for this research and the anonymity of my participation; that I have been given satisfactory answers to my inquiries concerning research procedures and other matters; and that I have been advised that I am free to withdraw my consent and to discontinue participation in the research or activity at any time without prejudice.

I agree to participate in one or more **electronically recorded** interviews for this research. I understand that such interviews and related materials will be kept completely anonymous, and that the results of this study may be published in any form that may serve its best.

I agree that any information obtained from this research may be used in any way thought best for this study.

……………………………………… ………………

Signature of Interviewee **Date**

APPENDIX B:

DATASET

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Age	Gende	Race	Healthcare Acce	Education Lev	Annual Income (US	Employment Statu	Housing Quali	Pollution Lev	Crime Rat	Nutrition Acces	Physical Activity Lev	Health Statu	Die	Exercise Lev
2	6	2	4	2	5	4	2	5	1	1	3	5	5	3	1
3	5	1	1	1	5	4	2	5	3	3	1	4	5	3	5
4	4	3	1	1	1	1	3	5	4	1	3	3	5	3	5
5	2	2	3	3	4	3	3	4	3	3	4	1	5	1	5
6	3	4	2	3	1	3	4	5	4	3	5	4	2	3	4
7	4	4	1	5	6	5	2	4	3	5	3	1	4	1	2
8	1	1	2	2	3	2	2	3	1	2	1	2	4	3	3
9	6	4	5	4	5	5	2	2	5	2	5	2	4	1	1
10	5	4	3	4	4	1	1	4	2	4	2	4	3	2	3
11	2	4	1	5	3	1	3	4	1	1	5	5	1	1	2
12	4	2	2	4	1	4	1	3	5	5	3	3	4	2	1
13	6	1	5	3	5	1	5	2	2	2	2	5	3	1	1
14	6	1	1	3	6	4	6	2	2	2	2	4	1	1	3
15	6	4	1	5	5	4	2	2	4	3	2	4	4	2	3
16	5	4	5	4	2	2	1	3	5	4	3	5	2	3	1
17	3	1	1	3	6	5	3	1	5	1	3	1	1	4	2
18	5	3	5	3	2	3	4	3	3	5	2	3	5	1	1
19	4	1	2	1	5	4	2	4	2	5	4	3	5	3	4
20	1	1	2	4	2	5	6	2	5	3	5	5	4	4	5
21	2	3	4	4	2	2	6	5	5	2	1	3	5	2	2
22	3	4	5	1	1	1	2	2	4	4	5	2	5	2	2
23	3	2	1	4	2	4	3	5	3	2	5	1	1	5	4
24	1	3	2	3	4	4	5	3	2	4	4	1	5	1	4
25	3	2	2	3	6	5	3	1	3	2	1	1	1	1	3
26	3	4	4	1	2	5	4	2	4	2	1	1	1	2	1

	P	Q	R	S	T	U	V	W	X	Y	Z
1	Smoking Hab	Alcohol Consumptic	Stress Lev	Socioeconomic Stat	Living Conditio	Sanitatic	Environmental Hazard Exposu	Access to Healthca	Quality of Car	Cost of Car	Health Policies Awaren
2	4	2	3	3	4	4	2	3	3	5	1
3	2	3	3	1	2	4	1	3	4	5	2
4	4	3	3	2	3	1	2	4	2	3	1
5	4	4	2	4	1	5	5	1	2	2	3
6	4	1	4	3	4	1	2	3	4	5	3
7	2	2	5	2	5	3	2	1	3	4	4
8	2	2	1	1	2	1	2	3	1	1	4
9	2	4	3	4	1	2	3	1	3	1	2
10	3	1	2	2	2	2	1	5	5	2	3
11	2	2	2	3	5	3	5	2	1	2	3
12	4	2	3	2	4	1	4	1	1	5	2
13	1	4	5	1	3	1	1	4	4	1	3
14	1	1	4	2	2	4	4	5	1	3	3
15	1	4	5	1	3	5	2	3	4	3	1
16	1	3	2	1	4	1	5	3	2	4	3
17	1	2	4	2	1	1	5	2	1	4	1
18	3	1	3	1	4	5	4	5	4	2	1
19	3	2	2	1	4	2	1	3	5	4	1
20	3	1	2	1	1	2	2	2	5	2	3
21	1	4	1	3	5	1	4	3	5	1	2
22	1	2	2	2	1	5	4	3	5	3	2
23	3	4	3	4	2	4	3	1	4	5	3
24	1	4	5	4	3	2	3	2	4	1	2
25	1	1	4	4	3	1	2	1	4	3	2
26	2	1	5	3	1	3	1	3	5	2	3

26	3	4	4	1	2	5	4	2	4	2	1	1	1	2	1
27	3	3	3	1	5	3	4	5	1	3	2	2	5	4	3
28	5	2	4	1	4	5	4	4	3	3	4	5	2	5	2
29	3	3	4	3	5	5	3	3	1	1	3	3	4	5	2
30	5	4	2	4	2	2	5	1	3	2	3	4	1	3	4
31	4	3	2	5	5	4	1	5	3	1	5	5	2	5	3
32	5	2	2	3	6	1	1	2	2	5	2	1	2	3	2
33	3	1	4	3	2	5	6	2	1	5	5	1	1	2	1
34	1	4	4	2	4	4	6	4	5	3	5	2	5	2	1
35	6	4	1	5	3	3	4	2	4	5	1	3	1	1	3
36	2	3	2	2	1	3	4	3	2	2	3	3	4	4	4
37	6	2	4	3	4	4	6	5	4	5	5	4	3	2	2
38	4	1	5	2	2	4	5	4	2	1	4	1	4	3	5
39	1	3	5	1	6	1	1	3	4	2	4	2	4	1	5
40	6	2	2	4	5	1	6	4	5	1	4	1	4	1	3
41	4	4	3	2	4	4	5	2	4	2	4	3	1	2	1
42	2	4	3	5	2	4	3	4	3	3	2	5	2	5	3
43	3	4	5	3	3	4	4	5	4	3	3	2	5	2	3
44	1	3	1	2	6	2	3	3	3	1	2	5	5	3	4
45	3	2	3	1	2	4	3	4	3	4	4	4	3	5	2
46	1	4	2	3	3	5	6	1	2	1	4	3	3	3	5
47	4	3	1	5	6	5	1	5	1	5	5	2	2	2	3
48	5	2	5	1	4	5	3	2	4	2	3	3	3	1	4
49	4	2	2	5	6	3	6	5	1	1	3	1	4	4	1
50	1	4	1	5	3	1	6	1	5	3	2	3	3	1	5

27	3	3	2	2	3	1	5	4	5	5	2
28	2	3	2	1	5	4	4	3	1	2	3
29	2	4	3	3	5	2	1	5	4	1	1
30	3	4	3	4	5	3	4	1	1	1	4
31	1	2	2	2	5	5	1	5	3	4	3
32	3	2	3	2	2	4	1	2	2	1	1
33	1	2	4	4	1	4	2	2	3	5	4
34	1	1	2	2	4	2	2	3	4	1	1
35	4	2	5	3	2	2	5	5	1	3	2
36	3	1	2	3	4	1	5	5	5	5	1
37	2	3	1	1	5	1	5	4	3	3	4
38	2	2	3	3	3	4	5	3	4	3	1
39	2	1	2	4	3	2	1	2	1	2	4
40	3	3	2	3	2	5	2	5	5	1	4
41	4	1	4	3	2	5	2	2	1	3	1
42	3	4	4	4	1	3	5	5	5	1	1
43	4	4	1	1	4	2	3	3	3	3	4
44	4	3	2	2	3	3	2	4	3	4	2
45	1	4	5	4	3	1	3	5	1	2	4
46	4	3	4	2	4	3	4	2	5	1	1
47	2	3	4	4	4	1	4	4	1	4	3
48	3	1	2	1	3	2	2	4	2	4	2
49	2	2	5	2	3	5	5	2	1	5	4
50	1	2	3	3	4	5	5	3	5	3	1

51	6	4	4	1	3	2	3	2	1	1	1	2	1	4	3
52	3	1	5	1	3	5	1	3	4	1	5	5	5	4	1
53	6	1	5	2	4	5	6	4	2	5	1	4	3	5	3
54	4	2	2	1	1	2	1	1	3	5	1	3	1	4	5
55	5	3	2	2	4	3	5	4	2	5	3	2	5	1	4
56	4	4	2	3	3	1	3	1	5	2	4	1	3	4	5
57	4	2	5	2	4	2	3	2	3	5	4	1	5	4	3
58	6	1	3	2	5	1	3	3	4	5	2	4	2	5	1
59	5	1	4	4	4	5	2	1	4	2	5	1	5	1	3
60	5	4	3	2	3	5	2	2	5	5	4	2	4	5	5
61	1	1	1	1	6	4	1	2	4	1	1	3	1	4	5
62	2	1	1	4	1	4	2	1	4	1	4	4	1	5	4
63	5	4	3	4	4	5	1	1	4	4	1	1	5	4	3
64	6	4	2	5	5	4	4	5	1	2	1	4	5	1	2
65	6	1	3	4	1	3	1	5	5	3	1	2	3	3	1
66	4	4	5	5	2	1	3	3	2	5	1	3	5	5	1
67	4	4	2	3	3	2	4	1	3	3	3	4	4	4	4
68	1	3	4	2	2	1	1	5	4	4	3	1	1	3	5
69	6	2	5	2	5	2	6	3	1	5	5	1	1	3	2
70	6	3	5	5	2	2	4	3	4	5	4	2	5	3	5
71	2	2	5	5	2	4	6	3	3	2	4	1	3	2	1
72	5	4	4	2	5	1	5	2	2	1	5	2	2	4	5
73	6	1	1	4	5	4	4	2	5	5	5	4	4	4	5
74	6	1	2	5	4	4	3	4	5	3	4	5	2	2	5

51	4	4	3	4	4	5	5	1	5	5	3
52	1	1	3	1	4	4	2	4	5	3	1
53	3	2	5	4	2	4	4	5	2	3	3
54	4	3	2	4	5	3	2	1	2	2	3
55	1	2	3	1	2	3	1	2	4	4	2
56	4	2	3	3	2	4	2	1	2	2	2
57	2	2	5	3	1	4	1	4	3	5	4
58	1	1	4	3	4	1	5	2	1	3	2
59	4	4	1	1	3	4	2	4	4	4	4
60	1	1	2	1	1	5	1	2	3	4	1
61	2	4	4	3	5	4	1	1	1	1	2
62	3	1	1	4	5	5	2	3	4	4	2
63	4	1	2	1	5	1	4	2	3	5	1
64	3	1	2	1	1	3	3	4	1	4	2
65	1	2	5	3	5	4	1	5	5	4	1
66	4	1	5	2	3	2	2	4	1	2	2
67	4	3	4	3	2	4	1	4	3	3	2
68	3	4	4	2	5	2	4	4	1	3	4
69	3	4	4	2	2	1	4	4	4	4	3
70	4	4	2	1	5	5	3	4	3	5	2
71	2	2	2	2	5	2	5	2	2	3	4
72	2	4	3	4	5	4	1	5	4	5	2
73	1	1	2	2	3	3	3	3	3	1	1
74	2	1	4	2	4	4	1	5	2	3	4

APPENDIX C:

QUESTIONNAIRES

Age

1. 18-24 Years
2. 25-34 Years
3. 35-44 Years
4. 45-54 Years
5. 55-64 Years
6. 65 more than Years

Gender

1. Male
2. Female
3. Non-Binary
4. Others

Race

1. Asian
2. Black or African American
3. Native American
4. White
5. Others

Healthcare Access

1. Very Poor
2. Poor
3. Moderate
4. Good

5. Very Good

Education Level

1. Primary
2. Secondary
3. High School
4. Undergraduate
5. Graduate
6. Postgraduate

Annual Income (USD)

1. Below $50,000
2. $50,001 - $80,000
3. $80,001 - $100,000
4. $100,001 - $150,000
5. $150,001 - $200,000
6. Above $200,000

Employment Status

1. Part-time
2. Employed
3. Self-employed
4. Unemployed
5. Student
6. Retired

Housing Quality

1. Very Poor
2. Poor

3. Moderate
4. Good
5. Very Good

Pollution Level

1. Very Low
2. Low
3. Moderate
4. High
5. Very High

Crime Rate

1. Very Low
2. Low
3. Moderate
4. High
5. Very High

Nutrition Access

1. Very Poor
2. Poor
3. Moderate
4. Good
5. Very Good

Physical Activity Level

1. Very Low
2. Low
3. Moderate

4. High
5. Very High

Health Status

1. Very Unhealthy
2. Unhealthy
3. Moderate
4. Healthy
5. Very Healthy

Diet

1. Very Unhealthy
2. Unhealthy
3. Moderate
4. Healthy
5. Very Healthy

Exercise Level

1. Frequent
2. Occasional
3. Rare
4. Regular
5. None

Smoking Habit

1. Yes
2. No
3. Frequent
4. Occasional

Alcohol Consumption

1. Yes
2. No
3. Frequent
4. Occasional

Stress Level

1. Very Low
2. Low
3. Moderate
4. High
5. Very High

Socioeconomic Status

1. Low
2. Middle
3. Upper Middle
4. High

Living Conditions

1. Very Poor
2. Poor
3. Moderate
4. Good
5. Very Good

Sanitation

1. Very Poor
2. Poor

3. Moderate
4. Good
5. Very Good

Environmental Hazard Exposure

1. Very Low
2. Low
3. Moderate
4. High
5. Very High

Access to Healthcare

1. Very Difficult
2. Difficult
3. Moderate
4. Easy
5. Very Easy

Quality of Care

1. Very Poor
2. Poor
3. Moderate
4. Good
5. Very Good

Cost of Care

1. Very Affordable
2. Affordable
3. Moderate

4. Expensive
5. Very Expensive

Health Policies Awareness

1. Unaware
2. Aware
3. Partially Aware
4. Very Aware

REFERENCES

Adler, N. E., & Stewart, J. (2010). Preface to The Biology of Disadvantage: Socioeconomic Status and Health. In *Annals of the New York Academy of Sciences*. https://doi.org/10.1111/j.1749-6632.2009.05385.x

Alexeeff, S. E., Liao, N. S., Liu, X., Van Den Eeden, S. K., & Sidney, S. (2021). Long-term PM2.5 exposure and risks of ischemic heart disease and stroke events: Review and meta-analysis. *Journal of the American Heart Association*. https://doi.org/10.1161/JAHA.120.016890

Alzubi, J., Nayyar, A., & Kumar, A. (2018). Machine Learning from Theory to Algorithms: An Overview. *Journal of Physics: Conference Series*. https://doi.org/10.1088/1742-6596/1142/1/012012

Amaro, H. (2014). The action is upstream: place-based approaches for achieving population health and health equity. *American Journal of Public Health*. https://doi.org/10.2105/AJPH.2014.302032

An, K., Salyer, J., Brown, R. E., Kao, H. F. S., Starkweather, A., & Shim, I. (2016). Salivary Biomarkers of Chronic Psychosocial Stress and CVD Risks: A Systematic Review. In *Biological Research for Nursing*. https://doi.org/10.1177/1099800415604437

Ananth, C. V., Misra, D. P., Demissie, K., & Smulian, J. C. (2001). Rates of preterm delivery among black women and white women in the United States over two decades: An age-period-cohort analysis. *American Journal of Epidemiology*. https://doi.org/10.1093/aje/154.7.657

Andvik, C. (2010). *Social Determinants of Health in Very Poor Ruralities Striving and Thriving in Dire Conditions : in a Poor Rural District of Ghana*. *May*.

Bailey, K. T., Cook, J. T., Ettinger de Cuba, S., Casey, P. H., Chilton, M., Coleman, S. M., Cutts, D. B., Heeren, T. C., Rose-Jacobs, R., Black, M. M., & Frank, D. A. (2016). Development of an Index of Subsidized Housing Availability and its Relationship to Housing Insecurity. *Housing Policy Debate*. https://doi.org/10.1080/10511482.2015.1015042

Bak, B., Skrobala, A., Adamska, A., & Malicki, J. (2022). What information can we gain from performing adaptive radiotherapy of head and neck cancer patients from the past 10 years? In *Cancer/Radiotherapie*. https://doi.org/10.1016/j.canrad.2021.08.019

Barnidge, E., Stenmark, S., & Seligman, H. (2017). Clinic-to-community models to address food insecurity. In *JAMA Pediatrics*. https://doi.org/10.1001/jamapediatrics.2017.0067

Barraza, N., Moro, S., Ferreyra, M., & de la Peña, A. (2019). Mutual information and sensitivity analysis for feature selection in customer targeting: A comparative study. *Journal of Information Science*. https://doi.org/10.1177/0165551518770967

Bean, D. M., Stringer, C., Beeknoo, N., Teo, J., & Dobson, R. J. B. (2017). Network analysis of patient flow in two UK acute care hospitals identifies key sub-networks for A&E performance. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0185912

Bellodi, E., Zese, R., Riguzzi, F., & Lamma, E. (2022). Introduction to Machine Learning. In *Machine Learning and Non-volatile Memories*. https://doi.org/10.1007/978-3-031-03841-9_1

Bennett, D. E., Lee, J. H., Richards, P. S., & Inglehart, M. R. (2010). General dentists and periodontal referrals. *The Journal of the Michigan Dental Association*.

Berkman, N. D., Sheridan, S. L., Donahue, K. E., Halpern, D. J., & Crotty, K. (2011). Low Health Literacy and Health Outcomes: An Updated Systematic Review. *Annals of Internal Medicine*, *155*(2), 97. https://doi.org/10.7326/0003-4819-155-2-201107190-00005

Berry, L. L., Deming, K. A., & Danaher, T. S. (2018). Improving Nonclinical and Clinical-Support Services: Lessons From Oncology. *Mayo Clinic Proceedings: Innovations, Quality & Outcomes*. https://doi.org/10.1016/j.mayocpiqo.2018.05.002

Betof, N., Salkin, L. M., & Ferris, R. T. (1985). Why general dentists refer patients to periodontists. *Journal of Dental Practice Administration : JDPA : Official Publication of American Academy of Dental Practice Administration, Organization of Teachers of Dental Practice Administration, American Academy of Dental Group Practice*.

Bozic, K. J., & Wright, J. G. (2012). Value-based healthcare and orthopaedic surgery: Editorial comment. In *Clinical Orthopaedics and Related Research*. https://doi.org/10.1007/s11999-012-2267-x

Brach, C., Keller, D., Hernandez, L., Baur, C., Parker, R., Dreyer, B., Schyve, P., Lemerise, A. J., & Schillinger, D. (2012). Ten Attributes of Health Literate Health Care Organizations. *NAM Perspectives*, *02*(6). https://doi.org/10.31478/201206a

Bradley, E. H., Canavan, M., Rogan, E., Talbert-Slagle, K., Ndumele, C., Taylor, L., & Curry, L. A. (2016). Variation In Health Outcomes: The Role Of Spending On Social Services, Public Health, And Health Care, 2000–09. *Health Affairs*, *35*(5), 760–768. https://doi.org/10.1377/hlthaff.2015.0814

Bradley, E. H., Elkins, B. R., Herrin, J., & Elbel, B. (2011). Health and social services expenditures: Associations with health outcomes. *BMJ Quality and Safety*. https://doi.org/10.1136/bmjqs.2010.048363

Braveman, P. A., Cubbin, C., Egerter, S., Williams, D. R., & Pamuk, E. (2010). Socioeconomic disparities in health in the united States: What the patterns tell us. *American Journal of Public Health*. https://doi.org/10.2105/AJPH.2009.166082

Braveman, P. A., Egerter, S. A., Woolf, S. H., & Marks, J. S. (2011). When do we know enough to recommend action on the social determinants of health? *American Journal of Preventive Medicine*. https://doi.org/10.1016/j.amepre.2010.09.026

Braveman, P., Egerter, S., & Williams, D. R. (2011). The social determinants of health: Coming of age. *Annual Review of Public Health*. https://doi.org/10.1146/annurev-publhealth-031210-101218

Bureau of Labor Statistics. (2017). THE EMPLOYMENT SITUATION. *News Release, Bureau of Labor Statistics*, 1–39.

Call, K. T., McAlpine, D. D., Garcia, C. M., Shippee, N., Beebe, T., Adeniyi, T. C., & Shippee, T. (2014). Barriers to care in an ethnically diverse publicly insured population: Is health care reform enough? *Medical Care*. https://doi.org/10.1097/MLR.0000000000000172

Carlson, S. J., Andrews, M. S., & Bickel, G. W. (1999). Measuring food insecurity and hunger in the United States: Development of a national benchmark measure and prevalence estimates. *Journal of Nutrition*. https://doi.org/10.1093/jn/129.2.510s

Carroll-Scott, A., Gilstad-Hayden, K., Rosenthal, L., Peters, S. M., McCaslin, C., Joyce, R., & Ickovics, J. R. (2013). Disentangling neighborhood contextual associations with child body mass index, diet, and physical activity: The role of built, socioeconomic, and social environments. *Social Science and Medicine*. https://doi.org/10.1016/j.socscimed.2013.04.003

Caruso, C. C., Hitchcock, E. M., Robert B. Dick, Russo, J. M., & Schmit, J. M. (2004). Overtime and Extended Work Shifts: Recent findings on illness, injuries and health behaviours. *Niosh*.

Castañeda, H., Holmes, S. M., Madrigal, D. S., Young, M. E. D. T., Beyeler, N., & Quesada, J. (2015). Immigration as a social determinant of health. In *Annual Review of Public Health*. https://doi.org/10.1146/annurev-publhealth-032013-182419

Chen, M., Liu, Q., Chen, S., Liu, Y., Zhang, C., & Liu, R. (2019). XGBoost-Based Algorithm Interpretation and Application on Post-Fault Transient Stability Status Prediction of Power System. *IEEE Access*, *7*, 13149–13158. https://doi.org/10.1109/ACCESS.2019.2893448

Ciaburro, G., Iannace, G., Ali, M., Alabdulkarem, A., & Nuhait, A. (2021). An artificial neural network approach to modelling absorbent asphalts acoustic properties. *Journal of King Saud University - Engineering Sciences*. https://doi.org/10.1016/j.jksues.2020.07.002

Cole, M. B., Trivedi, A. N., Wright, B., & Carey, K. (2018). Health Insurance Coverage and Access to Care for Community Health Center Patients: Evidence Following the Affordable Care Act. *Journal of General Internal Medicine*. https://doi.org/10.1007/s11606-018-4499-2

Conrad, D. A. (2015). The Theory of Value-Based Payment Incentives and Their Application to Health Care. *Health Services Research*. https://doi.org/10.1111/1475-6773.12408

Corral, I., & Landrine, H. (2012). Racial discrimination and health-promoting vs damaging behaviors among African-American adults. *Journal of Health Psychology*. https://doi.org/10.1177/1359105311435429

Crowley, S. (2003). The Affordable Housing Crisis: Residential Mobility of Poor Families and School Mobility of Poor Children. *The Journal of Negro Education*. https://doi.org/10.2307/3211288

Cutler, D. M., & Lleras-Muney, A. (2006). Education and Health: Evaluating Theories and Evidence. *National Bureau of Economic Research*. https://doi.org/10.3386/w12352

Debaun, M. R., Chen, M. J., Bishop, J. A., Gardner, M. J., & Kamal, R. N. (2019). Orthopaedic Trauma Quality Measures for Value-Based Health Care Delivery: A Systematic Review. In *Journal of Orthopaedic Trauma*. https://doi.org/10.1097/BOT.0000000000001372

DeVoe, J. E., Bazemore, A. W., Cottrell, E. K., Likumahuwa-Ackman, S., Grandmont, J., Spach, N., & Gold, R. (2016). Perspectives in Primary Care: A Conceptual Framework and Path for Integrating Social Determinants of Health Into Primary Care Practice. In *Chinese General Practice*. https://doi.org/10.1370/afm.1903

Eamon, M. K. (2001). The effects of poverty on children's socioemotional development: An ecological systems analysis. *Social Work*. https://doi.org/10.1093/sw/46.3.256

Fazeli Dehkordy, S., Hall, K. S., Dalton, V. K., & Carlos, R. C. (2016). The Link between Everyday Discrimination, Healthcare Utilization, and Health Status among a National Sample of Women. *Journal of Women's Health*. https://doi.org/10.1089/jwh.2015.5522

Ferrer, R. L. (2023). Social Determinants of Health. In *Chronic Illness Care: Principles and Practice, Second Edition*. https://doi.org/10.1007/978-3-031-29171-5_40

Franck, C., Grandi, S. M., & Eisenberg, M. J. (2013). Agricultural subsidies and the american obesity epidemic. *American Journal of Preventive Medicine*. https://doi.org/10.1016/j.amepre.2013.04.010

Franklin, J. (2005). The elements of statistical learning: data mining, inference and prediction. In *Mathematical Intelligencer*. https://doi.org/10.1007/BF02985802

Freedman, V. A., Grafova, I. B., & Rogowski, J. (2011). Neighborhoods and chronic disease onset in later life. *American Journal of Public Health*. https://doi.org/10.2105/AJPH.2009.178640

Giuse, N. B., Koonce, T. Y., Kusnoor, S. V., Prather, A. A., Gottlieb, L. M., Huang, L. C., Phillips, S. E., Shyr, Y., Adler, N. E., & Stead, W. W. (2017). Institute of Medicine Measures of Social and Behavioral Determinants of Health: A Feasibility Study. *American Journal of Preventive Medicine*. https://doi.org/10.1016/j.amepre.2016.07.033

Glasgow, R. E., & Emmons, K. M. (2007). How can we increase translation of research into practice? Types of evidence needed. *Annual Review of Public Health*. https://doi.org/10.1146/annurev.publhealth.28.021406.144145

Goesling, B. (2007). The rising significance of education for health? *Social Forces*. https://doi.org/10.1353/sof.2007.0068

Gostin, L. O. (2016). Politics and Public Health: The Flint Drinking Water Crisis. *Hastings Center Report*. https://doi.org/10.1002/hast.598

Gottlieb, L., Hessler, D., Long, D., Amaya, A., & Adler, N. (2014). A randomized trial on screening for social determinants of health: The iScreen study. *Pediatrics*. https://doi.org/10.1542/peds.2014-1439

Gottlieb, L. M., Tirozzi, K. J., Manchanda, R., Burns, A. R., & Sandel, M. T. (2015). Moving electronic medical records upstream incorporating social determinants of health. *American Journal of Preventive Medicine*. https://doi.org/10.1016/j.amepre.2014.07.009

Gove, W. R., Hughes, M., & Galle, O. R. (1979). Overcrowding in the home: an empirical investigation of its possible pathological consequences. *American Sociological Review*. https://doi.org/10.2307/2094818

Gray, J. A. M. (2006). Redefining Health Care: Creating Value-Based Competition on Results. *BMJ*. https://doi.org/10.1136/bmj.333.7571.760

Grundy, E. (2005). Commentary: The McKeown debate: Time for burial. *International Journal of Epidemiology*. https://doi.org/10.1093/ije/dyh272

Gu, L., Zeng, D., Li, P., & Guo, S. (2014). Cost minimization for big data processing in geo-distributed data centers. *IEEE Transactions on Emerging Topics in Computing*. https://doi.org/10.1109/TETC.2014.2310456

Hacker, K., & Walker, D. K. (2013). Achieving population health in accountable care organizations. In *American Journal of Public Health*. https://doi.org/10.2105/AJPH.2013.301254

Hadley, J., Ayanian, J. Z., Butler, S., Davis, K., & Kronick, R. (2003). Sicker and poorer - The consequences of being uninsured: A review of the research on the relationship between health insurance, medical care use, health, work, and income. In *Medical Care Research and Review*. https://doi.org/10.1177/1077558703254101

Hager, L. D. (2002). Hearing protection. Didn't hear it coming ... noise and hearing in industrial accidents. *Occupational Health & Safety (Waco, Tex.)*.

Hahn, R. A. (2021). What is a social determinant of health? Back to basics. *Journal of Public Health Research*. https://doi.org/10.4081/jphr.2021.2324

Hartinger-Saunders, R. M., Rine, C. M., Nochajski, T., & Wieczorek, W. (2012). Neighborhood crime and perception of safety as predictors of victimization and offending among youth: A call for macro-level prevention and intervention models. *Children and Youth Services Review*, *34*(9), 1966–1973. https://doi.org/10.1016/j.childyouth.2012.05.020

Hatef, E., Ma, X., Rouhizadeh, M., Singh, G., Weiner, J. P., & Kharrazi, H. (2021). Assessing the Impact of Social Needs and Social Determinants of Health on Health Care Utilization: Using Patient-and Community-Level Data. *Population Health Management*. https://doi.org/10.1089/pop.2020.0043

He, Z., Pfaff, E., Guo, S. J., Guo, Y., Wu, Y., Tao, C., Stiglic, G., & Bian, J. (2023). Enriching Real-world Data with Social Determinants of Health for Health Outcomes and Health Equity: Successes, Challenges, and Opportunities. *Yearbook of Medical Informatics*. https://doi.org/10.1055/s-0043-1768732

Herd, P., Schoeni, R. F., & House, J. S. (2008). Upstream solutions: Does the supplemental security income program reduce disability in the elderly? *Milbank Quarterly*. https://doi.org/10.1111/j.1468-0009.2007.00512.x

Hoel, H., Sparks, K., & Cooper, C. L. (2001). The cost of violence/stress at work and the benefits of a violence/stress-free working environment. *Geneva: International Labour Organization*.

Hood, C. M., Gennuso, K. P., Swain, G. R., & Catlin, B. B. (2016). County Health Rankings: Relationships between Determinant Factors and Health Outcomes. *American Journal of Preventive Medicine*. https://doi.org/10.1016/j.amepre.2015.08.024

Hu, F. B., Manson, J. E., Stampfer, M. J., Colditz, G., Liu, S., Solomon, C. G., & Willett, W. C. (2001). Diet, Lifestyle, and the Risk of Type 2 Diabetes Mellitus in Women. *New England Journal of Medicine*. https://doi.org/10.1056/nejmoa010492

Huang, J., Guo, B., & Kim, Y. (2010). Food insecurity and disability: Do economic resources matter? *Social Science Research*. https://doi.org/10.1016/j.ssresearch.2009.07.002

Jack, K., Evans, C., Bramley, L., Cooper, J., Keane, T., Cope, M., & Hendron, E. (2022). Identifying and Understanding the Non-clinical Impacts of Delayed or Cancelled Surgery in Order to Inform Prioritisation Processes: A Scoping Review. In *International Journal of Environmental Research and Public Health*. https://doi.org/10.3390/ijerph19095542

JCHS Harvard University. (2020). The State of the Nation's Housing 2021. *Harvard Joint Center on Housing Studies*.

Jones, A. D., Ngure, F. M., Pelto, G., & Young, S. L. (2013). What are we assessing when we measure food security? A compendium and review of current metrics. In *Advances in Nutrition*. https://doi.org/10.3945/an.113.004119

Kang, S. (2016). Inequality and crime revisited: effects of local inequality and economic segregation on crime. *Journal of Population Economics*. https://doi.org/10.1007/s00148-015-0579-3

Kelly, M., Morgan, A., Bonnefoy, J., Butt, J., & Bergman, V. (2007). The social determinants of health: developing an evidence base for political action. *WHO Final Report to the Commission*.

Kevin S. Blake, Kellerson, R. L., & Simic, A. (2007). Measuring overcrowding in housing. *U.S. Department of Housing and Urban Development Office of Policy Development and Research*, 38.

Kingdom U. (2023). Poverty Rate by Age. *American Community Survey*.

Kirschner, K. L., Breslin, M. Lou, & Iezzoni, L. I. (2007). Structural impairments that limit access to health care for patients with disabilities. In *JAMA*. https://doi.org/10.1001/jama.297.10.1121

Krause, T. M., Schaefer, C., & Highfield, L. (2021). The association of social determinants of health with health outcomes. *American Journal of Managed Care*. https://doi.org/10.37765/ajmc.2021.88603

Kreindler, S. A. (2017a). Six ways not to improve patient flow: a qualitative study. *BMJ Quality & Safety*, *26*(5), 388–394. https://doi.org/10.1136/bmjqs-2016-005438

Kreindler, S. A. (2017b). The three paradoxes of patient flow: An explanatory case study. *BMC Health Services Research*. https://doi.org/10.1186/s12913-017-2416-8

Kreuter, M. W., Thompson, T., McQueen, A., & Garg, R. (2020). Addressing Social Needs in Health Care Settings: Evidence, Challenges, and Opportunities for Public Health. In *Annual Review of Public Health*. https://doi.org/10.1146/annurev-publhealth-090419-102204

Kriegel, J., Jehle, F., Dieck, M., & Tuttle-Weidinger, L. (2015). Optimizing patient flow in Austrian hospitals – Improvement of patient-centered care by coordinating hospital-wide patient trails. *International Journal of Healthcare Management*, *8*(2), 89–99. https://doi.org/10.1179/2047971914Y.0000000093

Kriegel, J., Jehle, F., Moser, H., & Tuttle-Weidinger, L. (2016). Patient logistics management of patient flows in hospitals: A comparison of Bavarian and Austrian hospitals. *International Journal of Healthcare Management*, *9*(4), 257–268. https://doi.org/10.1080/20479700.2015.1119370

Krieger, J., & Higgins, D. L. (2002). Housing and health: Time again for public health action. In *American Journal of Public Health*. https://doi.org/10.2105/AJPH.92.5.758

Kripalani, S., Wallston, K., Cavanaugh, K. L., Osborn, C. Y., Mulvaney, S., Scott, A. M.,

& Rothman, R. L. (2014). MEASURES TO ASSESS A HEALTH-LITERATE ORGANIZATION. *National Academies of Medicine*, *July*.

Krug, E. G., Mercy, J. A., Dahlberg, L. L., & Zwi, A. B. (2002). The world report on violence and health. *Lancet*. https://doi.org/10.1016/S0140-6736(02)11133-0

Kumar, S., Calvo, R., Avendano, M., Sivaramakrishnan, K., & Berkman, L. F. (2012). Social support, volunteering and health around the world: Cross-national evidence from 139 countries. *Social Science and Medicine*. https://doi.org/10.1016/j.socscimed.2011.11.017

Kunz, J. M., Lawinger, H., Miko, S., Gerdes, M., Thuneibat, M., Hannapel, E., & Roberts, V. A. (2024). Surveillance of Waterborne Disease Outbreaks Associated with Drinking Water — United States, 2015–2020. *MMWR Surveillance Summaries*. https://doi.org/10.15585/mmwr.ss7301a1

Kushel, M. B., Gupta, R., Gee, L., & Haas, J. S. (2006). Housing instability and food insecurity as barriers to health care among low-income Americans. *Journal of General Internal Medicine*. https://doi.org/10.1111/j.1525-1497.2005.00278.x

Laub, J. A. (1999). Assessing the servant organization; Development of the Organizational Leadership Assessment (OLA) model. Dissertation Abstracts International,. *Procedia - Social and Behavioral Sciences*.

Lavelle, M., Reedy, G. B., Attoe, C., Simpson, T., & Anderson, J. E. (2019). Beyond the clinical team: evaluating the human factors-oriented training of non-clinical professionals working in healthcare contexts. *Advances in Simulation*, *4*(1), 11. https://doi.org/10.1186/s41077-019-0101-1

Lee, J. H., Bennett, D. E., Richards, P. S., & Inglehart, M. R. (2009). Periodontal Referral Patterns of General Dentists: Lessons for Dental Education. *Journal of Dental Education*, *73*(2), 199–210. https://doi.org/10.1002/j.0022-0337.2009.73.2.tb04655.x

Linden, G. J. (1998). Variation in periodontal referral by general dental practitioners. *Journal of Clinical Periodontology*, *25*(8), 655–661. https://doi.org/10.1111/j.1600-051X.1998.tb02502.x

Linden, G. J., Stevenson, M., & Burke, F. J. T. (1999). Variation in periodontal referral in

2 regions in theUK. *Journal of Clinical Periodontology*, *26*(9), 590–595. https://doi.org/10.1034/j.1600-051X.1999.260905.x

Loucks, E. B., Gilman, S. E., Howe, C. J., Kawachi, I., Kubzansky, L. D., Rudd, R. E., Martin, L. T., Nandi, A., Wilhelm, A., & Buka, S. L. (2015). Education and Coronary Heart Disease Risk: Potential Mechanisms Such as Literacy, Perceived Constraints, and Depressive Symptoms. *Health Education and Behavior*. https://doi.org/10.1177/1090198114560020

Ludwig, J., Sanbonmatsu, L., Gennetian, L., Adam, E., Duncan, G. J., Katz, L. F., Kessler, R. C., Kling, J. R., Lindau, S. T., Whitaker, R. C., & McDade, T. W. (2011). Neighborhoods, Obesity, and Diabetes — A Randomized Social Experiment. *New England Journal of Medicine*. https://doi.org/10.1056/nejmsa1103216

Lukachko, A., Hatzenbuehler, M. L., & Keyes, K. M. (2014). Structural racism and myocardial infarction in the United States. *Social Science & Medicine*, *103*, 42–50. https://doi.org/10.1016/j.socscimed.2013.07.021

Luo, Y., Xu, J., Granberg, E., & Wentworth, W. M. (2012). A longitudinal study of social status, perceived discrimination, and physical and emotional health among older adults. *Research on Aging*. https://doi.org/10.1177/0164027511426151

Lybrand, K. E., & Althausen, P. L. (2018). The Role of Value-Based Implants in Orthopedic Trauma. In *Orthopedic Clinics of North America*. https://doi.org/10.1016/j.ocl.2018.05.005

Mackenbach, J. P., Kunst, A. E., Cavelaars, A. E. J. M., Groenhof, F., & Geurts, J. J. M. (1997). Socioeconomic inequalities in morbidity and mortality in western Europe. *Lancet*. https://doi.org/10.1016/S0140-6736(96)07226-1

Mackenbach, J. P., Stirbu, I., Roskam, A.-J. R., Schaap, M. M., Menvielle, G., Leinsalu, M., & Kunst, A. E. (2008). Socioeconomic Inequalities in Health in 22 European Countries. *New England Journal of Medicine*. https://doi.org/10.1056/nejmsa0707519

Majerol, M., Newkirk, V., & Garfield, R. (2015). The Uninsured: A Primer. Key Facts about Health Insurance and the Uninsured in the Era of Health Reform. *The Kaiser*

Commission on Medicaid and the Uninsured.

Malamou, T. (2015). Social determinants of health. In *Nosileftiki*. https://doi.org/10.1097/nor.0000000000000829

Manning, L., & Islam, M. S. (2023). A systematic review to identify the challenges to achieving effective patient flow in public hospitals. *International Journal of Health Planning and Management*. https://doi.org/10.1002/hpm.3626

Marmot, M. (2005). Social determinants of health inequalities. *Lancet*. https://doi.org/10.1016/S0140-6736(05)71146-6

Martin, J. K., Tuch, S. A., & Roman, P. M. (2003). Problem Drinking Patterns among African Americans: The Impacts of Reports of Discrimination, Perceptions of Prejudice, and "Risky" Coping Strategies. *Journal of Health and Social Behavior*, *44*(3), 408. https://doi.org/10.2307/1519787

Martinson, M. L. (2012). Income inequality in health at all ages: A comparison of the United States and England. *American Journal of Public Health*. https://doi.org/10.2105/AJPH.2012.300929

Mays, V. M., & Cochran, S. D. (2001). Mental health correlates of perceived discrimination among lesbian, gay, and bisexual adults in the United States. *American Journal of Public Health*. https://doi.org/10.2105/AJPH.91.11.1869

McGinnis, J. M., & Foege, W. H. (1993). Actual Causes of Death in the United States. *JAMA: The Journal of the American Medical Association*. https://doi.org/10.1001/jama.1993.03510180077038

McGinnis, J. M., Williams-Russo, P., & Knickman, J. R. (2002). The case for more active policy attention to health promotion. *Health Affairs*. https://doi.org/10.1377/hlthaff.21.2.78

McKeown, T., Record, R. G., & Turner, R. D. (1975). An Interpretation of the Decline of Mortality in England and Wales during the Twentieth Century. *Population Studies*. https://doi.org/10.1080/00324728.1975.10412707

Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2017). Deep learning for healthcare: Review, opportunities and challenges. *Briefings in Bioinformatics*.

https://doi.org/10.1093/bib/bbx044

Myers, V., Drory, Y., Goldbourt, U., & Gerber, Y. (2014). Multilevel socioeconomic status and incidence of frailty post myocardial infarction. *International Journal of Cardiology*. https://doi.org/10.1016/j.ijcard.2013.11.009

National Institute for Occupational Safety and Health. (2004). Worker Health Chartbook. In *Hear My Sad Story* (pp. 1–382). Cornell University Press.

Nord, M. (2007). Characteristics of Low-Income Households with Very Low Food Security: An Analysis of the USDA GPRA Food Security Indicator. *United States Department of Agriculture*, *May*. https://doi.org/10.2139/ssrn.1084839

Nord, M. (2009). Food Insecurity in Households With Children: Prevalence , Severity , and Household Characteristics. *USDA Economic Research Report*.

Palermo, T. M., & Beals-Erickson, S. E. (2015). Electronic health records and the value of psychology in the health care system. *PsycCRITIQUES*. https://doi.org/10.1037/a0039423

Pascoe, E. A., & Richman, L. S. (2009). Perceived Discrimination and Health: A Meta-Analytic Review. *Psychological Bulletin*. https://doi.org/10.1037/a0016059

PEI, L. W. (2024). *Social Determinants of Health*.

Platz, E. A., Willett, W. C., Colditz, G. A., Rimm, E. B., Spiegelman, D., & Giovannucci, E. (2000). Proportion of colon cancer risk that might be preventable in a cohort of middle-aged US men. *Cancer Causes and Control*. https://doi.org/10.1023/A:1008999232442

Porter, M. E. (2009). A Strategy for Health Care Reform — Toward a Value-Based System. *New England Journal of Medicine*. https://doi.org/10.1056/nejmp0904131

Porter, M. E., & Teisberg, E. O. (2007). How physicians can change the future of health care. In *JAMA*. https://doi.org/10.1001/jama.297.10.1103

Powell, L. M., Slater, S., Mirtcheva, D., Bao, Y., & Chaloupka, F. J. (2007). Food store availability and neighborhood characteristics in the United States. *Preventive Medicine*, *44*(3), 189–195. https://doi.org/10.1016/j.ypmed.2006.08.008

Prather, A. A., Gottlieb, L. M., Giuse, N. B., Koonce, T. Y., Kusnoor, S. V., Stead, W. W.,

& Adler, N. E. (2017). National Academy of Medicine Social and Behavioral Measures: Associations With Self-Reported Health. *American Journal of Preventive Medicine*. https://doi.org/10.1016/j.amepre.2017.02.010

Prus, S. G. (2011). Comparing social determinants of self-rated health across the United States and Canada. *Social Science and Medicine*. https://doi.org/10.1016/j.socscimed.2011.04.010

Pryor, C., & Gurewich, D. (2004). Getting Care But Paying the Price: How Medical Debt Leaves Many in Massachusetts Facing Tough Choices. *The Access Project,*.

Qin, S., Thompson, C., Bogomolov, T., Ward, D., & Hakendorf, P. (2017). Hospital occupancy and discharge strategies: a simulation-based study. *Internal Medicine Journal*. https://doi.org/10.1111/imj.13485

Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., & Yang, G. Z. (2017). Deep Learning for Health Informatics. *IEEE Journal of Biomedical and Health Informatics*. https://doi.org/10.1109/JBHI.2016.2636665

Ray, J. C., & Kusumoto, F. (2016). The transition to value-based care. In *Journal of Interventional Cardiac Electrophysiology*. https://doi.org/10.1007/s10840-016-0166-x

Repetti, R. L., & Wang, S.-W. (2014). Employment and Parenting. *Parenting*, *14*(2), 121–132. https://doi.org/10.1080/15295192.2014.914364

Sabbath, E. L., Mejía-Guevara, I., Noelke, C., & Berkman, L. F. (2015). The long-term mortality impact of combined job strain and family circumstances: A life course analysis of working American mothers. *Social Science and Medicine*. https://doi.org/10.1016/j.socscimed.2015.10.024

Saha, S., Vaidyanathan, A., Lo, F., Brown, C., & Hess, J. J. (2021). Short term physician visits and medication prescriptions for allergic disease associated with seasonal tree, grass, and weed pollen exposure across the United States. *Environmental Health*, *20*(1), 85. https://doi.org/10.1186/s12940-021-00766-3

Samuel, A. L. (2000). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*. https://doi.org/10.1147/rd.441.0206

Schaider, L. A., Swetschinski, L., Campbell, C., & Rudel, R. A. (2019). Environmental justice and drinking water quality: Are there socioeconomic disparities in nitrate levels in U.S. drinking water? *Environmental Health: A Global Access Science Source*. https://doi.org/10.1186/s12940-018-0442-6

Schieman, S., & Reid, S. (2009). Job authority and health: Unraveling the competing suppression and explanatory influences. *Social Science and Medicine*. https://doi.org/10.1016/j.socscimed.2009.08.038

Schnoor, J. L. (2016). Recognizing Drinking Water Pipes as Community Health Hazards. In *Journal of Chemical Education*. https://doi.org/10.1021/acs.jchemed.6b00218

Seo, V., Baggett, T. P., Thorndike, A. N., Hull, P., Hsu, J., Newhouse, J. P., & Fung, V. (2019). Access to care among Medicaid and uninsured patients in community health centers after the Affordable Care Act. *BMC Health Services Research*. https://doi.org/10.1186/s12913-019-4124-z

Shahid, N., Rappon, T., & Berta, W. (2019). Applications of artificial neural networks in health care organizational decision-making: A scoping review. In *PLoS ONE*. https://doi.org/10.1371/journal.pone.0212356

Shain, M., & Kramer, D. M. (2004). HEALTH PROMOTION IN THE WORKPLACE: FRAMING THE CONCEPT; REVIEWING THE EVIDENCE. *Occupational and Environmental Medicine*, *61*(7), 643–648. https://doi.org/10.1136/oem.2004.013193

Sharkey, J. R., Johnson, C. M., & Dean, W. R. (2011). Relationship of household food insecurity to health-related quality of life in a large sample of rural and urban women. *Women and Health*. https://doi.org/10.1080/03630242.2011.584367

Sharpe, G., Durham, J. A., & Preshaw, P. M. (2007). Attitudes regarding specialist referrals in periodontics. *British Dental Journal*, *202*(4), E11–E11. https://doi.org/10.1038/bdj.2007.141

Shavers, V. L., Fagan, P., Jones, D., Klein, W. M. P., Boyington, J., Moten, C., & Rorie, E. (2012). The state of research on racial/ethnic discrimination in the receipt of health care. In *American Journal of Public Health*. https://doi.org/10.2105/AJPH.2012.300773

Sheats, K. J., Irving, S. M., Mercy, J. A., Simon, T. R., Crosby, A. E., Ford, D. C., Merrick, M. T., Annor, F. B., & Morgan, R. E. (2018). Violence-Related Disparities Experienced by Black Youth and Young Adults: Opportunities for Prevention. *American Journal of Preventive Medicine*, *55*(4), 462–469. https://doi.org/10.1016/j.amepre.2018.05.017

Shrider, E. A., Kollar, M., Chen, F., & Semega, J. (2021). Income and Poverty in the United States: 2020. *U.S. Census Bureau, Current Population Reports, P60-273*.

Shukla, P., Krishna, C. R., & Patil, N. V. (2024). Iot traffic-based DDoS attacks detection mechanisms: A comprehensive review. *Journal of Supercomputing*. https://doi.org/10.1007/s11227-023-05843-7

Simmons, L. A., & Swanberg, J. E. (2009). Psychosocial work environment and depressive symptoms among US workers: Comparing working poor and working non-poor. *Social Psychiatry and Psychiatric Epidemiology*. https://doi.org/10.1007/s00127-008-0479-x

Singh, G. K., & Siahpush, M. (2006). Widening socioeconomic inequalities in US life expectancy, 1980–2000. *International Journal of Epidemiology*, *35*(4), 969–979. https://doi.org/10.1093/ije/dyl083

Smedley, B. D., Stith, A. Y., & Nelson, A. R. (2003). *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care* (W. (DC): N. A. P. (US) (ed.)). National Academies Press.

Social Determinants of Health. (2019). *The Importance of Social Determinants of Health Data*.

Song, Y. Y., & Lu, Y. (2015). Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry*. https://doi.org/10.11919/j.issn.1002-0829.215044

South, E. C., MacDonald, J. M., Tam, V. W., Ridgeway, G., & Branas, C. C. (2023). Effect of Abandoned Housing Interventions on Gun Violence, Perceptions of Safety, and Substance Use in Black Neighborhoods. *JAMA Internal Medicine*, *183*(1), 31. https://doi.org/10.1001/jamainternmed.2022.5460

Sutabri, T., Pandi Selvam, R., Shankar, K., Nguyen, P. T., Hashim, W., & Maseleno, A. (2019). Machine learning for healthcare diagnostics. *International Journal of Engineering and Advanced Technology*. https://doi.org/10.35940/ijeat.F1304.0886S219

Sutton, R. S., & Barto, A. G. (2005). Reinforcement Learning: An Introduction. *IEEE Transactions on Neural Networks*. https://doi.org/10.1109/tnn.1998.712192

Swope, C. B., & Hernández, D. (2019). Housing as a determinant of health equity: A conceptual model. In *Social Science and Medicine*. https://doi.org/10.1016/j.socscimed.2019.112571

Sykes, B. L., Piquero, A. R., & Gioviano, J. P. (2017). Adolescent Racial Discrimination and Parental Perceptions of Safety in American Neighborhoods and Schools. *Sociological Forum*, *32*(S1), 952–974. https://doi.org/10.1111/socf.12364

Talukder, M. A., Islam, M. M., Uddin, M. A., Hasan, K. F., Sharmin, S., Alyami, S. A., & Moni, M. A. (2024). Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction. *Journal of Big Data*. https://doi.org/10.1186/s40537-024-00886-w

Tessum, C. W., Apte, J. S., Goodkind, A. L., Muller, N. Z., Mullins, K. A., Paolella, D. A., Polasky, S., Springer, N. P., Thakrar, S. K., Marshall, J. D., & Hill, J. D. (2019). Inequity in consumption of goods and services adds to racial–ethnic disparities in air pollution exposure. *Proceedings of the National Academy of Sciences*, *116*(13), 6001–6006. https://doi.org/10.1073/pnas.1818859116

Turner, M. C., Krewski, D., Pope, C. A., Chen, Y., Gapstur, S. M., & Thun, M. J. (2011). Long-term ambient fine particulate matter air pollution and lung cancer in a large cohort of never-smokers. *American Journal of Respiratory and Critical Care Medicine*. https://doi.org/10.1164/rccm.201106-1011OC

U.S. DEPARTMENT OF AGRICULTURE. (2024). *Food Security and Nutrition Assistance*. Economic Research Service.

Ver Ploeg, M. (2009). Access to Affordable and Nutritious Food : Measuring and Understanding Food Deserts and Their Consequences Report to Congress.

Agriculture.

Weaver, N. L., Wray, R. J., Zellin, S., Gautam, K., & Jupka, K. (2012). Advancing organizational health literacy in health care organizations serving high-needs populations: A case study. *Journal of Health Communication*. https://doi.org/10.1080/10810730.2012.714442

Weich, S., Blanchard, M., Prince, M., Burton, E., Erens, B., & Sproston, K. (2002). Mental health and the built environment: Cross-sectional survey of individual and contextual risk factors for depression. *British Journal of Psychiatry*. https://doi.org/10.1192/bjp.180.5.428

Winasti, W., Elkhuizen, S., Berrevoets, L., van Merode, G., & Berden, H. (2018). Inpatient flow management: a systematic review. *International Journal of Health Care Quality Assurance*, *31*(7), 718–734. https://doi.org/10.1108/IJHCQA-03-2017-0054

Zemanovich, M. R., Bogacki, R. E., Abbott, D. M., Maynard, J. G., & Lanning, S. K. (2006). Demographic Variables Affecting Patient Referrals From General Practice Dentists to Periodontists. *Journal of Periodontology*, *77*(3), 341–349. https://doi.org/10.1902/jop.2006.050125

Zenk, S. N., Schulz, A. J., Israel, B. A., James, S. A., Bao, S., & Wilson, M. L. (2005). Neighborhood racial composition, neighborhood poverty, and the spatial accessibility of supermarkets in metropolitan Detroit. *American Journal of Public Health*. https://doi.org/10.2105/AJPH.2004.042150