Investigating the Impact of SDOH and Chronic Disease Risk Factors in Underserved Communities with System Dynamics and Machine Learning

Student: Ahmeed A. Yinusa
Supervisor: Dr. Misa Faezipour

August 17, 2023

Abstract

Chronic diseases are medical burdens affecting individuals, and their prevalence is responsible for the greater part of the accumulated deaths globally. In the US, the onsets and prevalences of these acute diseases disproportionately affect individuals differently, especially those in minority-populated communities resulting in adverse health outcomes such as disability, poor mental health, and mortality. These chronic condition burdens are often motivated by the adverse influence of the complex and dynamic interaction of social determinants of health (SDOH) and chronic disease risk factors in these underserved communities. Considering the complex nature of these, this study employed the system dynamics modeling technique to explore the root causes of the interacting factors and their impact on the minority populations and utilize machine learning to create predictive analyses of the simulated outputs of this model for predictive analysis. The system dynamics model projected the trends of the dynamic impact of these factors. Moreover, in the machine learning process, the Random Forest model outperformed the decision tree, gradient boosting models with the lowest evaluating metrics scores with MAE of 39.59 and MSE of 18091.15, R-squared 0.999997, Median AE of 14.83, and 0.00001 offering precision and predictive prowess. These integrated approaches provide a comprehensive view of chronic disease risk factors and SDOH for preparing for possible interventions and policy considerations in these communities.

Contents

1 Introduction

1.1 Overview

Chronic diseases are responsible for the greater part of worldwide mortalities and a significant portion of expenditure on healthcare [1]. Minority populations in Western nations are disproportionately affected by these diseases. The burden is regularly motivated by a complex hereditary, ecological, and behavioral interconnection [2]. However, socioeconomic level, education, neighborhood, physical environment, occupation, and social support networks significantly influence chronic disease risk [3]. The recognition emphasizes the importance of a holistic strategy for medical care that considers not just personal conduct and healthcare interventions but also the broader community aspects that affect health consequences [4].

This research explores the relationship between social determinants of health (SDOH) and the risk factors for chronic diseases among marginalized communities. The goal is to contribute to the increasing literature emphasizing how socioeconomic factors can affect health inequalities. Despite multiple investigations examining social determinants of health and chronic diseases separately, a knowledge deficit exists concerning the combined impact of these determinants on chronic disease susceptibility among these communities.

1.2 Research Motivation

The pressing challenges of the influence of SDOH on chronic disease incidence in minority populations motivate this research to develop an innovative interdisciplinary approach by integrating system dynamics modeling and machine learning to address these issues. System dynamics provides a practical framework for exploring causal relationships within the variables, while machine learning provides predictive analytics for actionable insights. By integrating these methodologies, the research aspires to uncover the root causes of these factors, develop predictive models, and inform policy and practice.

1.3 Research Questions

The following research questions provided the basis for the research:

1. Why is it essential to address how the interactions of SDOH and chronic disease risk factors influence the health outcomes in minority populations?

2. What underlying causes drive the health disparities in the incidence and prevalence of chronic diseases within minority populations?

3. How can the challenges of chronic diseases and health outcomes be addressed by integrating system dynamics modeling with machine learning to create a comprehensive framework to analyze these challenges?

4. What validation and calibration procedures assure the integrated model's robustness?

5. How accurately can the proposed model predict future trends in chronic disease prevalence within minority populations?

6. What factors predict chronic disease outcomes most accurately, and how are they evaluated in the model?

7. How will the research benefit the policymakers, healthcare system, chronic disease medical experts, and the population concerned?

1.4 Research Objectives

The research aims to utilize a system dynamics approach to model and simulate the interactions of SDOH and chronic disease risk factors and their impacts on health outcomes and disparities in minority populations in the US. The model will focus on five priority areas: *Economic Stability*, *Neighborhood and Physical Envinronment*, *Education*, *Community and Social Contexts*, and *Healthcare System*. This process will enable the assessment of the effectiveness of various healthcare interventions and the financial viability of SDOH-related initiatives, policies, and practices. Furthermore, the study aims to develop machine learning algorithms to analyze the numerical output data of the system dynamics model and find patterns illustrating how SDOH influences health outcomes in these underserved communities.

In addition to identifying specific policy options for the social determinants of health that have an impact on a range of chosen health outcomes, such as mortality, chronic diseases, disability, and unhealthy behavior, the model focuses on illuminating the causal pathways between significant population health risk factors and health outcomes.

1.5 Research Hypothesis

The hypotheses for this research are as follows:

- Minority chronic disease prevalence and management depend on SDOH.

- A system dynamics technique integrated with machine learning will provide a more comprehensive view of the interrelated processes that cause chronic diseases and minority healthcare inequalities than a linear approach.

- The research can improve healthcare equality for minority groups by modeling SDOH factors and chronic disease outcomes.

- Addressing SDOH factors and illuminating causal pathways between primary population health risk factors and health outcomes can significantly improve healthcare interventions in minority communities to reduce chronic disease incidence and progression.

- Understanding the interplay between SDOH and chronic disease risk factors can help minority chronic disease patients.

1.6 Research Contributions

This research makes several contributions to public health, machine learning, and the interdisciplinary study of chronic diseases within minority and underserved communities. The novel integration of system dynamics modeling with machine learning techniques explicitly applied to the analysis of Social Determinants of Health (SDOH) and chronic disease risk factors leads to the following substantive contributions: *Innovative methodological framework, enhanced understanding of SDOH, contribution to social equity in health, actionable policy insights, and interdisciplinary collaboration.*

1.7 Organization

This research paper is systematically organized into sections to provide a coherent and thorough exploration of the impact of SDOH and chronic disease risk factors in minority populations by employing system dynamics modeling and machine learning techniques. This section presents the rest of the research structure: The *literature review*, which provides the problem statement and the background information of the research, the *materials and methodology*; provides the techniques, software, and procedures employed in the research, *results, and discussions* provides the results of both the system dynamics modeling and the machine learning algorithms, the *conclusion* provides the concluding part of the research, and the *future work* provides the future direction of the research.

2 Literature Review

This section includes the problem statement, research background, and the related research performed by other researchers. It provides the background for the study and highlights the significance of the influence of social determinants of health in minority-populated regions.

2.1 Problem Statement

Chronic disease prevalence constitutes an alarming global health crisis, disproportionately impacting the minority population and underserved communities. These health crises and disparities in health outcomes in these communities are not merely due to biological factors and consequences but a result of the complex dynamic nature of SDOH factors. The adverse health outcomes result from the complex interplay between the SDOH and chronic disease risk factors observed in these populations. Despite significant efforts in addressing SDOH, current literature still needs to improve in identifying and capturing these complex interactions influencing the incidence and prevalence of these diseases in minority-populated regions. These challenges lie in the need for an integrated and multidimensional framework that can holistically address complex systems, such as the influences of these interacting factors. This gap hinders the implementation of focused interventions, informed policies, and individualized healthcare strategies that reduce chronic disease burden in underserved areas. As a result, this research seeks to address this critical problem by developing and validating an integrated approach that combines a system dynamic model and machine learning. This research aims to uncover the root causes of these interacting factors, predict their trends, and make practical recommendations to improve minority populations' adequate healthcare access and equity.

2.2 Background

Chronic diseases have considerable disparities in incidence and prevalence rates [5]. These disparities have existed for decades, but their influence has intensified across several disease categories [6]. The leading cause of death and disability in the United States is prolonged illness caused by chronic diseases. These diseases are a factor in the daily activities of 90 million people and are directly responsible for 7 out of 10 deaths [7]. To curtail health disparities, we must take a multidisciplinary approach when treating people with the most frequent chronic diseases. In the United States, chronic disease treatment accounts for 75% of total healthcare costs [8].

In 2007, the United States of America spent $2.2 trillion on medical care, which was more than any other country in the world had spent on health care at the time. Health disparities still exist regardless of continuous spending. In addition, a gap in understanding between the various demographic groups also persists [9]. Low-income Americans, as well as racial minorities, have a much higher incidence of disease. The variety of treatment options and access to therapy is severely restricted in these groups. As the unemployment rate increases, the gap between the rich and poor is already widening [10]. This difference will continue to rise at a significantly faster rate. As part of the effort to improve the healthcare system, it is imperative to reduce costs to save money on medical care. This implies safeguarding the patient's ability to select their healthcare practitioner, healthcare facilities, insurance coverage, and engaging in preventative care activities. Also, it ensures that everyone has access to high-quality health care that can be paid for by every American [11].

Furthermore, people of color and ethnic minorities, and underserved communities have a disproportionately higher incidence of major chronic diseases than people of other races and ethnicities. Diabetes, cancer, chronic kidney disease (CKD), heart disease, stroke, and HIV/AIDS are just a few of the diseases that fall into this category. Compared to the general population, where 39% of people have a chronic condition, one of the most significant disparities is found in the African American community, where 48% have a chronic disease [12].

2.2.1 Racial and Ethnic Diversity in the U.S. Population

As of 2021, 42% of all the people inside the US were from diverse ethnic backgrounds. This group included:

- 19% Hispanic
- 12% Black
- 6% Asian
- 1% of American Indians or Alaska Natives
- less than 1% were Native Hawaiians or Other Pacific Islanders
- 5% classified themselves as another racial category [13].

Additionally, this category includes individuals who recognize themselves as having multiple ancestries. The remaining 58% of individuals were of Caucasian descent—the share among those who identified as people of color has been growing over time. The largest growth is seen in people categorizing themselves as Hispanic or with Asian heritage. The racial diversity of society is anticipated to increase further. By 2050, people of color will represent over half of the residents. Changes to race/ethnicity data collection and reporting may affect social diversity indicators. Recent changes to these questions and how they were classified have increased the number of people identifying as other races or having various racial origins [13].

2.2.2 Understanding the Social Determinants of Health

Social determinants of health encompass the conditions in which individuals are born, grow, live, work, and age [14]. The phrase refers to health-related information, attitudes, beliefs, and behaviors, including smoking, nutrition, and alcohol intake. SDOH significantly influences the health of individuals and communities by shaping lifestyle decisions and behaviors that cumulatively lead to health or disease. Concurrently, these determinants are molded by public policies, implying that they are theoretically alterable [26]. Healthy People 2030 categorizes SDOH into five areas: *Economic Stability*, *Education Access and Quality*, *Health Care and Quality Access*, *Neighborhood and Built Environment*, and *Social and Community Context* [18].

1. **Economic Stability**

   Socioeconomic Status (SES) influences the United States' chronic disease, ethnicity, and race disparities. Income, wealth, employment status, and job type determine the availability of medical care, housing, and other chronic disease challenges. Recent research identified a persistent connection between increased income mobility and reduced chronic disease and mortality occurrences [17].

2. **Neighborhood and Physical Environment**

   The health and overall wellness of individuals significantly depend on where they live. Numerous Americans reside in neighborhoods exhibiting alarming crime rates, violence, and environmental pollution (specifically *air* and *water*). Along with numerous other detrimental factors jeopardizing their overall well-being and safety. Minorities and low-income people reside in unsafe environmental situations more often. Moreover, some individuals are subject to harmful workplace elements, such as passive smoking or high-decibel noises [18].

3. **Education**

Education improves economic and health resources. It makes it easy to analyze complex medical data, enhancing healthcare decision-making, income, and living standards. Further, studies show that higher education facilitates employment in safer, better-paying jobs. Financial security enables one to cater to one's health and consult professional medical experts [19].

4. **Community and Social Context**

   Within public health, determinants exist in various facets that shape one's well-being; one such crucial factor is the *social and community context*. This encompasses subjects like active participation in civic affairs or public initiatives to foster societal unity [20]. Regrettably, structural discrimination can arise whereby certain policies unfairly target specific groups, increasing stress levels while adversely affecting their mental health status. A pertinent example bringing this to light would be *New York City's infamous "stop and frisk"* policy [21], which undoubtedly amplified feelings of anxiety while also prompting higher vulnerability towards depressive conditions among those marginalized communities singled out by this practice [19].

5. **Healthcare System**

   Sound *health* requires *primary care*, including disease detection, treatment, and prevention [22]. However, insurance and provider constraints might hinder individual access to health care. Studies suggest that such health access reduces significant disease mortalities and increases diabetic medication utilization by uninsured patients [23]. Personal health management requires health literacy, which reduces frequent hospitalizations and deaths [24].

2.3 Related Research

Over the years, SDOH attention has been drawn to chronic disease risk factors, particularly in minority communities. Innovative methods are often needed to comprehend and address these complex interplays. This research examines how machine learning (ML) and system dynamics modeling may investigate and reduce SDOH's effects on minority chronic disease risk factors. Researchers and policymakers may better understand multidimensional relationships and devise targeted policies to minimize health inequalities by merging these computational tools with health data and social determinants indicators. Using ML and system dynamics modeling to explore SDOH and chronic diseases in minority communities has benefits, problems, and future objectives. This section of the research covers the literature.

2.3.1 The Crucial Role of Social Determinants in Health Outcomes

This section explores the literature on the significant impact of SDOH on the health outcomes of individuals. The following research studies explain the various influences of SDOH in different communities.

Bharmal et al. [26] delve into SDOH, emphasizing non-medical factors like social disadvantage, risk exposure, and social inequities significantly influencing health outcomes. This research differentiates three methodologies for studying these upstream SDOHs: The social disadvantage approach, which explores how various socio-economic elements and associated stress affect health The life course approach, which associates health with critical periods of risk exposure and considers the impact of social status on gene regulation The health equity approach investigates how socio-demographic factors and social capital can shape health. Their study further underscores several challenges in understanding SDOH, encompassing complex causal pathways, multiple intervening factors, research approaches, and funding limitations. Cockerham et al. [27] explore SDOH in chronic diseases, countering their previous characterization as secondary factors. Their research explores the health impacts of SDOH, including smoking and mortality. It also evaluates four key SDOH theories - the life course, fundamental cause, social capital, and health lifestyle theory - and analyzes the role of neighborhood disadvantage, social networks, and perceived discrimination in health disparities. Federico et al. [28] examine how housing, community safety, and healthcare access affect childhood asthma, a global chronic condition. Despite evidence that socioeconomic and environmental factors affect clinical treatment, integrating them is difficult. The best methods to reduce these variables are unknown. The paper also summarizes recent studies on these factors and proposes a thorough therapeutic strategy for detecting and treating them in asthmatic children. Raphael [29] critically evaluates the application of the SDOH concept within the Canadian policy landscape. Despite its recognition, the implementation is hindered by governmental approaches favoring market dynamics and welfare retrenchment and a reluctance among SDH researchers to identify policy implications. It identifies and analyzes various SDOH discourses to help advance the SDOH agenda in Canada and globally.

Singu et al.[30] examine susceptible demographic groups' SARS-CoV-2 pandemic impacts. Asthma, cardiovascular disease, hypertension, chronic renal sickness, obesity, and the elderly were covered. Socioeconomic status may alter the occurrence of some prior disorders and raise COVID-19 severity in these people. Medical SDOH is stressed in this review, emphasizing the examination of how these factors influence weaker people in emergencies. To guide government and ensure health equity. The study implies that focusing on basic causes can help address medical emergencies and protect everyone's health and safety, regardless of socioeconomic position.

2.3.2 SDOH and Health Outcomes in Minority populations

The section deeply explores the literature on the different impacts of SDOH in various minority communities.

The SDOH theory suggests that socioeconomic factors affect individual and community health. These factors significantly affect minority cancer research engagement. Nurse researchers must identify and overcome these barriers to increase research participation and improve minority cancer trial design and outcomes. In their study, Asare et al. [31] emphasize the need to include SDOH knowledge in cancer treatment research to promote inclusivity and efficacy.

Alcendor [32] compares COVID-19's effects on African Americans, Hispanics/Latinos, and non-Hispanic Whites in this research. The study underlines the virus's great transmissibility and toxicity, which causes major disease and death, especially in pre-existing patients. SARS-CoV-2 is a global concern, including in the U.S. Clinical factors that increase minority mortality are examined. The author also examines the structural, cultural, and sociological barriers ethnic minorities face in achieving health equity and proposes legislative solutions.

COVID-19 affected minority populations' SDOH and Cardiovascular Disease (CVD) incidence, morbidity, and mortality. Russo et al. [33] discuss how the pandemic affects SDOH and cardiovascular health. The epidemic increased unemployment, food insecurity, loneliness, processed food consumption, healthcare, income, physical activity, and hypertension control. Minorities and the underprivileged are unprotected. Culturally appropriate therapy can improve immigrants' and marginalized groups' health.

Fraiman and Wojcik [34] discuss genetic testing for rare genetic disorders in children. These disorders have several genetic variations. Genetic testing confirms the molecular diagnosis and terminates patients' and families' diagnostic journeys. The study argues that diagnostic testing for uncommon genetic illnesses in children is less investigated than for adult cancer risk. They also show pediatric genetic diagnosis inequalities and the need for prospective investigations to ensure precision medicine benefits all.

2.3.3 SDOH and Chronic Disease Risk Factors

The interactions between SDOH and chronic disease risk factors are complicated. These interactions can increase the possibility of chronic diseases or long-term ailments in any region, especially the minority populations that usually experience adverse situations because of their lifestyle, environment, and poor access to proper healthcare. To understand the impacts of the interactions, this research section explains the effects of the SDOH and chronic disease risk factors interactions on health outcomes.

Public health studies have connected chronic disease development to an individual's SDOH, including unhealthy behaviors, unfavorable environments, socioeconomic status (SES), income, education, and other lifestyle risk factors [35]. They have also linked this interaction to the effectiveness of clinical and community preventative services. These factors frequently act as risk factors for chronic diseases, adding complexity to the individual's health outcomes [36]. Healthcare institutions recognize the relevance of social and behavioral factors in patient health, making it possible to address these health challenges [37]. The following related works further elaborate on the interplay of SDOH with chronic disease risk factors.

Brakefield et al. [38] explore how SDOH influences COVID-19 infectivity, hospitalization, and mortality in marginalized U.S. populations. The evaluation disproportionately affects racial/ethnic minority communities, the elderly, and displaced/homeless persons. Race/ethnicity, income, housing, healthcare access, occupation, transportation, education, air quality, and food insecurity affect COVID-19 outcomes. SDOH indicators and health data are recommended for minority pandemic response and recovery.

Emeny et al. [39] explore how SDOH impacts health disparities. SDOH encompasses social, political, and economic factors greatly influencing health outcomes. The paper emphasizes the need to integrate SDOH into research, proposing a precision health framework to understand their role in chronic diseases. It discusses measuring SDOH and its associations with health outcomes. Considering SDOH in study design and analysis can inform prevention and treatment strategies for equitable care and population health.

2.3.4 Prevalence and Impact of Chronic Diseases in Minority Populations

Chronic disease, self-care, and family and social ties research on older ethnic minority men and women. In their study, Gallant et al. [40] examine sociocultural factors in self-care to improve health. They then evaluate the literature on chronic illness self-care habits among older African-Americans, Latinos, Asian-Americans, and American Indians in the U.S., examining sociological aspects such as residential, cultural, and socioeconomic patterns, family dynamics, and other social ties. Studying self-care addresses these societal inclinations.

2.3.5 Application of Machine Learning in Chronic Disease Prediction

This research section explains the literature on the application of machine learning in healthcare, particularly the prediction and management of chronic diseases. Moreover, the following research studies employ different machine learning algorithms to predict the health outcomes of individuals considering some medical factors and procedures.

Battineni et al. [41] review machine learning (ML) in Chronic disease (CDs) predictive models. CDs require lifelong therapy, increasing worldwide healthcare expenses. To discover cutting-edge ML-based CD diagnosis methods, the scientists evaluated 453 PubMed and CINAHL papers from 2015 to 2019. After rigorous study, 22 articles were chosen to demonstrate various modeling methodologies and their strengths and weaknesses in identifying distinct diseases. The results show that each real-time clinical practice method has merits and cons. The most popular CD classification and diagnosis models were Support Vector Machine (SVM), Linear Regression (LR), and clustering algorithms. In this research, Nusinovici et al. [42] compared logistic regression to machine learning (ML) in predicting hypertension (HTN), diabetes (DM), cardiovascular diseases (CVDs), and chronic kidney disease (CKD). The study screened Asian adults. They also evaluated other machine learning algorithms against Logistic regression. Logistic regression predicted CKD and DM well, while neural networks and SVM predicted CVD and HTN. The Logistic regression matched ML. Logistic regression predicted low-incidence major chronic diseases.

Chien Hsiang et al. [43] present an integrated detection system for chronic cardiac disease prediction that measures blood pressure, glucose, lipids, and heart rate. Support Vector Machine, Random Forests, k-Nearest Neighbors, XGBoost, and LightGBM were used to classify the data. Random Forests and k-Nearest Neighbors outperformed XGBoost and LightGBM with 88.52% prediction accuracy. The suggested technology will monitor users' physical health to help manage chronic heart conditions.

Convolutional neural network (CNN) for automatic feature extraction and disease prediction and K-nearest neighbor (KNN) for distance calculation identifies the precise match in the data set and the final disease prediction results to provide an overall patient symptom-based disease prognosis.

This broad disease prediction considers symptoms, lifestyle, and medical visits. Naïve Bayes, decision tree, and logistic regression are compared to the suggested approach [44].

Khalid et al. [45] evaluated algorithm accuracy and work analysis of machine learning classification. Their top four algorithms and hybrid model predicted UCI CKD. On the same dataset, gradient boosting (GB) has 99% accuracy, random forest 98%, decision tree classifier 96%, and their suggested hybrid model gave 100% accuracy. Naïve Bayes, decision tree, K-nearest neighbor, random forest, support vector machine, LDA, GB, and neural network predict CKD. GB, Gaussian Naïve Bayes, decision tree, and random forest accuracy are compared.

Chen et al. [46] explore the optimizing machine learning predicts chronic disease outbreaks in disease-prone areas. Improved prediction models are tested on 2013–2015 central China hospital data. The study proposed Convolutional neural network (CNN)-based multimodal sickness risk prediction using hospital structured and unstructured data. Their method has 94.8% prediction accuracy and converges faster than the CNN-based unimodal disease risk prediction algorithm.

2.3.6 Employing System Dynamics Modeling in Healthcare

The system dynamics model is a powerful and exploratory concept that addresses complex and dynamic systems and scenarios. Its applications have been employed in several fields, such as healthcare, manufacturing, education, and management. Hence, this research section explores different studies that applied the system dynamics model to address the complexity of various factors and events in healthcare.

System dynamics modeling is employed by Yinusa et al. [47] in their research to address and model CKD incidence, prevalence, and health inequities that influence its management. Health interventions for minorities with CKD inspired this method. The model's graphical results reveal that dynamic factors affecting CKD incidence and prevalence are related. Hence, healthcare disparities complicate this disease's care and management interventions.

Loyo et al. [48] encourage preventive care and manage limited resources using a System Dynamics Model of Cardiovascular Disease Risks (SD model). The SD model considers disease prevalence, risk factors, the local environment, and population health. The model examines upstream cardiovascular event prevention. Interactive simulations set stakeholder policy and strategy. The study simulates interventions. SD influenced Austin's Chronic Disease Prevention Coalition's intervention goals.

Ciplak and Barton [49] discuss Istanbul's hospital waste management using system dynamics modeling. It emphasizes safe disposal and segregation to protect healthcare workers, patients, and the environment. Healthcare waste makes up 65% of municipal waste. However, appropriate segregation may cut healthcare waste by thousands of tons annually. The research stresses better waste management to minimize costs and waste amounts and supports autoclaving and other cost-effective healthcare waste treatment methods.

Furthermore, SDOH and chronic diseases in minority populations have been extensively researched, as seen in the literature review (section 2). However, gaps still need to be filled in the fusion of these research areas with *machine learning* and *system dynamics modeling* approaches. Moreover, some data scientists and healthcare experts have separately employed machine learning to predict chronic diseases and system dynamics modeling to examine health and healthcare disparities. However, these approaches can be combined better to deeply understand the interplays between SDOH and chronic diseases. This combined approach is adopted in this study to address the interactions of SDOH and chronic disease risk factors and their impacts in the minority-populated regions in the US.

3 Materials and Methodology

This section involves the research materials and methods. Thus, the research aimed to create a unique methodological framework integrating *system dynamics modeling* with *machine learning* approaches. This integration sheds light on how SDOH influences the incidence and prevalence of chronic diseases among minority communities in the United States.

Moreover, system dynamics modeling is employed as an *exploratory modeling tool* to understand the complicated interactions and feedback loops between SDOH factors and health outcomes, particularly in these minority populations. Regarding the impact of SDOH on general health outcomes, it offers graphic representations that show non-linear correlations and temporal time delays. Further, it enables us to simulate long-term influences resulting from various determinants of health and visualize how alterations within one factor can propagate throughout an entire system over time.

Generally, machine learning utilizes data patterns to predict and decide, making it a unique *predictive modeling tool*. To discover patterns and correlations, an algorithm is trained using training data. These insights are used to test data to predict. Thus, for the research, the simulation output data from system dynamics modeling is employed as the data input for the machine algorithms. Moreover, some system dynamic modeling parameters serve as features, while one or two serve as targets for the machine learning predictive process.

3.1 Research Process Design

Figure 1 shows the process design of the research. This process design explains the concepts, steps, and procedures employed in the research. The interactions of SDOH and chronic disease risk factors are complicated and non-linear. This situation makes the influence of these factors adverse in every region in the world, including the United States of America, and especially in the minority-populated regions. With this, a system dynamic model is appropriate in this scenario. The system dynamic technique considers the complex interacting nature of the interacting factors and variables in these areas. After studying and observing the system, a research idea was conceptualized, then there was a need to address the situation by employing a system dynamics model using Vensim software for exploratory analysis. Afterward, a Python library, PySD, was employed to translate the Vensim model and simulation into a Pythonic environment. Moreover, the system dynamics modeling output was saved and fed into the machine learning process using different algorithms for predictive analysis.

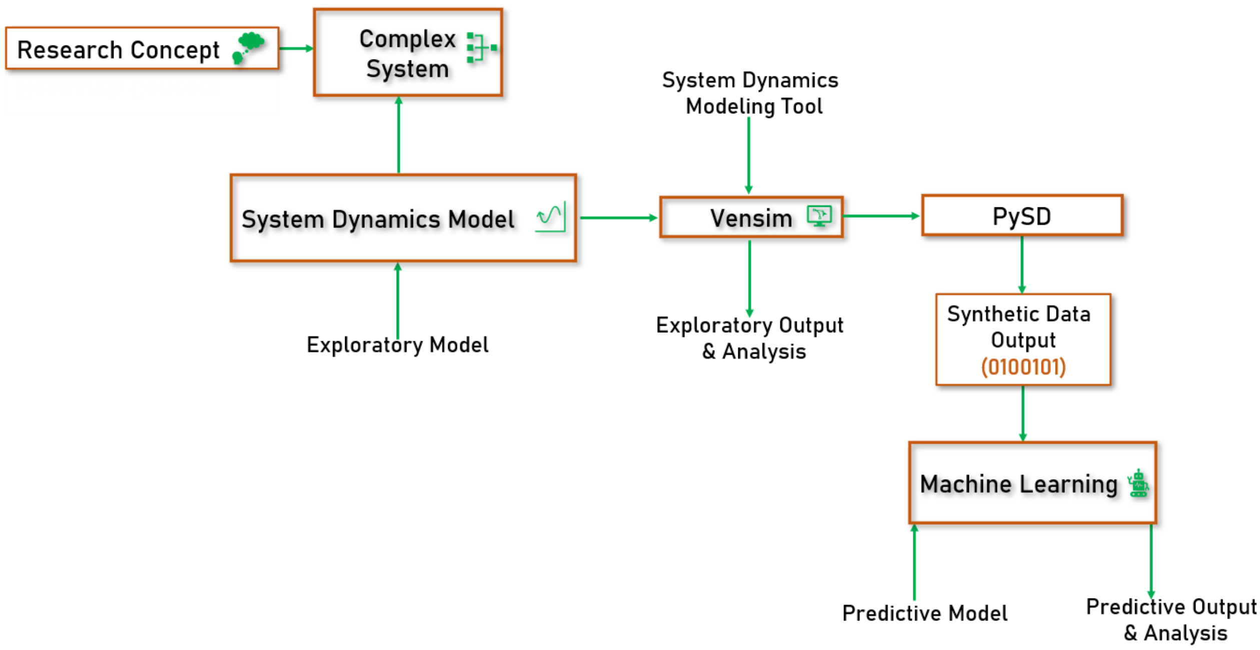

Figure 1: The Research Process Design.

3.2 Materials

This section presents the software and tools employed in the research.

3.2.1 Software and Tools

This research employed carefully selected software to facilitate the execution of the integrative methodological framework. The *Vensim software* [51] developed and simulated the complex dynamic model and diagrams for the SD modeling procedure. Due to its versatility and compatibility with numerous data types and formats, the PySD library was employed as a *Pythonic environment* for the model to advance further and modify the SD model for data cleaning, preprocessing, and analysis.

3.3 Python System Dynamics (PySD) Library

PySD is a *Python library* that converts models from *Stella*® [50] or *Vensim*® [51] (*Commercial system dynamics modeling tools*) into Python, first released in 2014 by James Houghton. It includes functional elements for parsing, implementing, building, and solving models. PySD also enables users to import and modify model inputs, break models into submodules, isolate parts of a model for individual running, and store intermediate simulation results [52]. An essential upgrade in version 3.0.0 was the separation of the parsing and building processes, which allows the output models to be written in any programming language using an Abstract Model Representation. Despite being well-established, PySD continues to evolve, introducing new features regularly.

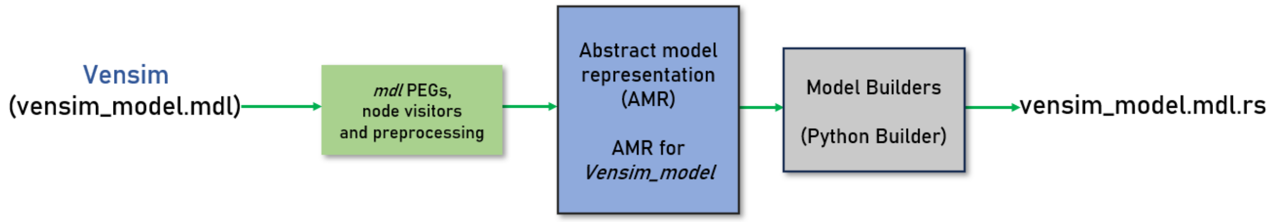

Figure 2: Illustration of the enhanced parsing-building logic introduced in PySD v3.0.0. "PEG" denotes Parsing Expression Grammar.

3.4 Research Methodology

This section explores the methodologies and techniques employed in the research.

3.4.1 System Dynamics Modeling

System Dynamics (SD) is a computational technique employed to model and simulate the dynamics and challenges of complex systems over time. Professor Jay W. Forrester [53] of the Massachusetts Institute of Technology (MIT) established the foundation of this technique in the 1950s. Several public and private industries have employed SD for decision-making and policy design. SD comprises *stock*, *flows*, *variables*, *parameters*, and other *auxiliaries* as its building *components*. The stocks are the accumulative quantities or states that take on a specific value each time. The flows are the rates at which quantities change over time. They are the mathematical derivatives in the model, dictating how the stocks, or cumulative quantities, increase or decrease. The flows are usually in two forms: inflow (*increases the stock over time*) and outflow (*decreases the stock over time*). Moreover, variables represent intermediate calculations or values that are not stocks or flows but are still crucial for the model's functioning, and parameters define the simulation's external circumstances. These elements play significant roles in the SD modeling process by enabling the formation of *feedback loops*, where the condition of the stock variables feeds back to affect the model's flow of information [52].

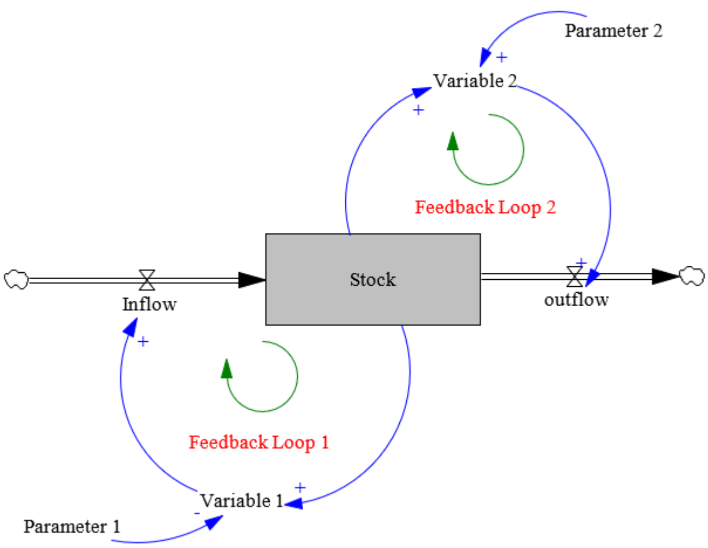

Figure 3: Visualization of System Dynamics Model Components in Domain-Specific Modeling Environments

Feedback Loops in SD

They represent different types of feedback that can occur in a system; these include:

- Reinforcing Loop (*Positive Feedback Loop* (R)): This type of loop amplifies system behavior. The effect of a change in the system is an increase in the same direction. For example, *poor physical activity* can lead to *obesity*, which in turn can result in more *health issues*. This is a reinforcing loop because the lack of exercise leads to obesity, especially with a poor diet, which further increases health complications. Reinforcing loops can lead to exponential growth or decline. Figure 4 shows the reinforcing loop diagram for the previous example. The arrows represent the direction of the loop, and the *positive sign* (+) indicates an increment in the forward direction. Analogically, the reinforcing loop is always in a clockwise direction format.

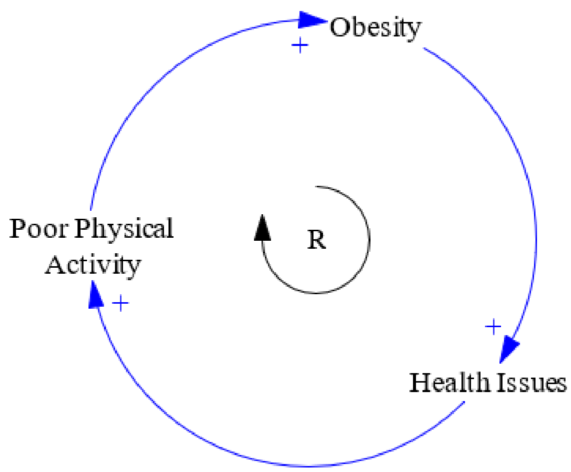

Figure 4: The Reinforcement Loop Sample Visualization in SD

- Balancing Loop (*Negative Feedback Loop* (B)): This type of loop seeks to maintain a desired state or balance in the system. It works to counteract any deviation from the goal. For example, increased health awareness can increase efforts to improve air and water quality, decreasing poor air and water quality and reducing chronic disease. This is a balancing loop because improving air and water quality can reduce chronic diseases in many ways. Figure 5 shows the balancing loop diagram for the previous example. The arrows represent the direction of the loop, and the *negative sign* (-) indicates a decrement in the forward direction. Analogically, the balancing loop is always in an anticlockwise direction format.

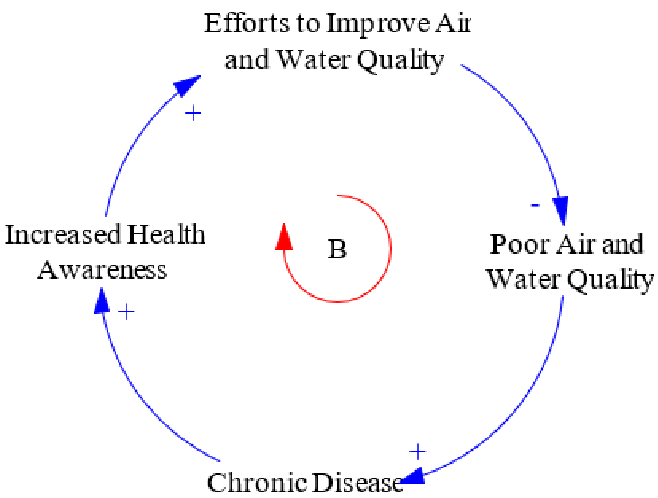

Figure 5: The Balance Loop Sample Visualization in SD

3.4.2 The Application of System Dynamics Modeling in the Research

The system dynamics model is applied in this research to address the dynamicity and complexity of the interactions of SDOH and the chronic disease risk factors and the impacts that usually arise from these interactions in minority populations leading to various adverse health outcomes. The effects of these interactions among these factors affect the entire population in the US, but it disproportionately affects the minority population the more. The application of system dynamics modeling comes in two forms: the *causal loop model or diagram* and the *stock and flow model* explained in the previous section.

3.4.3 Data Sources and Variable Selection

The research modeling variables (*SDOH and chronic disease risk factors*) are carefully selected for the causal loop diagrams and the stock and flow model. This process is based on reviewing empirical studies, comprehensive literature, and expert consultations. Also, observations within the targeted population in understanding the relevant variables for the specific communities

3.4.4 Model Variable Relationships

Identifying, arranging, and establishing relationships between the variables in the model is essential to deeply explore the root causes of the challenges with SDOH and chronic disease risk factors in minority populations. The relationships between these variables in the causal loop diagrams and the stock and flow model were established in three ways: these are *Theoretical Alignments*, *Sensitivity Analysis*, and *Comparison with Existing models.*

- **Theoretical Alignments**: This ensures that the variables are related based on their alignment with existing theories, epidemiology, social science, models, and principles within public health.
- **Sensitivity Analysis**: This establishes and confirms the relationships' magnitude by assessing how changes in one variable affect others.
- **Comparison with Existing Models**: In this case, previously established frameworks and models were consulted to ensure that the credibility and consistency of the model align with the established relationships of the variables.

3.4.5 Model Simulation and Validation Analysis

In system dynamics modeling, validation is essential in ensuring the developed model represents the real-world system it intends to simulate. There are various validation techniques in system dynamics, such as structural validation, behavioral validation, comparative validation, operational validation, policy validation, documentation, and transparency, and each has its significance [65]. This research employed *Structural validation, Behavioral validation, and Operational validation.*

- **Structural validation**: With this technique, the research employed the *Face Validity* approach by involving experts examining the model's structure to observe whether it aligns with their understanding of the model and makes sense with the actual word it simulates. Still under structural Validation, *Dimensional Consistency* was employed to ensure that the units in the model match the stocks and flows' units correctly to avoid unit errors in the Vensim software. Also, parameter values were checked to depict the real-world system's essence accurately.
- **Behavioral validation**: With this method, the research employed the *Sensitivity Analysis approach* to assess how changes (variations) in

parameters affect the model's behavior. This process was undertaken to ensure the robustness of the model. This validation technique also employed an *Extreme Conditions Testing approach* by subjecting the model to extreme conditions to observe and examine how it behaves and identify any unrealistic outcomes.

- **Operational validation**: With this approach, the research employed *Boundary Adequacy Testing* to ensure that the model's significant variables and relationships are within defined boundaries.

1. **The Causal Loop Diagram of the Interplay between SDOH and Chronic Disease Risk Factors**

   As explained in the previous section, SDOH is categorized into *Economic Stability*, *Neighboorhood and Physical Enviropnment*, *Education*, *Commuinty and Social Context*, *Healthcare System*, and *Health Behaviors*. Tables 1, 2, and 3 present the respective SDOH category. Table 4 presents some *chronic disease risk factors* in the following causal loop diagram.

Economic Stability	**Neighborhood and Physical Environment**
Poverty	Air and Water Quality
Unemployment	Residential Segregation
Health Expenditure	Access to Exercise Opportunities
Health Insurance	Housing Problems

Table 1: Social Determinants of Health Factors

Education	**Community and Social Context**
School Completion	Social Associations
Inequalities in Education	Social Isolation
Health Knowledge	School Segregation
School Enrollment	Health Disparities

Table 2: Social Determinants of Health Factors

Healthcare System	**Health Behaviors**
Chronic Diseases	Alcohol Consumption
Demand for Healthcare	Drug Abuse
Access to Care	Obesity
Provider Linguistic and Cultural Competency	Physical Inactivity
Awareness and Efforts to Improve Cultural Competency	-
Communication with Healthcare Providers	-
Health Disparities	-

Table 3: Social Determinants of Health Factors

Chronic Disease Risk Factors
High Blood Pressure/Cholesterol
Increased Rates of Low Birth Weight
Stress

Table 4: Chronic Disease Risk Factors

2. **SDOH with Chronic Disease Risk Factors Interplay**

   The causal loop diagram for the interactions of the SDOH and chronic disease risk factors in the minority populations was developed by selecting the relevant variables associated with the model as previously mentioned in sections 3.4.3 and 3.4.4.

   Figure 6 shows the causal loop diagram of the interaction of SDOH and chronic disease risk factors. Generally, some chronic disease risk factors still serve as social determents of health factors; as a result, there are complex interactions of these factors. Moreover, employing a linear approach to address the impact of SDOH and chronic disease risk factors in minority population regions is inappropriate because the approach does not consider the reinforcing and balancing loops that would have occurred among the factors. Since the situation is complex and dynamic, a technique such as the system dynamics model is better than the linear approach. This research employs a causal loop diagram to address the interplays and the causal effects among these factors. Following that, the complex interacting factors are modeled and simulated in the minority populations.

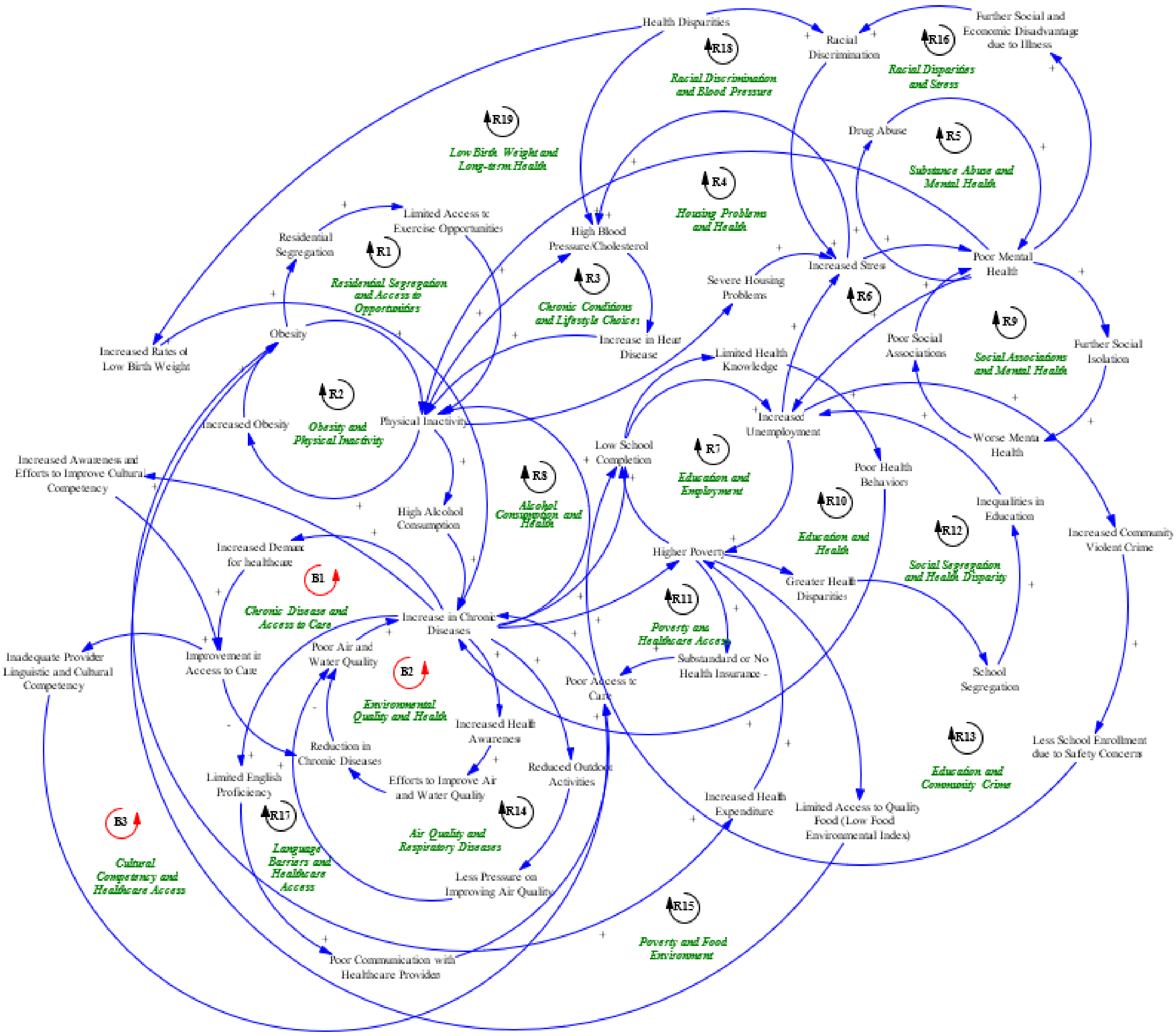

Figure 6: SDOH and Chronic Disease Risk Factors Interactions.

- **Residential Segregation and Access to Opportunities (R1)**: This loop suggests that *Residential Segregation* leads to *Limited Access to Exercise Opportunities.* Afterward, leading to physical inactivity, then leading to obesity.
- **Obesity and Physical Inactivity (R2)**: In this loop, *Physical Inactivity* leads to *Obesity*, which in turn leads to I*ncreased Obesity*, then back to *physical inactivity.*
- **Chronic Conditions and Lifestyle Choices (R3)**: This rein-

forcement loop suggests that *High blood pressure and High Cholesterol* leads to an increase in *Heart disease*, which leads to *Physical Inactivity* and then back to *High Blood Pressure and High Cholesterol*.

- **Housing Problems and Health (R4)**: This loop illustrates the that *Severe Housing Problems* lead to *Increased Stress* and then to *Poor Mental Health*, then to *Physical Inactivity*, and *Severe Housing Problems*.
- **Substance Abuse and Mental Health (R5)**: *Drug Abuse* presumably leads to *Poor Mental Health*, then back to *Drug Abuse* creating a reinforcing loop.
- **Mental Health Issues and Unemployment (R6)**: In this loop, *Increased Stress* results in *Poor Mental Health*, then *Increased Unemployment*, and then back to *Increased Stress*.
- **Education and Employment (R7)**: This loop suggests that *Low School Completion* leads to *Increased unemployment* and *Higher poverty*, then back to *Low School Completion*. This could further create a reinforcing loop where the situation continues to worsen.
- **Alcohol Consumption and Health (R8)**: *High Alcohol Consumption* presumably leads to an *Increased in Chronic Diseases*, then to *Physical Inactivity*, then back to *High Alcohol Consumption*, creating a reinforcing loop.
- **Social Associations and Mental Health (R9)**: In this loop, the *Poor Social Associations* lead to *Poor Mental Health*, which further leads to *Social Isolation* and *Worse Mental Health*, then back to *Poor Social Associations*.
- **Education and Health (R10)**: *Limited Health Knowledge* leads to *Poor Health Behaviors*, leading to *Increased in Chronic Diseases*, which results in *Low School Completion*, then back to *Limited Health Knowledge* . This is a reinforcing loop, suggesting that a lack of education can lead to behaviors that negatively impact health.
- **Poverty and Healthcare Access (R11)**: In this loop, *High Poverty* leads to *Substandard or No Health Insurance*, which leads to *Poor Access to Care*, then resulting in *Increased in Chronic Diseases*, then back to *High Poverty*.
- **Social Segregation and Health Disparity (R12)**: In this loop, *School Segregation* leads to *Inequalities in Education*, which

in turn leads to *Increased Unemployment*, then to *Higher Poverty*, then to *Greater Health Disparities*, resulting back to *School Segregation.* This reinforcing loop suggests that social segregation can lead to health disparities and unemployment.

- **Education and Community Crime (R13)**: *Increased Community Violent Crime* leads to *Less School Enrollment due to Safety Concerns*, which results in *Low School Completion*, which leads to *Increased Unemployment*, then back to *Increased Community Violent Crime.* This reinforcing loop indicates that increased crime can lead to decreased school enrollment.
- **Air Quality and Respiratory Diseases (R14)**: This loop suggests that *Reduced Outdoor Activities* lead to *Less pressure on Improving Air Quality*, leading to *Poor Air and Water Quality*, then leads to *Increased in Chronic Diseases*, then back to *Reduced Outdoor Activities.*
- **Poverty and Food Environment (R15)**: *Limited Access to Quality Food (Low Food Environmental Index)* leads to *Obesity*, leads to *Increased Health Expenditure*, leads to *Higher Poverty*, then back to *Limited Access to Quality Food (Low Food Environmental Index).* This is a reinforcing loop, suggesting that poverty can lead to a poor food environment, which in turn can lead to increased health costs.
- **Racial Disparities and Stress (R16)**: *Racial Discrimination* leads to *Increased Stress*, which in turn leads to *Poor Mental Health*, then resulting in *Further Social and Economic Disadvantage due to Illness*, then back to *Racial Discrimination.* This reinforcing loop suggests that racial discrimination can increase stress and economic disadvantage.
- **Language Barriers and Healthcare Access (R17)**: *Limited English Proficiency* leads to *Poor Communication with Healthcare Providers*, leading to *Poor Access to Care* and *Increased in Chronic Diseases.* This reinforcing loop indicates that language barriers can lead to poor healthcare access.
- **Racial Discrimination and Blood Pressure (R18)**: The loop suggests that *Racial Discrimination* can lead to *Health Disparities* and *Increased Stress*, which can specifically increase *High Blood and Cholesterol rates.*
- **Low Birth Weight and Long-term Health (R19)**: *Health Disparities* can lead to *Increased Rates of Low Birth Weight*, which

in turn leads to long-term health issues (*Increased in Chronic Diseases*). This reinforcing loop suggests that low birth weight can have long-term health implications.

- **Chronic Disease and Access to Care (B1)**: In this loop, an *Increase in Chronic Diseases* leads to *Increased Healthcare Demand and Improved Access to Care*, which leads to a *Reduction in Chronic Diseases.*
- **Environmental Quality and Health (B2)**: *Poor Air and Water Quality* lead to *Increased in Chronic Diseases.* However, *Increased Health Awareness* leads to *Efforts to Improve Air and Water Quality*, which in turn leads to a *Reduction in Chronic Diseases.* This is a balancing loop aiming to maintain a certain level of health.
- **Cultural Competency and Healthcare Access (B3)**: This balancing loop indicates that *Inadequate Provider Linguistic and Cultural Competency* leads to *Poor Access to Care* which in turn leads to *Increase in Chronic Diseases*, which results in an *Increase in Demand for Healthcare*, then leads to *Improvement in Access to Care*, finally back to *Inadequate Provider Linguistic and Cultural Competency* .

3. **The Causal Diagram of the Interaction of the Chronic Disease Risk Factors**

   Table 5 shows the variables involved in the model's causal loop diagram of the chronic disease risk factors in minority populations.

Chronic Disease Risk Factors	
High Alcohol Consumption	Physical Inactivity
Obesity	Higher Risk of Chronic Diseases
Poor Prevalence of Smoking	Physical Distress
Poor Food Environmental Index	Poor Physical Health
Poor Diet and Chronic Diseases	High-Stress Levels
Poor Mental Health	Poor Access to Exercise Opportunities
Further Limitation in Exercise	Limited Exercise and Obesity
High Blood Pressure	Increase in Heart Diseases
High Rates of Substance Abuse	Increased Use of Substances
High Hormone Levels	Genetic Susceptibility to Chronic Diseases
High Blood Cholesterol	Aging

Table 5: Chronic Disease Risk Factors variables in the Model

4. **Chronic Risk Factors Interactions in the Minority Populations**:

   Chronic disease risk factors enhance the possibility of developing chronic disease. They initiate and further contribute to the incidence and prevalence of these chronic diseases, thereby negatively influencing individuals' health and quality of life. Generally, these risk factors are categorized into two groups: *non-modifiable chronic risk* and *modifiable chronic disease risk factors.*

   (a) The *non-modifiable risk factors* are the ones that cannot be changed or imagined, or altered to improve an individual's health, and they include:
      - Individual's Age
      - The gender type of individual
      - Individual's genetic composition
      - Individual's ethnicity or race

   (b) The *modifiable risk factors* are the ones that can be managed or changed to improve an individual's health, and they include:
      - Unhealthy diet
      - High tobacco intake
      - Physical inactivity
      - Obesity
      - High cholesterol
      - High blood pressure
      - High alcohol consumption

   The causal loop diagram for the interactions of the chronic disease risk factors in the minority populations was also developed by selecting the relevant variables associated with the model as previously mentioned in sections 3.4.3 and 3.4.4.

   Figure 7 shows the causal loop diagram of the model's interaction of chronic disease risk factors. This diagram includes the possible reinforcing loops that exist in these interplays. The interplays involve the modifiable and non-modifiable risk factors leading to chronic diseases in minority populations.

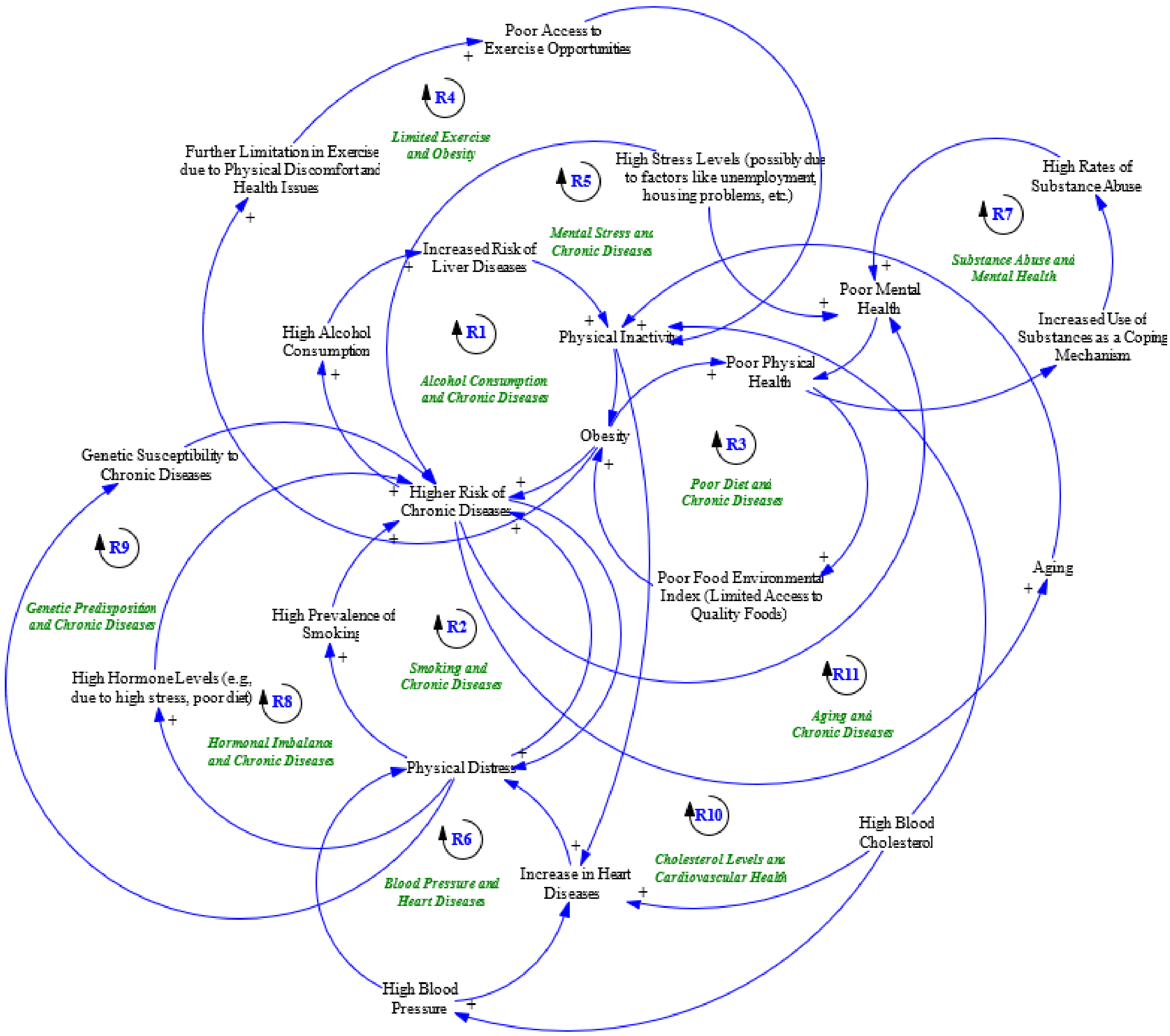

Figure 7: Chronic Disease Risk Factors Interactions.

- **Alcohol Consumption and Chronic Diseases (R1)**: This loop indicates that *High Alcohol Consumption* increases the *Risk of Liver Diseases*, which in turn increases *Obesity*, then leads to *Higher Risk of Chronic Diseases*, then, back to *High Alcohol Consumption*.
- **Smoking and Chronic Diseases (R2)**: This loop suggest that *High prevalence of smoking* leads to *Higher Risk of Chronic Diseases* which in turn results in *Physical Distress*, and then back to *High prevalence of smoking*.

- **Poor Diet and Chronic Diseases (R3)**: *Poor Food Environmental Index (Limited Access to Quality Foods)* leads to *Obesity* which leads to then back to *Poor Food Environmental Index (Limited Access to Quality Foods).*
- **Limited Exercise and Obesity (R4)**: In this loop, *Poor Access to Exercise Opportunities* which leads to *Physical Inactivity*, then leads to *Obesity*, which leads to *Further Limitation in Exercise due to Physical Discomfort and Health Issues* then this leads back to *Poor Access to Exercise Opportunities.*
- **Mental Stress and Chronic Diseases (R5)**: This loop suggests that *High-Stress Levels* (possibly due to factors like unemployment, housing problems, etc.) lead to *Poor Mental Health*, which increases *Poor Physical Health*; this leads to *Obesity*, then leads to a *Higher Risk of Chronic Diseases*, which then leads back to *High-Stress Levels.*
- **Blood Pressure and Heart Diseases (R6)**: In this loop, *High Blood Pressure* increases the *Risk of Heart Diseases* and then leads to an increase *Physical Distress.*
- **Substance Abuse and Mental Health (R7)**: In this loop, *High Rates of Substance Abuse* leads to *Poor Mental Health*, then which leads to *Poor Physical Health*, which in turn leads to *Poor Physical Health*, resulting in *Increased use of Substances as a Coping Mechanism* and then back to *High Rates of Substance Abuse.*
- **Hormonal Imbalance and Chronic Diseases (R8)**: In this loop, *High Hormone Levels* (e.g., due to high stress, poor diet) lead to a *Higher Risk of Chronic Diseases* due to *Higher Physical Distress.*
- **Genetic Predisposition and Chronic Diseases (R9)**: This loop indicates that *Genetic Susceptibility to Chronic Diseases* results in *Higher Risk of Chronic Diseases* dues to *Higher Physical Distress* and *High Hormone Levels.*
- **Cholesterol Levels and Cardiovascular Health (R10)**: This loop suggests that *High Blood Cholesterol* levels lead to an increase in *Heart Diseases* due to *Physical Inactivity.*
- **Aging and Chronic Diseases (R11)**: *Aging* increases *Physical Inactivity*, which in turn leads to *Obesity*, then resulting in a *Higher Risk of Chronic Diseases.*

3.4.6 The Stock and Flow Model for the Research

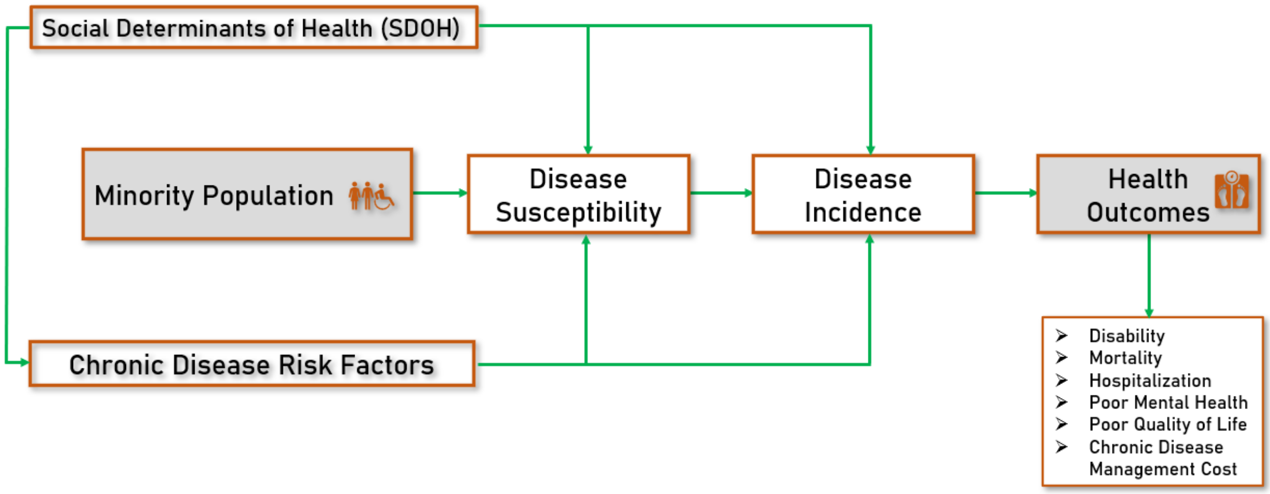

Figure 8: The System Dynamics Model Design Process.

Figure 8 shows the process design of the system dynamics model. This diagram presents how the SDOH factors influence the chronic disease risk factors in minority populations in the US. The interactions of SDOH factors and chronic disease risk factors play significant roles in the incidence of different chronic diseases (*Diabetes, Cancer, hypertension, Stroke, Asthma*, and *Chronic Obstructive Pulmonary Disease (COPD)*), and the health outcomes of the minority populations. In some cases, some chronic disease risk factors also serve as SDOH factors. This instance makes this interplay a dynamic and complex process. Moreover, the model starts with some of the minority populations being susceptible to the influence of the SDOH and chronic disease risk factors, then transiting into the susceptible minority populations part. The factors still influence these susceptible populations leading to the transition into the incidence of chronic diseases, making the susceptible populations become minority populations with two or more of these chronic diseases. Also, at this stage, these interactive factors influence these minority population regions, leading to adverse health outcomes like *disability, mortality, hospitalization, poor mental health, poor quality of life*, and *higher chronic disease management cost.*

3.4.7 Research Stocks, flows, Variables, and Interventions Tabulation

This section uses tables to explain the stocks, flows, variables, and interventions of the system dynamics model in the research. These tables enable the ease of identifying the parameters in the model. The variables are split into SDOH and chronic disease risk factors. Figures 6, 7, 8, 9 and 10 present the stocks, flows, SDOH variables, chronic disease risk factors, and intervention variables tables below.

Model Stocks
Minority Population
Susceptible Minority Population
Minority Population with Two or More Chronic Diseases
Recovered Minority Population from Two or More Chronic Diseases
Hospitalized Minority Population
Minority Population with Poor Mental Health
Proportion of the Minority Population with Mortality
Minority Population with Poor Quality of Life
Minority Population with Disability

Table 6: The Stocks employed in the Model

Model Flows
Susceptibility Transition
Chronic Disease Onset
Susceptibility After Recovery
Recovery After from Two or More Chronic Diseases
Poor Quality of Life
Mortality
Poor Mental Health Mortality
Hospitalization from Poor Mental Health
Disability Prevalence
Mental Health Prevalence

Table 7: The Flows employed in the Model

SDOH
Low School Completion
Increased Unemployment
Low Income
Higher Poverty
Poor Access to Quality Care
Limited Access to Care
Poor Air and Water Quality
Health Issues
Limited Access to Quality Food
Inadequate Provider Linguistic and Cultural Competency
Limited English Proficiency
Substandard or No Health Insurance
School Segregation
Increased Community Violent Crime
Less School Enrollment due to Safety Concerns
Poor Communication with Healthcare Providers
Poor Social Associations
Residential Segregation
Racial Discrimination

Table 8: SDOH factors employed in the Model

Chronic Disease Risk Factors
Alcohol Consumption Level (High)
Physical Distress
High Blood Sugar
High Blood Pressure Incidence
High Cholesterol Incidence
Smoking (Tobacco Abuse)
Obesity
Physical Inactivity
Age
Poor Mental Health Status
High Hormone Levels (Imbalance)
Poor Diet (Malnutrition)
Drug/Substance Abuse

Table 9: Chronic Disease Risk Factors employed in the Model

Interventions
Limited English Proficiency Interventions
Poor Air and Water Quality Interventions
Health Issues Interventions
Limited Access to Quality Food Interventions
Poor Communication with Healthcare Providers Interventions
Substandard or No Health Insurance Interventions
Increased Community Violent Crime Interventions
Less School Enrollment due to Safety Concerns Interventions
Poor Access to Quality Care Interventions
School Segregation Interventions
Inadequate Provider Linguistic and Cultural Competency Interventions
Racial Discrimination Interventions
Smoking (Tobacco Abuse) Interventions
Physical Distress Interventions
Physical Inactivity Interventions
Alcohol Consumption Level (High) Interventions
Drug/Substance Abuse Interventions
Obesity Interventions
High Blood Sugar Interventions
Poor Mental Health Status Interventions
High Hormone Levels (Imbalance) Interventions
High Blood Pressure Incidence Interventions
High Cholesterol Incidence Interventions
Poor Diet (Malnutrition) Interventions
Limited Access to Care Interventions

Table 10: The Intervention variables employed in the Model

3.4.8 The Stock and Flow Model for the Research

In the research, the system and the various variables involved are modeled and simulated using the stock and flow model to understand the influence of the interacting factors further. Table 11 presents Vensim's setting parameters in the modeling process. The model starts with an initial time of 2020 and a final time of 2070 in years, indicating 50 years for the projection of the influence of the interacting SDOH and chronic disease risk factors in minority populations. The time step is adopted to observe the interacting trends of the model critically, and the integration type employed is the Euler method. Moreover, Figure 9 shows the main system dynamics model of the research.

This model involves the previous causal loop diagrams and the combination of stock and flow parameters to deeply understand the research's complex and dynamic nature. This model considers all the variables involved in the impact of SDOH and chronic disease risk factors on these minority populations and the possible interventions to alleviate the adverse health outcomes of the interactions of these factors.

Parameter	**Value**
INITIAL TIME	2020
FINAL TIME	2070
TIME STEP	0.0078125
Units of Time	Year
Integration Type	Euler

Table 11: Time boundaries for the model

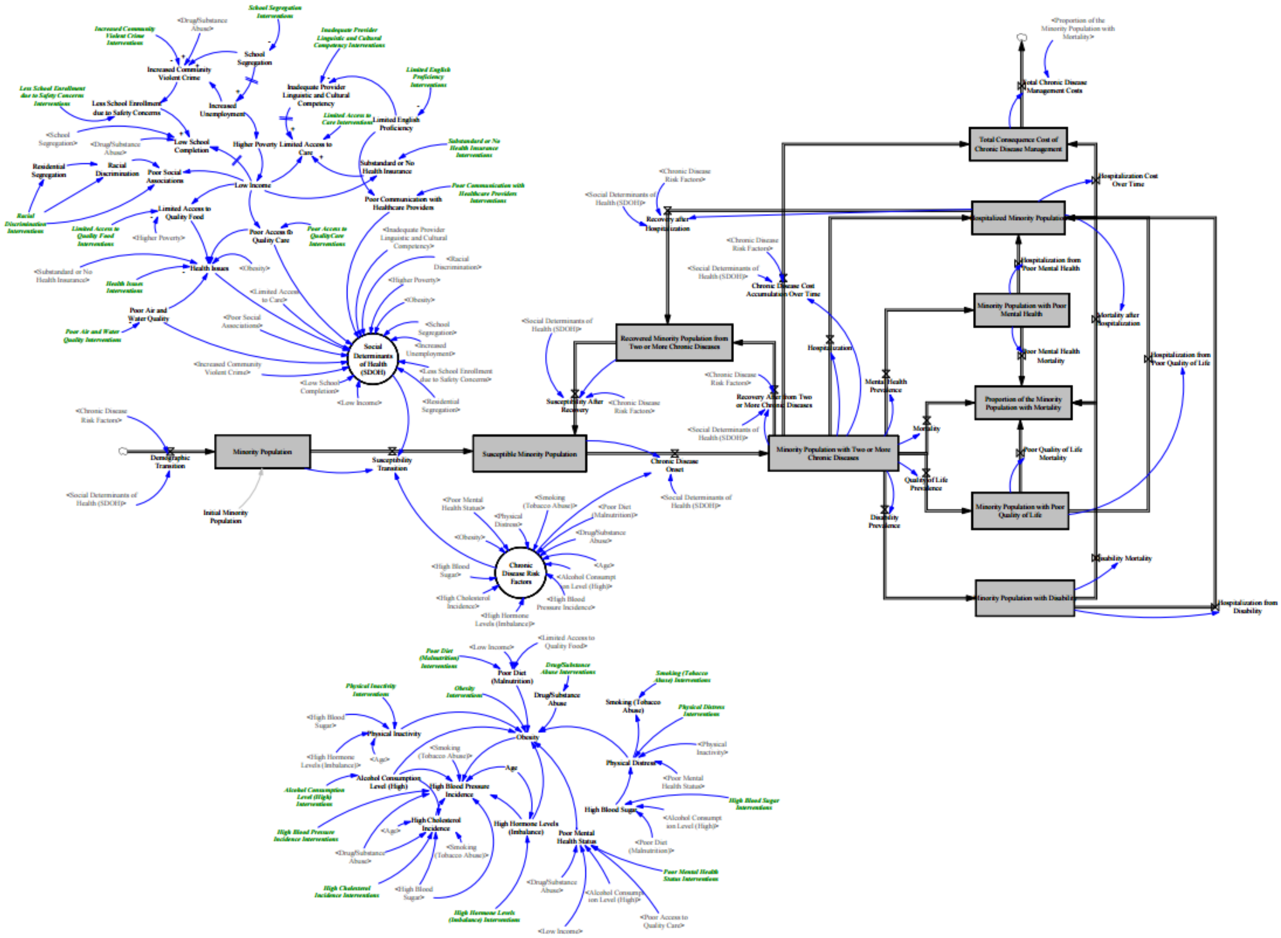

Figure 9: The System Dynamics Model using Stocks and Flows.

Figures 10, 11, 12, 13, and 14 show the various sections of the main system dynamics model for more visibility.

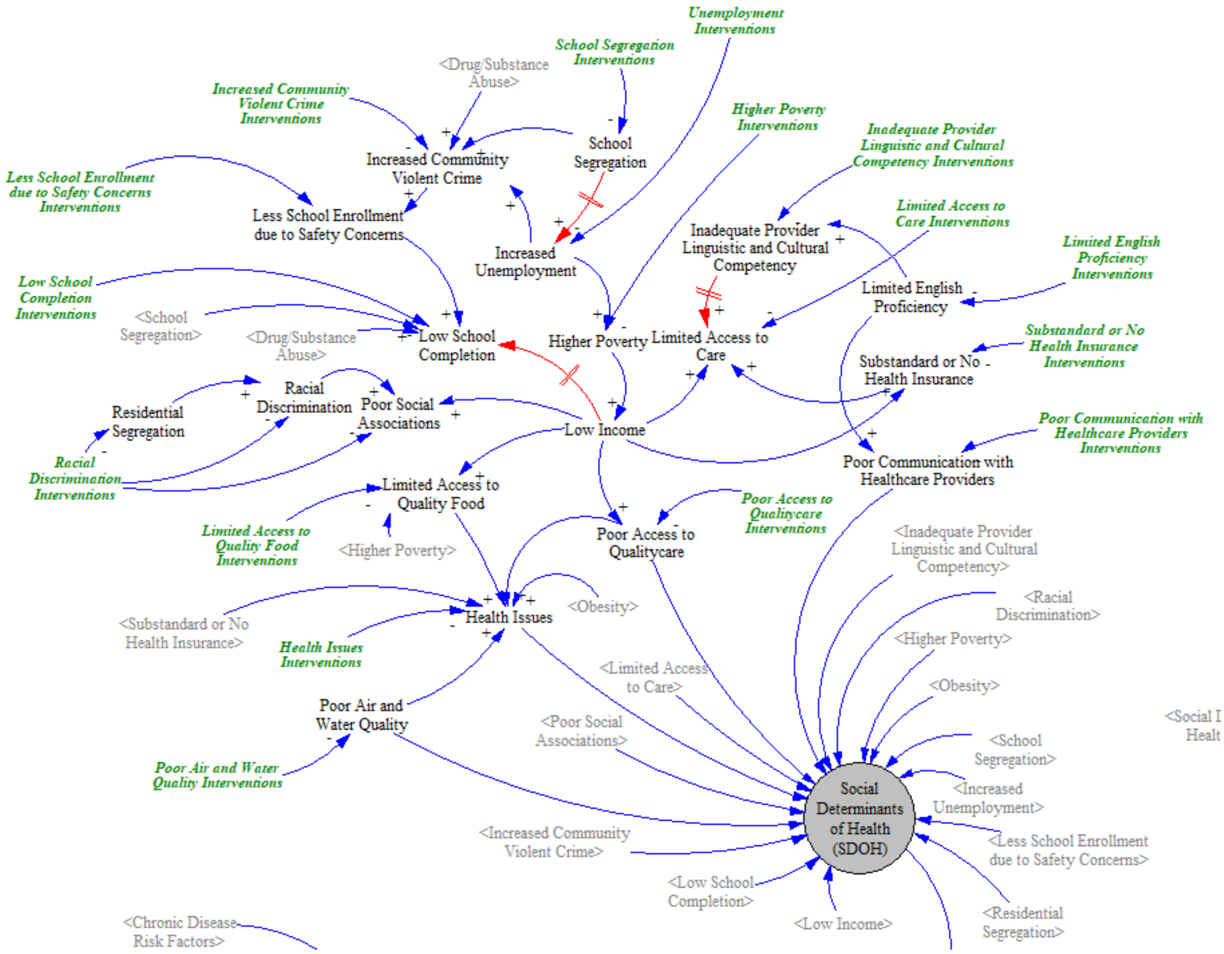

Figure 10: The System Dynamics Model Section 1.

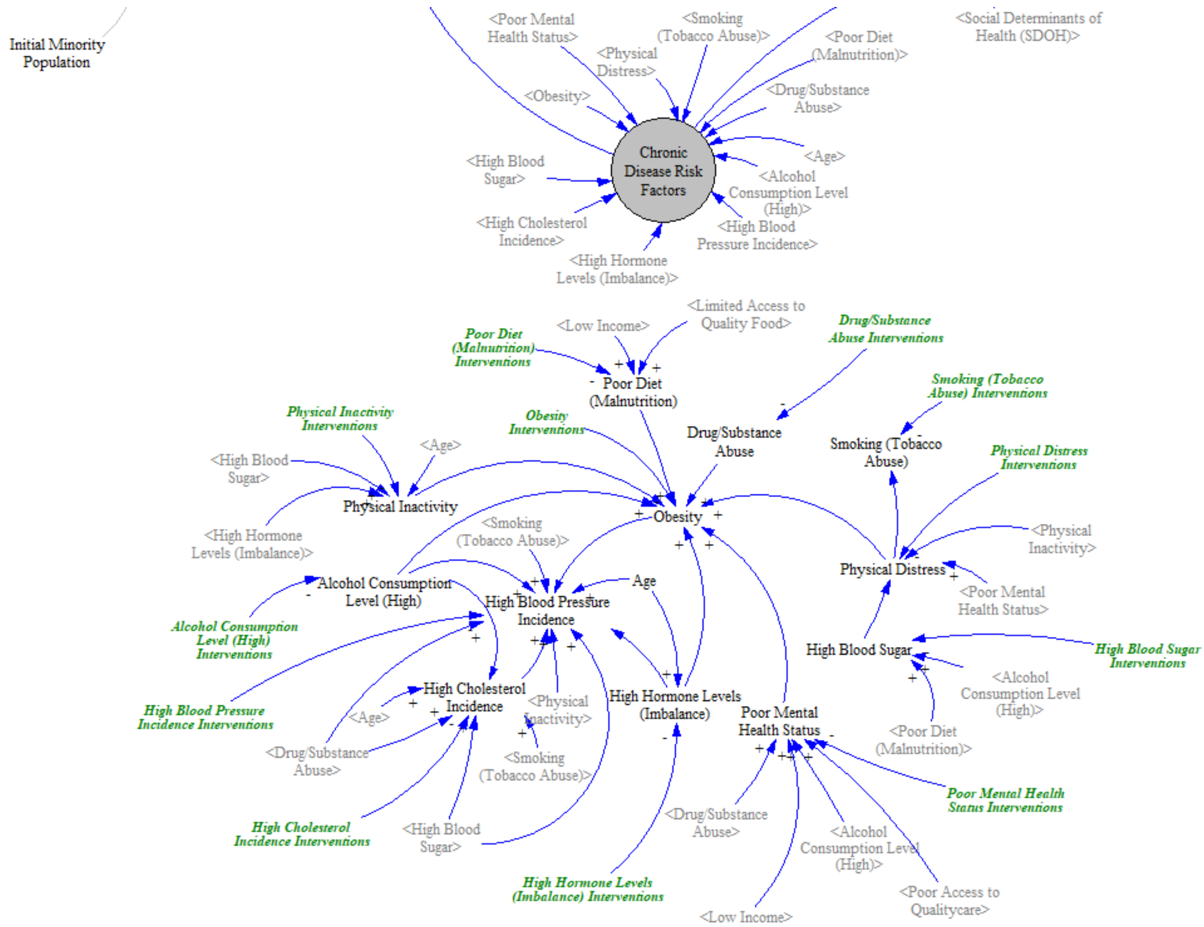

Figure 11: The System Dynamics Model Section 2.

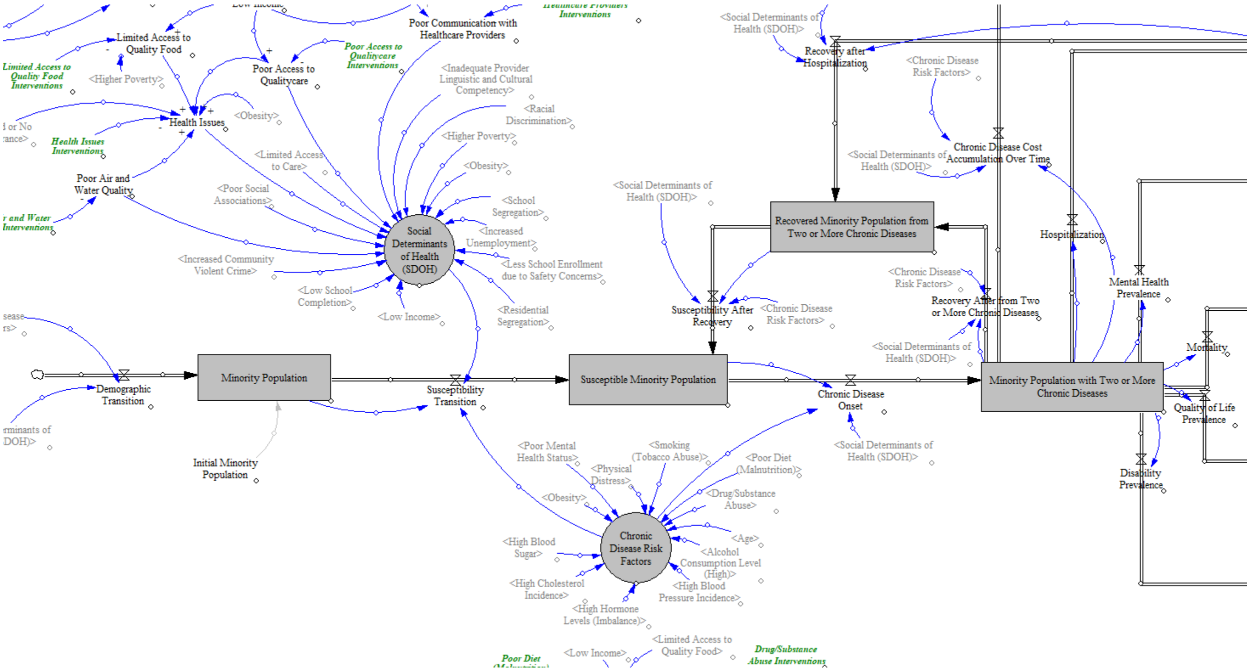

Figure 12: The System Dynamics Model Section 3.

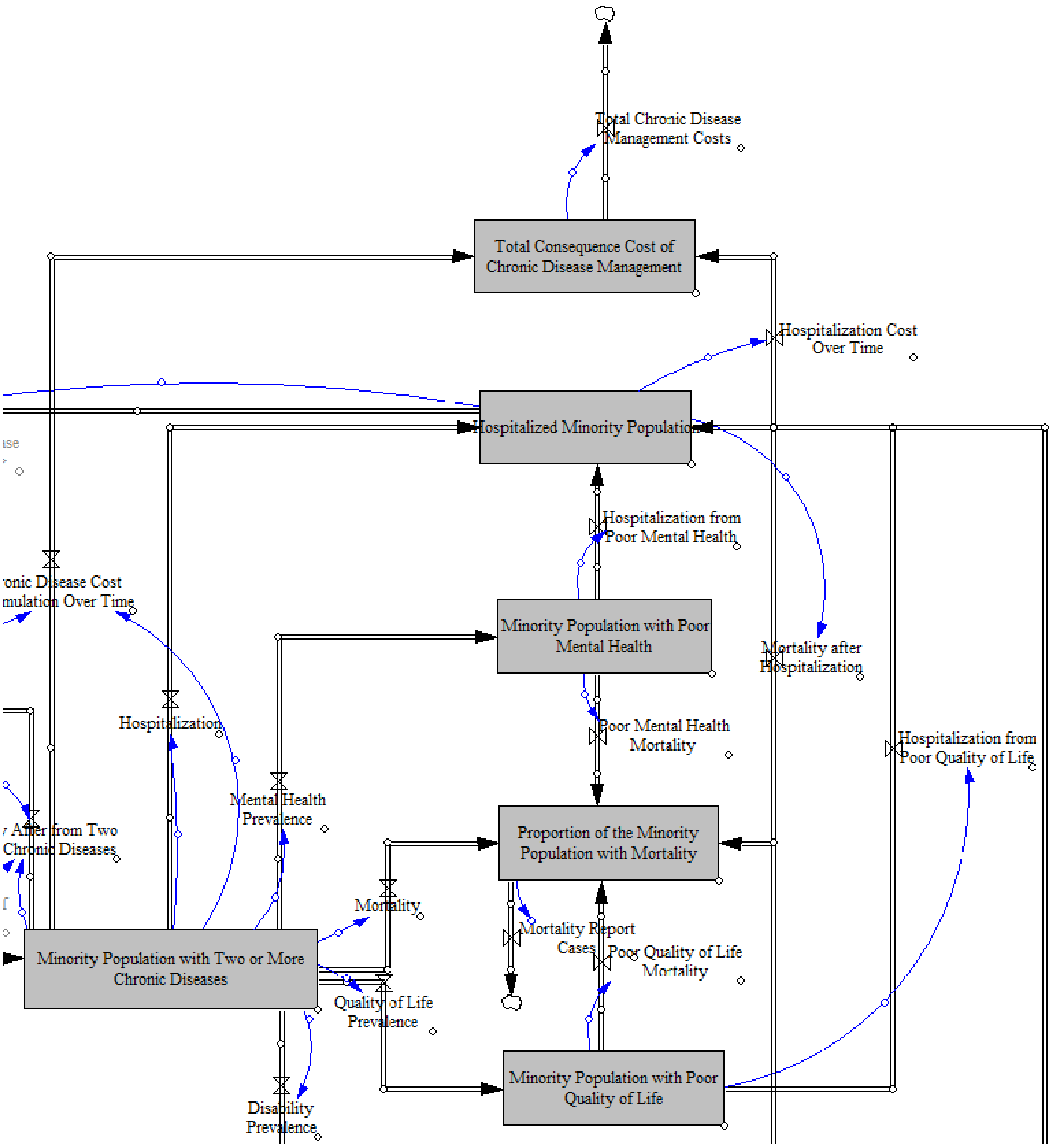

Figure 13: The System Dynamics Model Section 4.

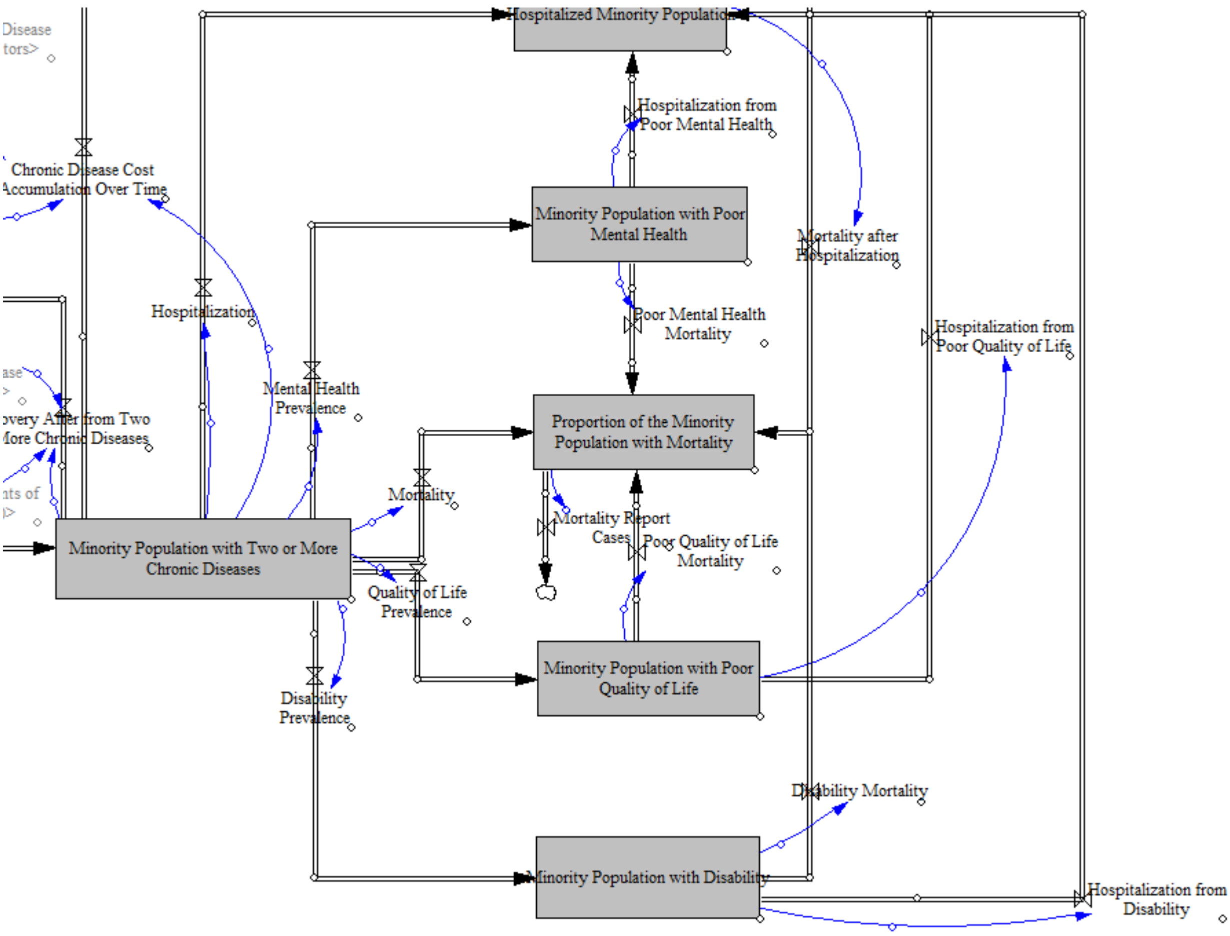

Figure 14: The System Dynamics Model Section 5.

3.4.9 System Dynamics Model Equations Analysis

A system dynamics model's general differential and integral form equations are described below. These equations connect the concepts of stocks (S), flows (inflows I and outflows O), and time (t).

The differential form is:

$$\frac{dS}{dt} = I - O \tag{1}$$

Here:

- S represents the stock or the state of the system at a given time.

- I represents the inflow, or how much is added to the stock per unit of time.
- O represents the outflow, or how much is subtracted from the stock per unit of time.
- $\frac{dS}{dt}$ represents the rate of change of the stock over time, which is equal to the inflow minus the outflow.

The integral form of the equation is:

$$S(t) = S(t_0) + \int_{t_0}^{t} (I(\tau) - O(\tau))d\tau \tag{2}$$

Here:

- $S(t)$ represents the stock at a given time.
- $S(t_0)$ represents the initial value of the stock at the starting time t_0.
- $I(\tau)$ and $O(\tau)$ represent the inflow and outflow at time τ respectively.
- The integral $\int_{t_0}^{t}(I(\tau) - O(\tau))d\tau$ represents the accumulated net inflow (*inflows minus outflows*) over the time period from t_0 to t.

In the integral form, we start with an initial stock level $S(t_0)$ and add up (*integrate*) all the net inflows over the period from t_0 to t. This gives us the stock level at time t. If the *inflows* exceed the *outflows* during this period, the stock will increase; if the *outflows* exceed the *inflows*, the stock will decrease.

3.4.10 The Equations of the Stock and Flow Model of the Research

This section illustrates the equations of the model in the research. Each parameter (in this case, *the stocks, flows, auxiliary variable*, and *interventions*) has its equation to render the simulation without errors. Generally, in the Vensim system dynamics software, the stock is usually in integral form, and the other parameters have their numerical equations, as seen in the previous section. In the following equations, the w_n represents the weight of the influence or impact of each variable in the model, where n is an integer (1, 2, 3, ... n). For example, w_1 could represent 10% of the impact of SDOH or chronic disease risk factors in the minority population in the model.

The following equations present the model's demographic, susceptible, and chronic disease onset equations over time, indicating the chronic condition's incidence in minority populations.

1.

$$\textbf{Demographic Transition} = w_1 \times \text{Social Determinants of Health (SDOH)} \\ \times w_2 \times \text{Chronic Disease Risk Factors}$$

2.

$$\textbf{Minority Population}(t) = \text{Minority Population}(t_0) + \int_{t_0}^{t} \text{Demographic Transition}(\tau) \\ - \int_{t_0}^{t} \text{Susceptibility Transition}(\tau) d\tau$$

3.

$$\textbf{Susceptibility Transition} = w_1 \times \text{Minority Population} \\ \times w_2 \times \text{Social Determinants of Health (SDOH)} \\ \times w_3 \times \text{Chronic Disease Risk Factors}$$

4.

$$\textbf{Susceptible Minority Population}(t) = \text{Susceptible Minority Population}(t_0) \\ + \int_{t_0}^{t} \text{Susceptibility After Recovery}(\tau) \\ + \int_{t_0}^{t} \text{Susceptibility Transition}(\tau) \\ - \int_{t_0}^{t} \text{Chronic Disease Onset}(\tau) d\tau$$

5.

$$\textbf{Chronic Disease Onset} = w_1 \times \text{Susceptible Minority Population} \\ \times w_2 \times \text{Social Determinants of Health (SDOH)} \\ \times w_3 \times \text{Chronic Disease Risk Factors}$$

6.

$$\textbf{Minority Population with Two or More Chronic Diseases}(t) = \\ \text{Minority Population with Two or More Chronic Diseases}(t_0) \\ + \int_{t_0}^{t} \text{Chronic Disease Onset}(\tau) d\tau \\ - \int_{t_0}^{t} \text{Chronic Disease Cost Accumulation Over Time}(\tau) d\tau \\ - \int_{t_0}^{t} \text{Disability Prevalence}(\tau) d\tau \\ - \int_{t_0}^{t} \text{Hospitalization}(\tau) d\tau \\ - \int_{t_0}^{t} \text{Mental Health Prevalence}(\tau) d\tau \\ - \int_{t_0}^{t} \text{Mortality}(\tau) d\tau \\ - \int_{t_0}^{t} \text{Quality of Life Prevalence}(\tau) d\tau \\ - \int_{t_0}^{t} \text{Recovery After from Two or More Chronic Diseases}(\tau) d\tau$$

The following equations are for the health outcomes due to the interplay of the SDOH factors and the chronic disease risk factors in minority populations.

7.

$$\textbf{Recovered Minority Population from Two or More Chronic Diseases}(t) = \\ \text{Recovered Minority Population from Two or More Chronic Diseases}(t_0) \\ + \int_{t_0}^{t} \text{Recovery After from Two or More Chronic Diseases}(\tau) d\tau \\ + \int_{t_0}^{t} \text{Recovery After Hospitalization}(\tau) d\tau \\ - \int_{t_0}^{t} \text{Susceptibility After Recovery}(\tau) d\tau$$

8.

$$\textbf{Susceptibility After Recovery} = \\ w_1 \times \text{Recovered Minority Population from Two or More Chronic Diseases} \\ \times w_2 \times \text{Chronic Disease Risk Factors} \\ \times w_3 \times \text{"Social Determinants of Health (SDOH)"}$$

9.

$$\text{Recovery After from Two or More Chronic Diseases} = \\ w_1 \times \frac{\text{Minority Population with Two or More Chronic Diseases}}{\text{"Social Determinants of Health (SDOH)"} + \text{Chronic Disease Risk Factors}}$$

10.

$$\textbf{Recovery after Hospitalization} = \\ w_1 \times \frac{\text{Hospitalized Minority Population} \times w_2}{\text{Chronic Disease Risk Factors} + \text{"Social Determinants of Health (SDOH)"}}$$

11.

$$\textbf{Hospitalized Minority Population}(t) = \text{Hospitalized Minority Population}(t_0) \\ + \int_{t_0}^{t} \text{Hospitalization}(\tau) d\tau \\ + \int_{t_0}^{t} \text{Hospitalization from Disability}(\tau) d\tau \\ + \int_{t_0}^{t} \text{Hospitalization from Poor Mental Health}(\tau) d\tau \\ + \int_{t_0}^{t} \text{Hospitalization from Poor Quality of Life}(\tau) d\tau \\ - \int_{t_0}^{t} \text{Hospitalization Cost Over Time}(\tau) d\tau \\ - \int_{t_0}^{t} \text{Mortality after Hospitalization}(\tau) d\tau \\ - \int_{t_0}^{t} \text{Recovery after Hospitalization}(\tau) d\tau$$

12.

$$\begin{aligned}
&\textbf{Total Consequence Cost of Chronic Disease Management}(t) \\
&\quad = \text{Total Consequence Cost of Chronic Disease Management}(t_0) \\
&\quad + \int_{t_0}^{t} \text{Chronic Disease Cost Accumulation Over Time}(\tau)d\tau \\
&\quad + \int_{t_0}^{t} \text{Hospitalization Cost Over Time}(\tau)d\tau \\
&\quad - \int_{t_0}^{t} \text{Total Chronic Disease Management Costs}(\tau)d\tau.
\end{aligned}$$

13.

$$\textbf{Total Chronic Disease Management Costs} = w_1 \times \frac{\text{Total Consequence Cost of Chronic Disease Management} \times w_2}{\text{Proportion of the Minority Population with Mortality}}$$

14.

$$\begin{aligned}
&\textbf{Minority Population with Poor Mental Health}(t) \\
&\quad = \text{Minority Population with Poor Mental Health}(t_0) \\
&\quad + \int_{t_0}^{t} \text{Mental Health Prevalence}(\tau)d\tau \\
&\quad - \int_{t_0}^{t} \text{Hospitalization from Poor Mental Health}(\tau)d\tau \\
&\quad - \int_{t_0}^{t} \text{Poor Mental Health Mortality}(\tau)d\tau.
\end{aligned}$$

15.

$$\begin{aligned}
&\textbf{Proportion of the Minority Population with Mortality}(t) \\
&= \text{Proportion of the Minority Population with Mortality}(t_0) \\
&+ \int_{t_0}^{t} \text{Disability Mortality}(\tau) d\tau \\
&+ \int_{t_0}^{t} \text{Mortality}(\tau) d\tau \\
&+ \int_{t_0}^{t} \text{Mortality after Hospitalization}(\tau) d\tau \\
&+ \int_{t_0}^{t} \text{Poor Mental Health Mortality}(\tau) d\tau \\
&+ \int_{t_0}^{t} \text{Poor Quality of Life Mortality}(\tau) d\tau.
\end{aligned}$$

16.

$$\begin{aligned}
&\textbf{Minority Population with Poor Quality of Life}(t) \\
&= \text{Minority Population with Poor Quality of Life}(t_0) \\
&+ \int_{t_0}^{t} \text{Quality of Life Prevalence}(\tau) d\tau \\
&- \int_{t_0}^{t} \text{Hospitalization from Poor Quality of Life}(\tau) d\tau \\
&- \int_{t_0}^{t} \text{Poor Quality of Life Mortality}(\tau) d\tau.
\end{aligned}$$

17.

$$\begin{aligned}
&\textbf{Minority Population with Disability}(t) \\
&= \text{Minority Population with Disability}(t_0) \\
&+ \int_{t_0}^{t} \text{Disability Prevalence}(\tau) d\tau \\
&- \int_{t_0}^{t} \text{Disability Mortality}(\tau) d\tau \\
&- \int_{t_0}^{t} \text{Hospitalization from Disability}(\tau) d\tau.
\end{aligned}$$

3.5 Majority and Minority Populations Distribution in the United States

Table 12 displays the population demography in 2022 [54]. The demography shows the population category based on the percentages and millions. From the table, the *White alone* race is the majority population, and the others are the minority populations. Hence, the *Native Hawaiian and Other Pacific Islander alone* is the lowest minority population this year. In the research, each population demography can be entered into the model to simulate the interaction of SDOH and chronic disease risk factors. However, since the research is about minority populations, only these populations are considered in the model to see the effect of the interacting factors in these populations.

Race/Ethnicity	**Percentage**	**Estimated Population**
White alone	75.5%	251,591,802
Black or African American alone	13.6%	45,287,112
American Indian and Alaska Native alone	1.3%	4,332,738
Asian alone	6.3%	20,997,116
Native Hawaiian and Other Pacific Islander alone	0.3%	999,862
Two or More Races	3.0%	9,998,627
Hispanic or Latino	19.1%	63,677,974
White alone, not Hispanic or Latino	58.9%	196,202,392

Table 12: Population Distribution in the United States (2022).

3.6 The Application of PySD in the Research

This section provides a comprehensive understanding of the PySD in the research. The *PySD library* facilitates the efficient translation and initiation of the Vensim-based system dynamics model in a Pythonic environment. The dynamic model manipulation was made more accessible by the library's ability to read Vensim models and turn them into Python scripts. It also assured compatibility with several powerful Python libraries. Moreover, the Python library Pandas [55] was employed to manipulate and analyze the data from the model. The Pandas' high-performance, user-friendly data structures, and analysis made handling the model's data accessible. The research data visualizations were displayed using Matplotlib [56] and Seaborn [57] Python plotting libraries. The Seaborn library is a statistical graphics package built on Matplotlib, which allowed the creation of visually beautiful and instructional statistical graphics. Matplotlib offered a dependable platform on which we could build a variety of plots.

Algorithm 1 Run System Dynamics Model and Save Output to CSV

```
procedure RunSDModel
    model ← read_vensim('SDOH-Chronic Disease Risk Factors.mdl')
    output ← model.run()
    Print(output)
    output.to_csv('SDOH-Chronic Disease Risk Factors.csv')
end procedure
```

The Algorithm 1 initiates the ***SDOH-Chronic Disease Risk Factors.mdl*** file. It implements the ***read_ vensim***, the system dynamics model from the Vensim software. The ***model.run*** function runs the model and saves the result in the output variable, which shows the model's dynamics over time. This output rapidly evaluates the model for the programming process. Moreover, the output is saved in a ***csv*** format as ***SDOH-Chronic Disease Risk Factors.csv***.

3.7 The Application of Machine Learning (ML) in the Research

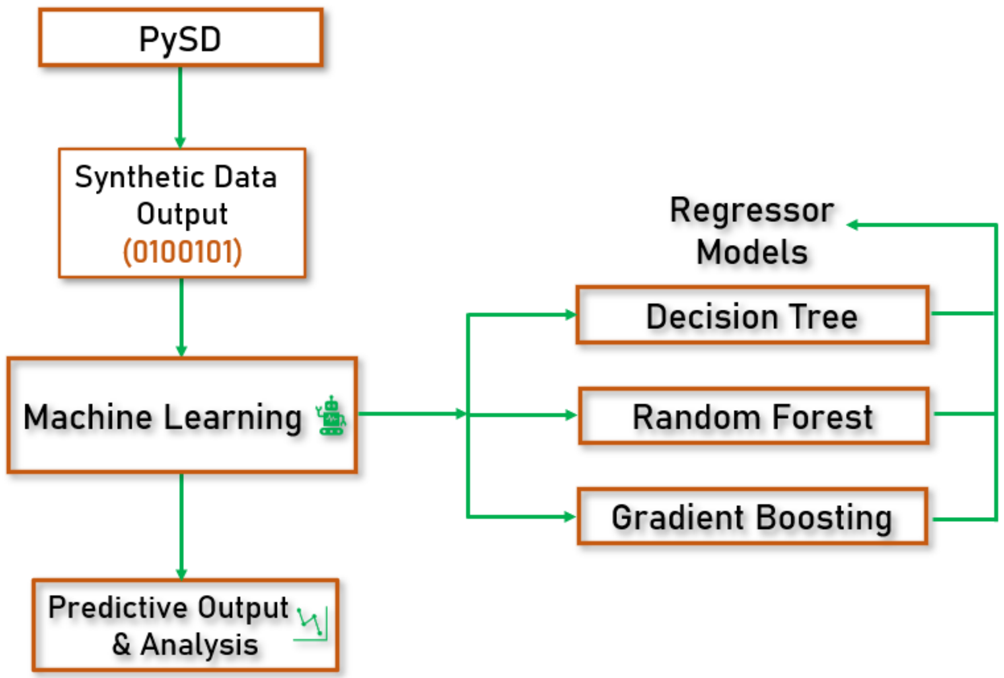

Figure 15: The Research Machine Learning Procedure.

Figure 15 describes how the *machine learning process* is arranged in the research. After the *PySD library* has been utilized to translate the *Vensim model* into the Pythobc environment, the numeric outputs of the simulation were saved and then employed in the machine learning procedure. From the figure, three machine learning regressor algorithms, *Decision Tree, Random Forest, and Gradient Boosting*, are used in the research for the variable prediction and analysis.

4 Results and Discussions

This section of the research explains the graphical and tabular results from both system dynamics modeling and machine learning algorithms, along with the discussions of these results from the models. The following subsections explore the outputs of each trend from the *exploratory* and *predictive* models.

4.1 Chronic Disease Susceptibility, Onset, and Prevalence Over Time

This section explains the susceptibility, incidence, and prevalence of the minority populations with time due to the impact of the SDOH and the chronic disease risk factors.

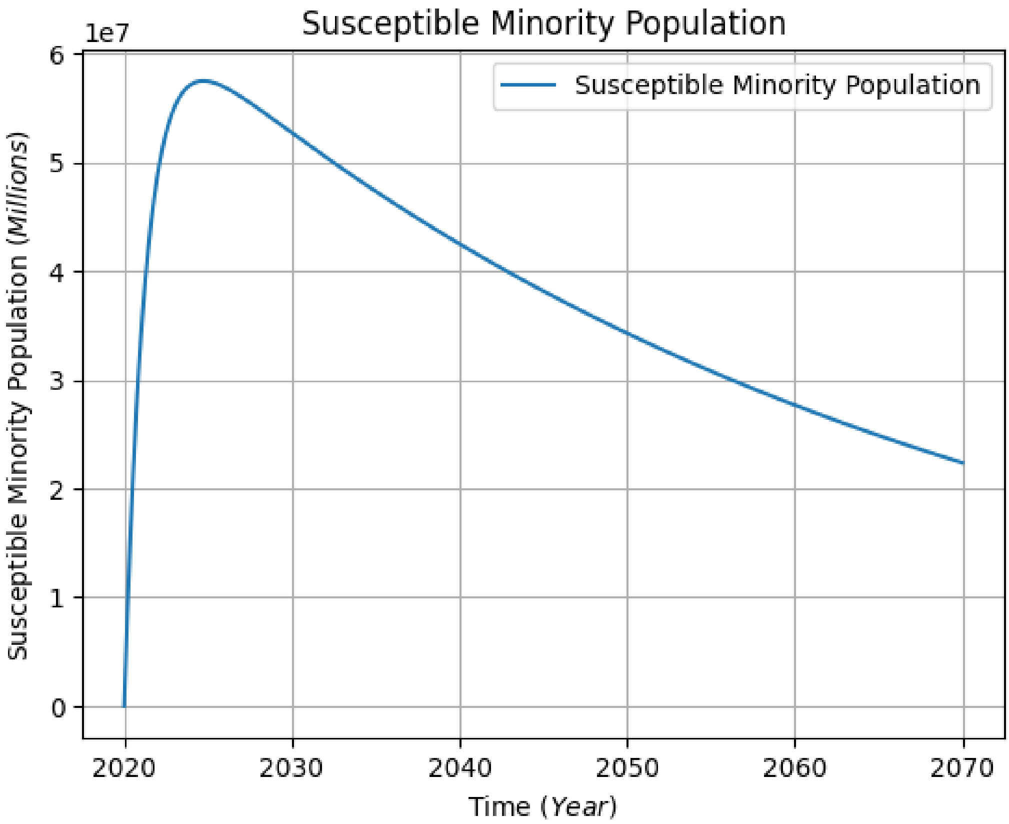

Figure 16: Susceptible Minority Population Over Time

In Figure 16, the graph displays the trend of the susceptible minority population in the United States over time. These minority populations often experience the inappropriate burden of chronic diseases and the impact of SDOH based on ethnicity, socioeconomic status, race, environmental exposure, and access to healthcare. Minority communities are, therefore, more susceptible to disparities in health or adverse health outcomes, making them vulnerable to various chronic health issues. The plot indicates an increase in the susceptible minority population from 0 in 2020 to around 5.5 (*Million*) between this year and 2030, simultaneously showing the increase's pace. this trend declined to about 2.2 (*Million*) by 2070.This circumstance may suggest that several intervention practices, such as social justice and health equity for these groups, have improved or that the definition and assessment of vulnerability have evolved.

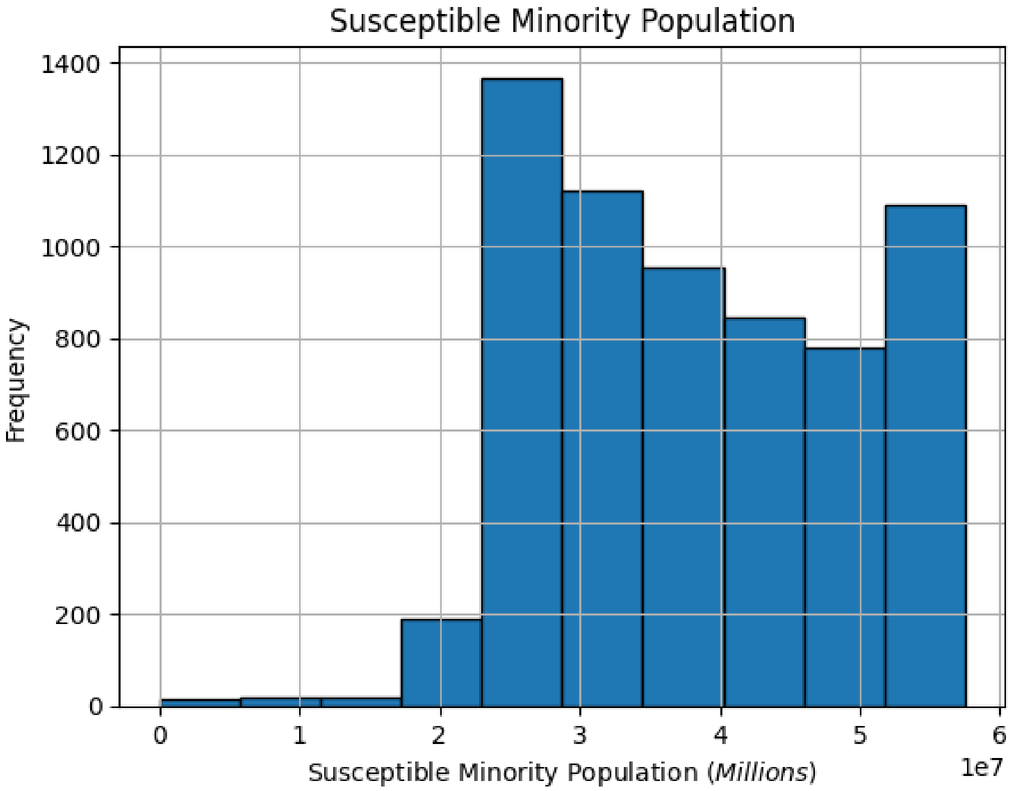

Figure 17: Susceptible Minority Population Distribution

Figure 17 depicts the *susceptible minority population's distribution* rather than the previous line plot. The distribution's fifth bar shows that the most susceptible minority population size is 20–30 (*Million*). This analysis implies that many states or regions have numerous susceptible populations with health issues. The first bar in the plot represents the least standard susceptible minority population size, between 0 and 2 (*Million*). Moreover, the graph shows that the distribution is left-skewed, meaning more states or regions have higher vulnerable minority populations based on the influence of SDOH and chronic disease risk factors in these regions. This analysis suggests that US racial and ethnic groups may have unequal access to health treatment and prevention. The line plot in Figure 16 and Figure 17 shows the nature and trend of chronic disease susceptibility in the minority population according to the model.

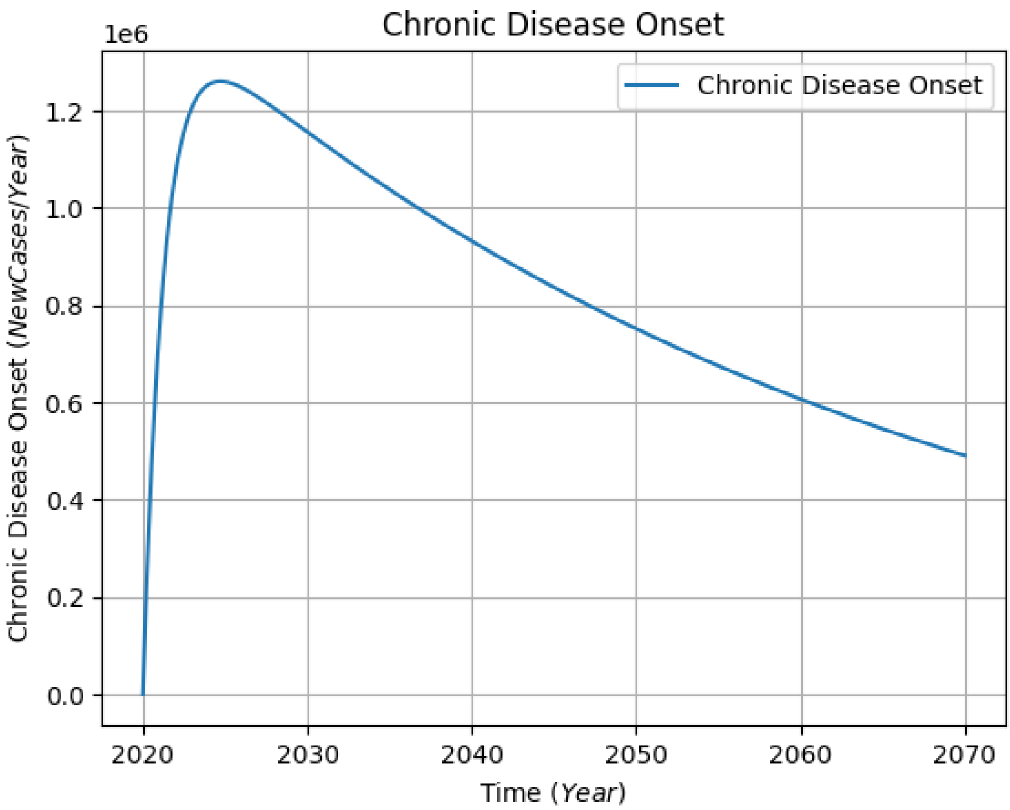

Figure 18: Chronic Disease Onset Over Time

Figure 18 shows the line plots of the onset of chronic diseases in the minority populations in the US over time. It projected the rise and decline of the incidence of these chronic conditions based on the new cases per year in these populations from 2020 to 2070, indicating 50 years forecast as mentioned previously. As seen from the plot, the number of (*New Cases/Year*) rises from 0 in (*Million*) to around a peak of 1.22 (*New Cases/Year*) in (*Million*) between the years 2020 to 2030. This rise could result from poor management of chronic disease risk factors and the influence of SDOH in these communities. Afterward, there is a further decrease in the curve from the year 2030 to the year 2070 at 1.19 (*New Cases/Year*) in (*Million*), which is due to some intervention measures in these minority communities.

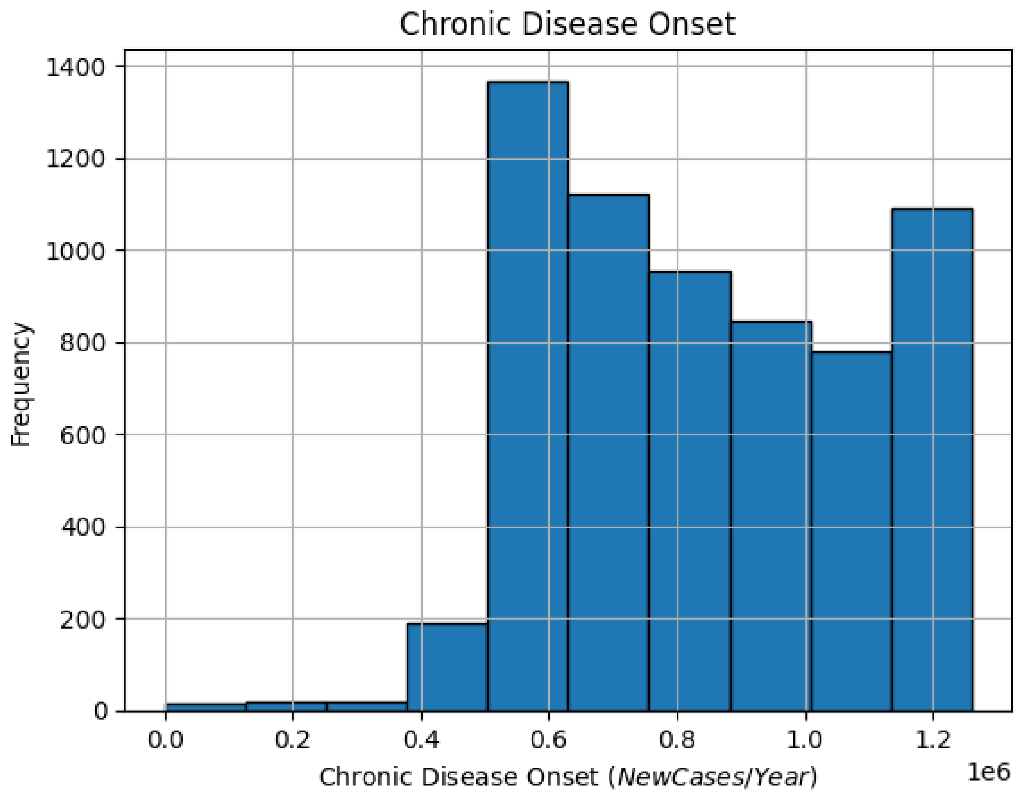

Figure 19: Chronic Disease Onset Distribution

According to Figure 19, the most *onset of chronic diseases* in the distribution is between 0.4 ($Million$) and 0.7 ($Million$), which corresponds to the graph's highest bar. This situation indicates that many states yearly have numerous chronic disease cases in their minority population areas. The lowest bar in the distribution represents the least common chronic disease onset, which is between 0 ($Million$) and 0.4 ($Million$). Hence, few states or regions have low yearly chronic disease cases. Moreover, the distribution is left-skewed, showing that more states have higher rates of chronic disease onset than lower ones. This analysis may highlight U.S. chronic illness prevention and management issues. Lifestyle, environment, age, and healthcare access often cause these issues.

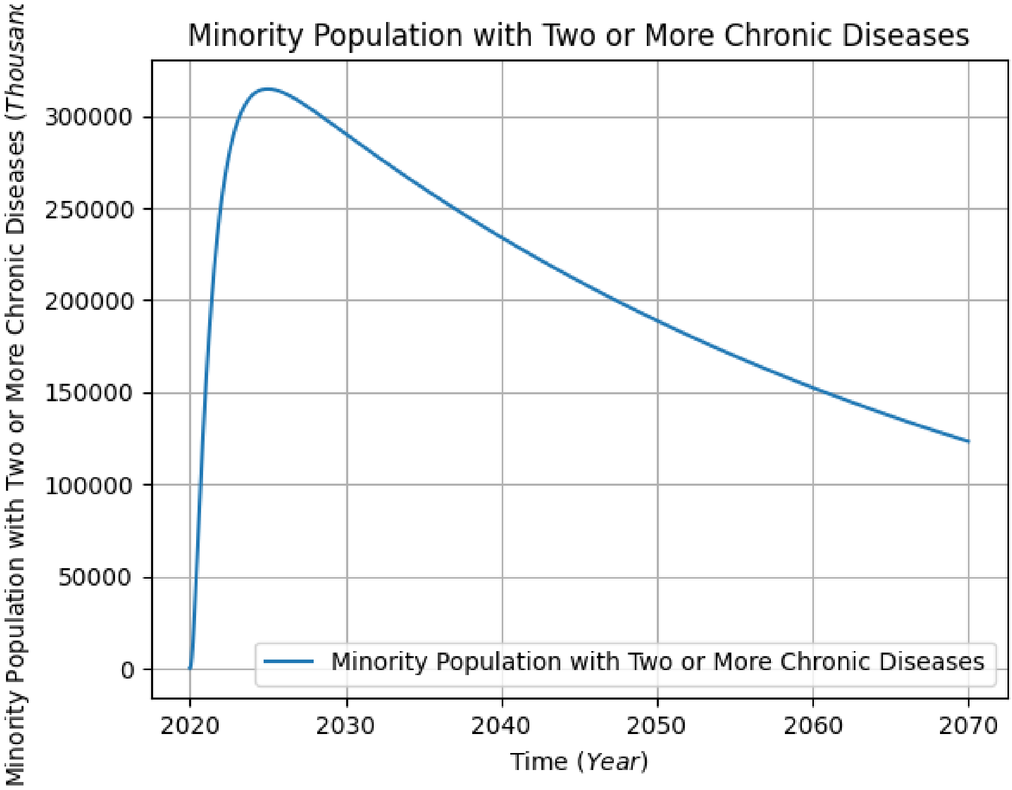

Figure 20: Minority Population with Two or More Chronic Diseases Over Time

Figure 20 displays the *minority population with two or more chronic diseases* from 2020 to 2070 in 10-year increments for 50 years. These projected minority population with two or more chronic diseases declines differently each year, as seen in the graph. Between 2020 and 2025, the minority population will increase to about 350000 (*Thousand*) as its peak from 0 (*Thousand*). Afterward, a gradual decline occurs from around 2027 to 2030. Then moving forward, the curve declines through years 2040, 2050, 2060, until 2070, indicating that the line steepens with time. This declining effect experienced by these minority populations with these chronic diseases may be due to future initiatives or policies that may lessen the risk or burden of chronic diseases among racial and ethnic minorities.

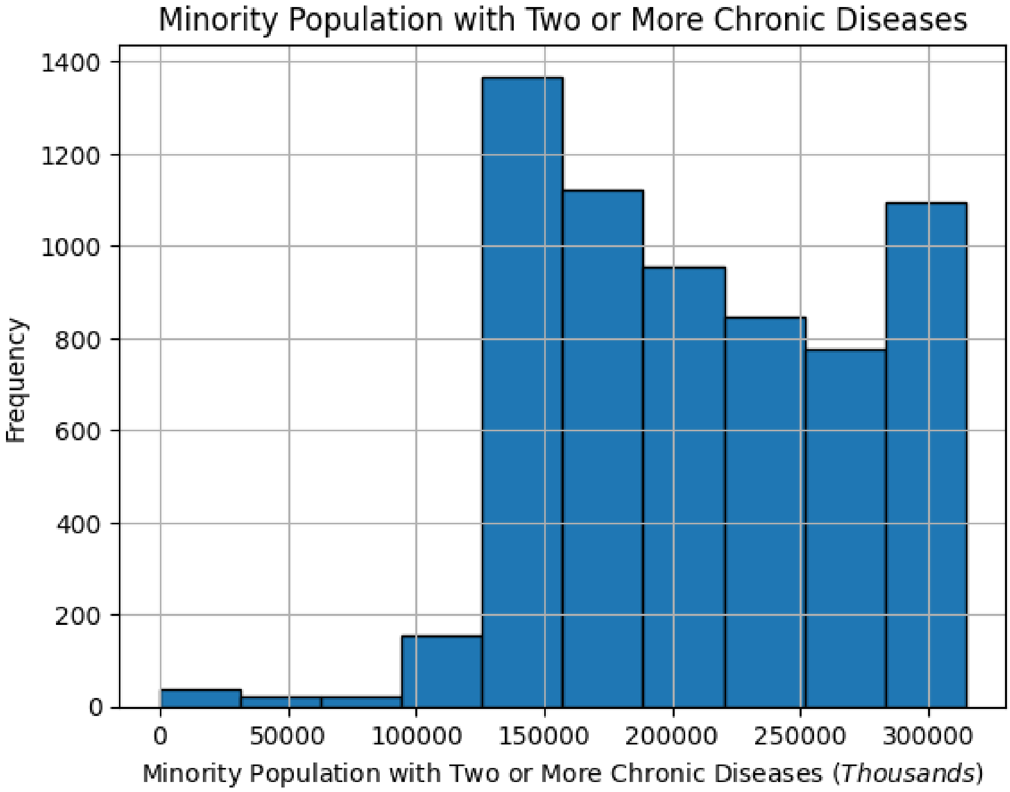

Figure 21: Minority Population with Two or More Chronic Diseases Over Time

Figure 21 displays the distribution of minority populations with two or more chronic diseases. The horizontal axis represents the number of these minority populations, and the vertical axis represents the frequency of the populations. The graph shows that the highest is around 150000 ($Thousand$), and the lowest is between 50000 ($Thousand$) and 100000 ($Thousand$). The graph shows that the distribution is left-skewed, indicating that more regions or states have higher minority populations with two or more chronic diseases than others. This situation may be due to the intervention procedures undertaken at the beginning of the year, which may have reduced the minority population type. Further, with time, the interventions declined, making the minority populations increase.

4.2 Implementing Interventions' Effect on SDOH and Chronic Disease Risks Factors in Minority Populations Over Time (*Sensitivity Analysis*).

This section explains the *sensitivity analysis* of the model by varying the input intervention parameters. This procedure is undertaken to view the impact of the various interventions in the interplay between SDOH and chronic disease risk factors in minority populations. The following subsections below explain more about the influence of these interventions.

4.2.1 Health Issues Intervention Effect on the Monirty Population

Health interventions encompass activities or approaches to evaluate, enhance, sustain, advance, or alter health for individuals or a whole community. These interventions comprise educational or healthcare initiatives, policy alterations, environmental enhancements, or campaigns to promote health. Complex interventions consist of several separate or interrelated elements [59].

Similarly, chronic disease management interventions preserve patient lives and minimize costs with well-designed healthcare treatments. These comprehensive therapies focus on early diagnosis and treatment of ongoing issues, slowing disease development. In this sense, self-care patients should actively collaborate with doctors. This partnership is vital since chronically ill patients need several therapies. By maximizing these actions, healthcare assets are used effectively, improving patient outcomes [60].

The intervention values in the graphs below illustrate how much effort or resources are required to address how the interactions of SDOH and chronic disease risk factors affect minority populations in the US. The various intervention values (0.1, 0.2, 0.3, 0.4, and 0.5) represent the percentage of how these influences impact these populations. As an illustration, an intervention value of 0.1 corresponds to 10%, 0.2 to 20%, 0.3 to 30%, 0.4 to 40%, and 0.5 to 50%.

4.2.2 Health Issues Intervention Effect on Monirty Population with Two or More Chronic Diseases

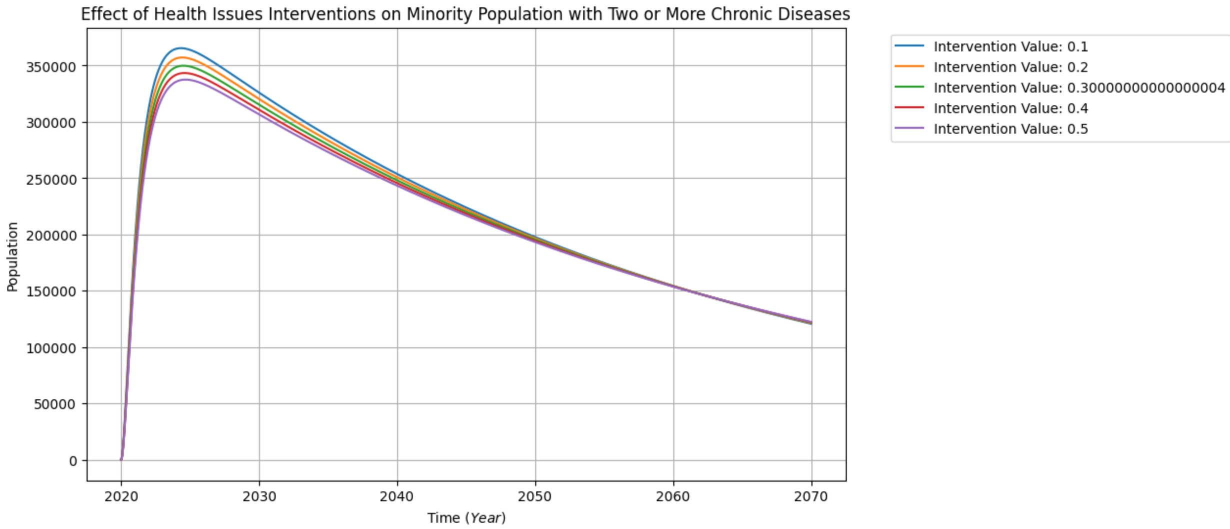

Figure 22: Health Issues Intervention Effect on Monirty Population with Two or More Chronic Diseases

Figure 22 shows the health issues intervention effects on minority populations with two or more chronic diseases. Generally, these interventions aim to reduce the number of these minority communities. As illustrated initially in the previous section, the varying intervention values represent the effect of these interventions on the SDOH and chronic diseases in these populations. The horizontal axis represents the year duration in the model from 2020 to 2027, and the vertical axis represents the number of the minority population with two or more chronic diseases. Starting in 2020, we can see that these interventions' influences are latent from 0 (*Thousands*) to around 250000 (*Thousands*) of the population. After that, there is a visible divergence and a peak of 350000 (*Thousands*), indicating the effects of these interventions between 2020 and 2025. These signify the purpose of the interventions by reducing these populations. Furthermore, we can see a decline and a convergence in the intervention effects between 2030 and 2070. The decline can be due to the following factors: *inadequate adherence, delayed outcomes, chronic disease development, inadequate intensity or duration, comorbidities, socioeconomic considerations, measurement difficulties, hereditary factors, and psychological concerns.*

4.2.3 Health Issues Intervention Effect on Recovered Monirty Population from Two or More Chronic Diseases

Recovery from chronic diseases is usually influenced by the nature of the diseases, the patient's health, lifestyle, and early discovery [58]. Some chronic disease conditions, such as *Type 2 Diabetes* and *hypertension*, can be effectively treated with dietary adjustments and proactive medical care. Moreover, *Chronic Obstructive Pulmonary Disease* (COPD), *Heart Disease*, and a few *autoimmune diseases* can benefit from proper care and early intervention. Also, early detection and management are essential for *Chronic Kidney Disease* (CKD) and some *malignancies* for earlier recovery. Generally, effective management of chronic diseases depends on regular checkups, treatment compliance, and good lifestyle choices.

Figure 23 shows the apparent increase in the recovery rate of the minority populations from two or more chronic diseases such as *kidney diseases, cancer, diabetes, Alzheimer's disease, kidney diseases, autoimmune diseases, chronic respiratory diseases, heart disease, stroke, and osteoporosis*, based on the effect of an increase in health issues interventions (*medication, physical therapy, lifestyle changes, alternative and complementary therapies, rehabilitation, psychotherapy, vaccination, surgery, health education, and palliative care* in the model over 50 years.

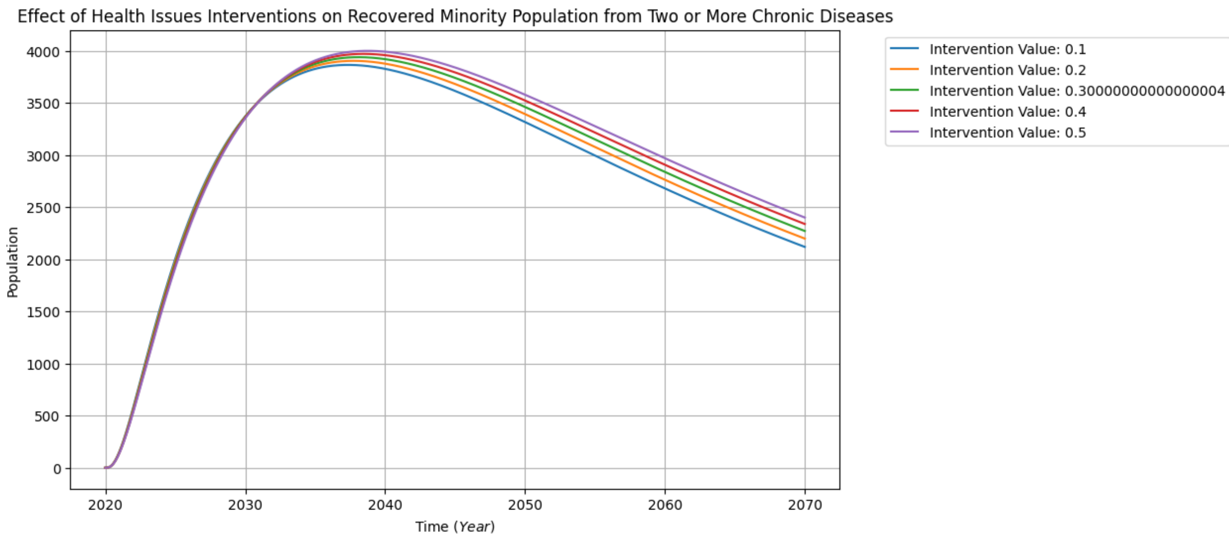

Figure 23: Health Issues Intervention Effect on Recovered Monirty Population from Two or More Chronic Diseases

In figure 23, starting in 2020, all five recovery rates due to the health issues interventions influence the minority population with chronic diseases to rise rapidly from 0 to 4000 (*Thousands*) between the years 2020 and 2035. However, there were few significant effects since the curves indicate a narrower flow. Afterward, the model projects a significant increase in recovery rates, with the 50% intervention rising the most and the 10% the least. Between 2035-2070, the model projects a gradual decline in minority population chronic disease recovery at approximately 2200 (*Thousands*), but with rises in the interventions' influences. Notably, the 50% and 10% interventions diverge significantly, suggesting further interventions may alleviate the prevalence of chronic diseases in these communities. Despite the apparent decline in the recovered minority populations from these diseases in the projected years, further interventions still diverge, signifying that with more interventions, there will be more recovered minority populations from two or more chronic disease conditions.

4.2.4 Obesity Intervention Effect on the Monirty Populations

Effective and proactive interventions can prevent obesity. These interventions come in various ways, such as regular physical exercises, eating healthier food, reducing stress, improving sleep, and lowering high cholesterol. Multicomponent obesity prevention treatments by medical professionals who emphasize proper diet intake and physical exercise through educational, environmental, and behavioral activities or interventions such as counseling and coaching [61] would significantly alleviate obesity.

4.2.5 Obesity Interventions Effect on Monirty Population with Two or More Chronic Diseases

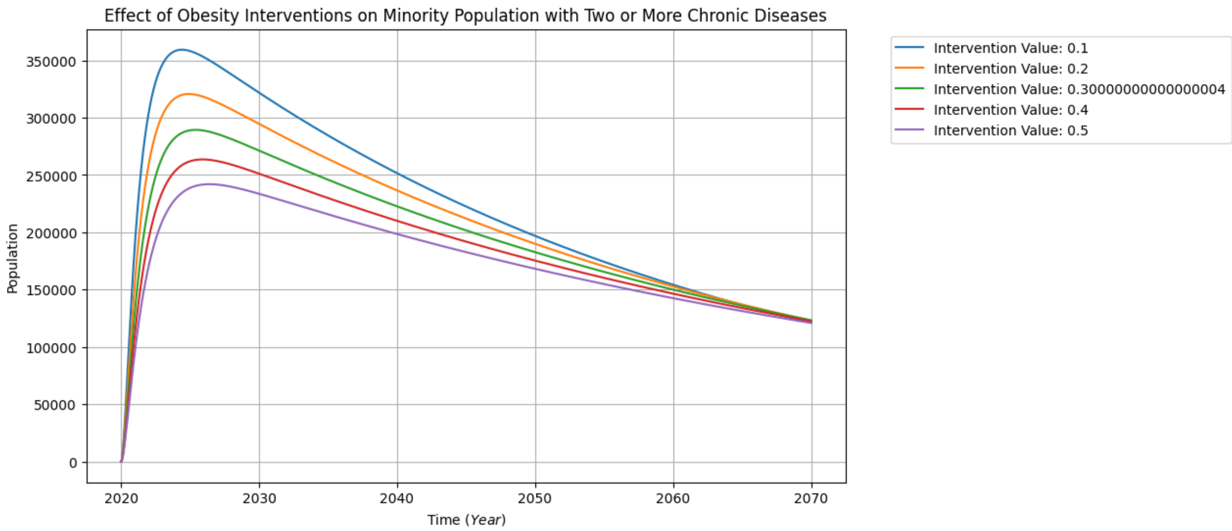

Figure 24: Obesity Interventions Effect on Monirty Population with Two or More Chronic Diseases

Figure 24 shows that in 2020, all five intervention values started simultaneously, indicating that minority populations with two or more chronic diseases had a comparable population regardless of intervention value. The blue curve represents the initial obesity intervention which has a 10% influence on these minority populations as a risk factor. With further increases in the interventions from 20% to 50%, we can see a decrease in the minority populations. Moreover, starting in the year 2020 in the model, the minority population with the disease increased from 0 (*Thousand*) to about 360,000 (*Thousand*), indicating an increase in the effect of the interaction of SDOH and the chronic disease risk factors, specifically, obesity with a further decline of these populations to about 130000 (*Thousand*) over time to the year 2070. With the initialization of the obesity interventions based on their influences in percentages (10%, 20%, 30%, 40%, 50%), we can see declining diverging curves of the intervention values, with the 20% reducing the initial peak of the 10% influence on the population from 360000(*Thousand*) to around 320000 (*Thousand*), then with the other intervention values decreasing the number of the people further to (30%:280000 (*Thousand*)), (40%:260000 (*Thousand*)), (50%:240000 (*Thousand*)). In the year 2070, all the intervention values converged, following the trend of the initial projection of the model. All the obesity interventions indicated a decrease in the minority population with two or more chronic diseases.

4.2.6 Obesity Interventions Effect on Recovered Minority Population with Two or More Chronic Diseases

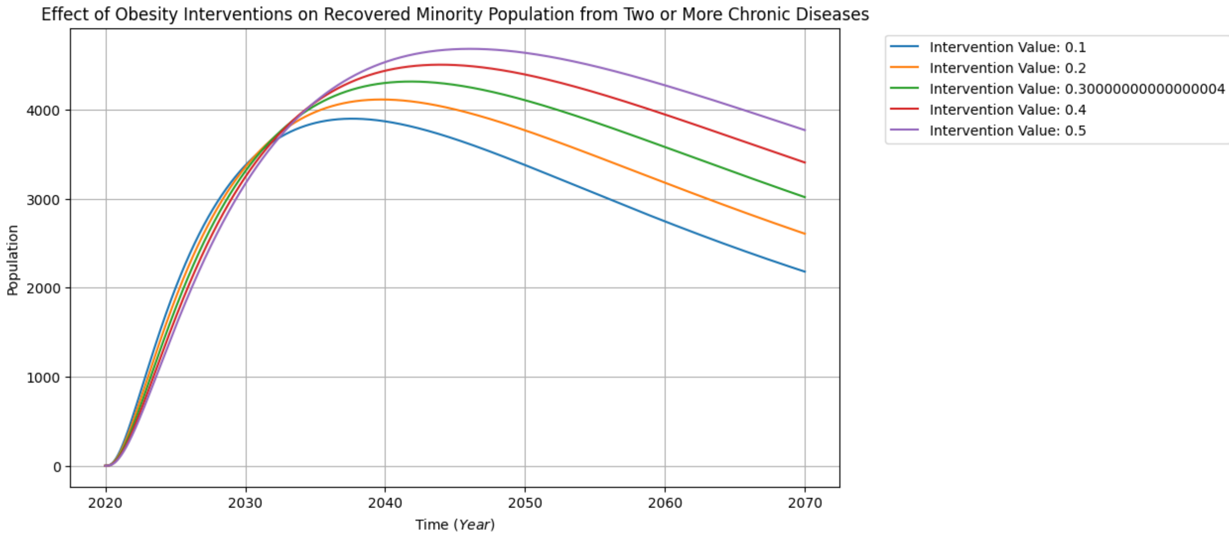

Figure 25: Obesity Interventions Effect on Recovered Minority Population with Two or More Chronic Diseases

In Figure 25, starting at 2020 in the graph, the obesity intervention values initially beginning with the 10% influence decreases the number of the recovered minority population from two or more chronic diseases from around 3000 (*Thousand*) in 2027 to about 2800 (*Thousand*) at 2029, in this case, the intervention effects are still latent at this point. Afterward, around 2034, we can see an intersection of these interventions at a point corresponding to 3700 (*Thousand*) of the recovered minority population, where their influence began to take effect in reducing the number of the recovered minority population from chronic diseases. After that, with further increments in obesity interventions, we can see visible increments (*diverging intervention values, 20%, 30%,40%,50%*) in the recovered population from 2040, 2050, 2060, to 2070. In this case, the intervention processes indicate the increase in recovered minority populations from two or more chronic diseases due to the influence of the various intervention measures.

4.3 Machine Learning Model Predication and Analysis

As mentioned in section 3.6, the PySD library is employed to translate the system dynamics modeling process into Python scripted process. Afterward, the output simulated data from the modeling is saved and fed into the machine learning process. According to the machine learning process, the variables of the system dynamics model can be modeled as a *classification* or a *regression* model taking one of the variables as a target. Also, these variables can be modeled as a *time-series forecasting* model based on the year of simulation of the system dynamics model. In this research, the regression model is adopted based on the nature of the variable. Considering the predictive model analysis for the impact of the SDOH and the chronic disease risks interacting factors in the minority populated regions, the *Minority population with two or more chronic diseases* is taken as the target, and the rest of the variables are taken as features. Three regression algorithms are employed in the machine learning procedure for the predictive analysis. The algorithms are *Random Forest*, *Gradient Boosting*, and *Decision Tree.* The algorithms are explained in the subsequent sections 4.3.1, 4.3.2, and 4.3.3.

4.3.1 Random Forest

A *Random Forest technique* is a popular supervised machine learning technique for Classification and Regression challenges in machine learning. It has multiple decision trees and outputs the mode of the individual trees' classes (classification) or mean prediction (regression) [62].

1. **Bootstrap Sampling**

$$D_i = \{(x_1^i, y_1^i), (x_2^i, y_2^i), \ldots, (x_n^i, y_n^i)\}$$

Where:

- D_i: i-th bootstrap sample
- x: features
- y: labels

Using a random sampling method that incorporates replacements, the equation above produces several subsets of the initially provided dataset.

2. **Final Prediction for Classification**

$$y = \text{mode}\{y_1^1, y_1^2, \ldots, y_1^T\}$$

Where:

- y: final prediction
- T: total number of trees

This equation can aggregate the predictions using the median from the several distinct decision trees used for categorization.

3. **Final Prediction for Regression**

$$y = \frac{1}{T}\sum_{i=1}^{T} y_1^i$$

Where:

- y: final prediction
- T: total number of trees

This equation aggregates the results of the trees ' predictions by taking the mean of all the individual decision trees used for regression.

4.3.2 Gradient Boosting

Gradient Boosting builds models incrementally by fitting weak learners to the negative gradient of the loss function [64].

1. **Initialization**

$$F_0(x) = \text{arg min}_\gamma \sum_{i=1}^{N} L(y_i, \gamma)$$

Where:

- $F_0(x)$: initial model
- L: loss function
- y_i: actual label
- γ: constant

This equation sets the initial model, usually a constant value that minimizes the loss function.

2. **Gradient Descent Step**

$$r_{it} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{t-1}(x)}$$

Where:

- r_{it}: residual
- L: loss function
- y_i: actual label
- $F(x)$: model

This equation calculates the negative gradient of the loss function, guiding the direction for the next weak learner.

3. **Update**

$$F_t(x) = F_{t-1}(x) + \rho_t h_t(x)$$

Where:

- $F_t(x)$: updated model
- ρ_t: learning rate
- $h_t(x)$: weak learner

This equation updates the model by adding a fraction of the newly fitted weak learner.

4.3.3 Decision Tree

Decision Trees partition the feature space into regions and assign a label (or value) to each region [63].

1. **Split Criterion (e.g., Gini Impurity)**

$$G = 1 - \sum_{i=1}^{k} p_i^2$$

Where:

- G: Gini impurity
- p_i: probability of class i

This equation determines the optimal characteristic to divide the data based on the impurity of the set of items.

2. **Information Gain**

$$IG = H(D) - \sum \frac{|D_v|}{|D|} H(D_v)$$

Where:

- IG: information gain
- $H(D)$: entropy of the dataset
- D_v: subset of the dataset

This equation determines the optimal way to split the data since it estimates the entropy decrease resulting from dividing the data.

3. **Tree Pruning (e.g., Cost Complexity Pruning)**

$$C_\alpha(T) = C(T) + \alpha|T|$$

Where:

- $C_\alpha(T)$: cost complexity
- $C(T)$: cost of the tree
- α: complexity parameter
- $|T|$: number of terminal nodes

This equation decreases the forest by cutting away branches that do not contribute significant information for prediction.

4.3.4 Analysishe of Actual and Predicted Outputs of the Model

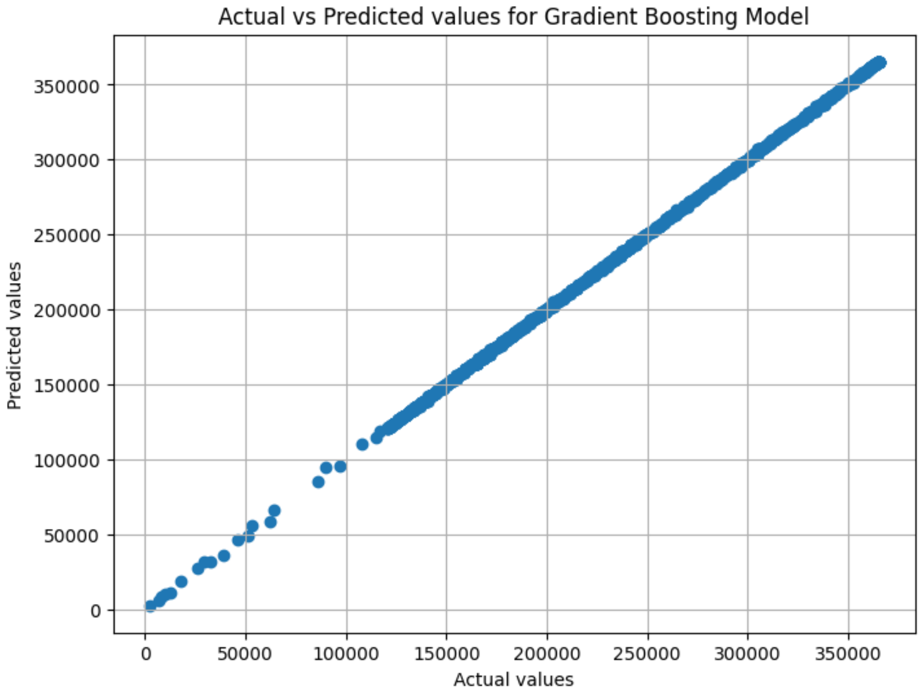

Figure 26: Actual and Predicted Values for Gradient Boost Machine Learning Model

Figure 26 shows a scattered diagram of the target variable's actual values and predicted values for the *gradient boosting model* from input features. The actual values are on the horizontal axis, while the predicted values are on the vertical axis, with a maximum value of 350000. All the data points from the plot are data collection observations from the modeling process. A blue straight line from the origin with few data point-to-line spacing fits well. Generally, the best-fit slope shows how closely variables are related. In this case, positive slopes indicate a positive correlation, implying that if one variable increases, the other also rises. In contrast, negative slopes suggest a negative correlation between variables, meaning one increases and the other decreases. Moreover, unrelated variables (*variables that have no relationship*) have zero correlation or slope. From the plot, the line of best-fit slopes upward, indicating a good correlation between actual and predicted values. This process implies that the *gradient-boosting model* predicts the data trends. However, as seen from the earlier part of the plot, data points distant from the line of best fit indicate expected errors. This situation may be due to overfitting or underfitting, data noise, outliers, missing values, and other factors that could cause these errors.

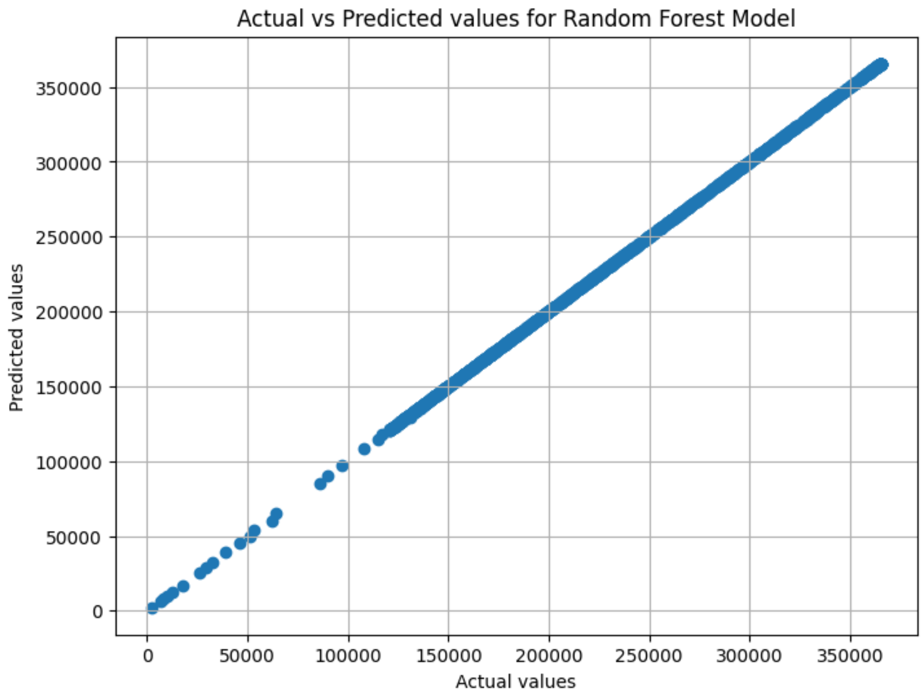

Figure 27: Actual and Predicted Values for Random Forest Machine Learning Model

Figure 27 shows the scatter diagram for the line of best fit with a positive slope, indicating a positive correlation between the actual and predicted values using the *random forest model*. The plot suggests that the model performs well in predicting the data trends. As seen from the scatter plot, specific few data points are distant from the line of best fit, indicating fewer prediction errors. Compared to the *gradient boosting model* in Figure 26, the *random forest model* performs better because the dispersed points are few and still in the range of the actual values.

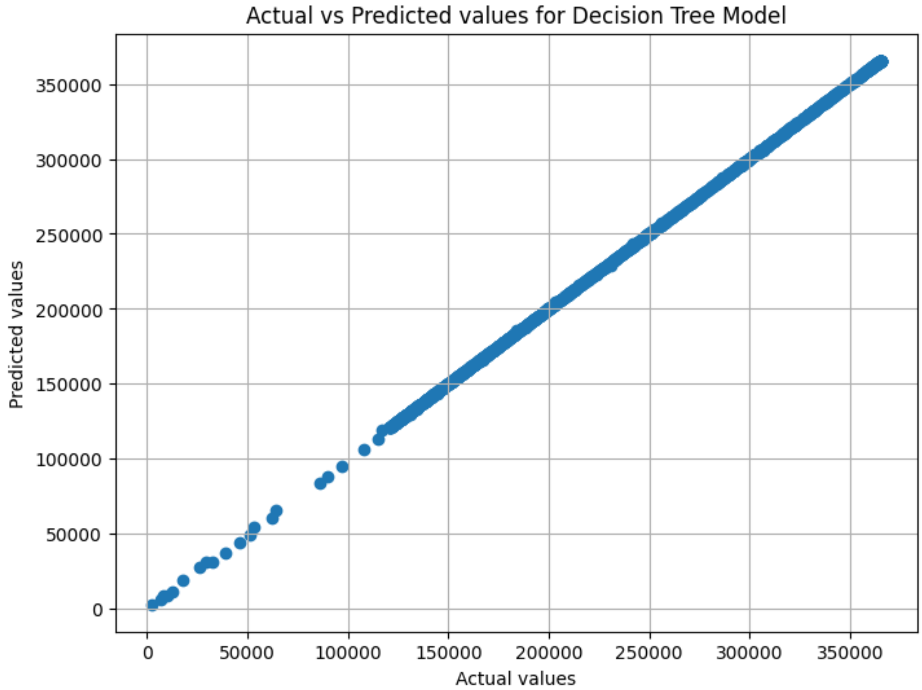

Figure 28: Actual and Predicted Values for Decision Tree Machine Learning Model

Figure 28 shows the scatter diagram of a positive correlation between the actual and predicted values using the *decision tree model*. The line of bit fit of the scatter plot implies that the model performs well in predicting the data trends. However, as seen from the plot, few data points are distant from the line of best fit in the model, indicating prediction errors. Moreover, compared to the *gradient boosting model* in Figure 26, the *decision tree model* performs better because the dispersed points are few and still in the range of the actual values. However, the *random forest* in Figure 27 performs better than the *decision tree model* because the line of best fit in the former captures more data than the latter, making it a better model for the data.

4.3.5 The Learning Curve Analysis of the Models

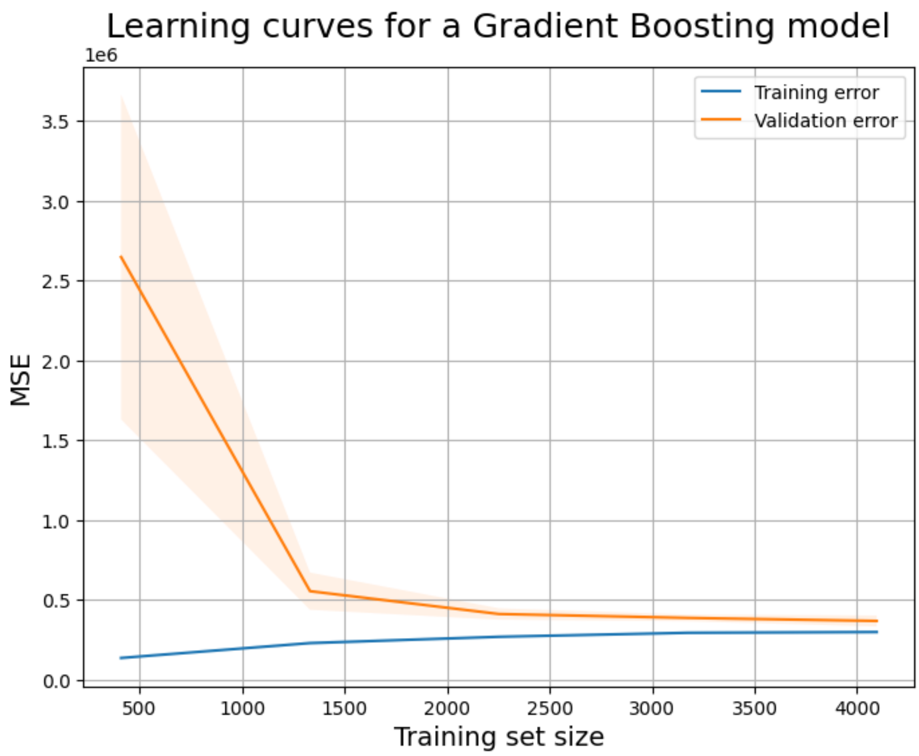

Figure 29: Gradient Boosting Learning Curve

Figure 29 is a *learning curve* that shows the prediction of a continuous regression result employing a *gradient boosting algorithm* from input features. The plot's axes are the *training set size* (number of data points required to train the model) on the horizontal axis and the *mean squared error* (MSE) on the vertical axis, which measures how well the model fits the data. Lower MSE determines the best algorithm (generally, reducing errors makes predictions accurate). The orange line from the learning curve depicts the *validation error*, which is the MSE on a separate data set not used for training data. At the same time, the blue line depicts the training error that is used for training data. The training error decreases rapidly, as seen from the plot, as the training set size increases because the model can learn more from more data. However, the validation error increases gradually as the training data increases due to the overfitting of the model, indicating that it can not adjust to new data. This situation suggests that the *gradient boosting model* needs to be more sophisticated and memorize training data instead of broad patterns. Moreover, the model may perform well on training data but could improve on unseen data, making it unsuitable for generating predictions on new data.

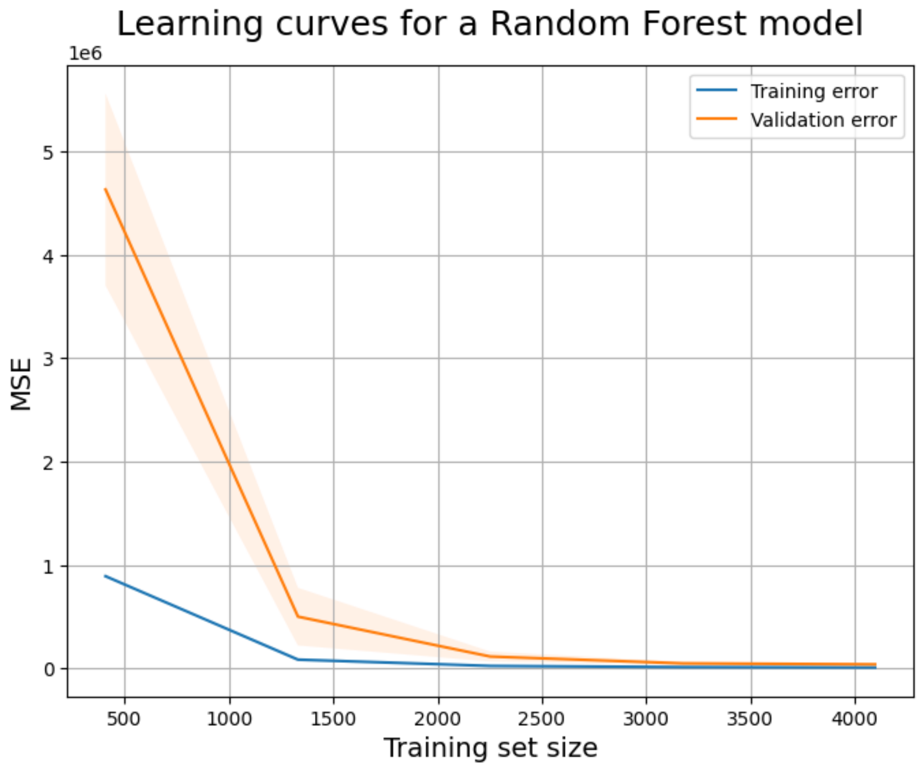

Figure 30: Random Forest Learning Curve

Figure 30 represents the *random forest learning curve*. It shows that as the training set size increases, the training and *validation error* decrease at different rates. The orange line (*validation error*) decreases more rapidly than the blue line (*training error*) because the model can learn more from more data. However, further along, the curve, the gap between the two lines remains relatively constant, indicating that the model does not overfit or underfit the data as more are introduced. Compared to the *gradient boosting model* in Figure 29, the *random forest model* performs well, has a better balance between complexity and simplicity of the modeling process, and can capture general patterns from the data.

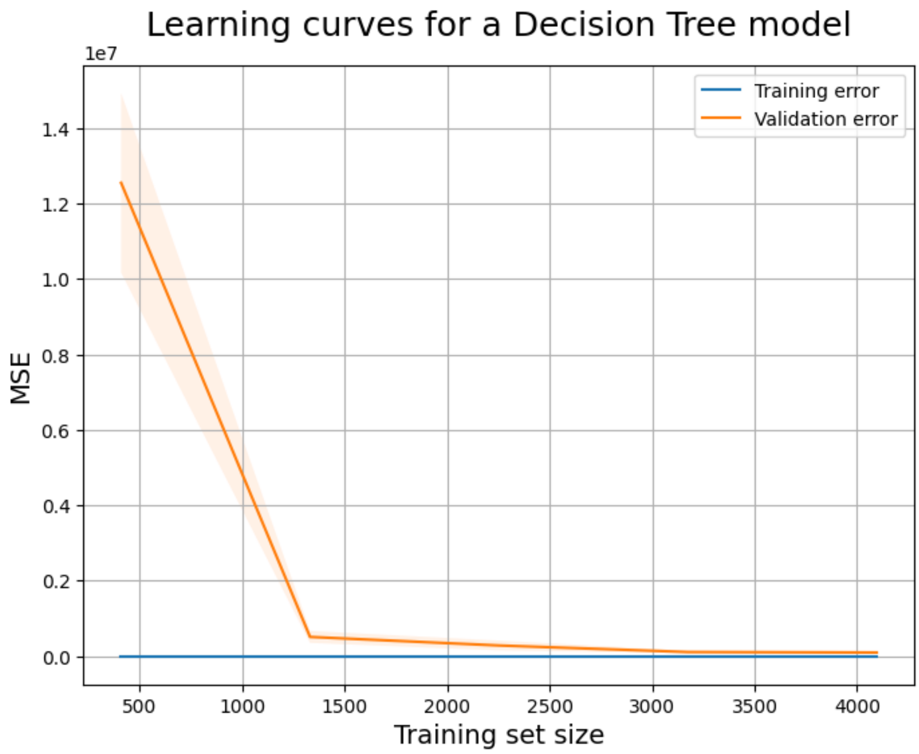

Figure 31: Decision Tree Learning Curve

Figure 31 represents the *decision tree learning curve.* As previously explained in Figures 29 and 30, as the *training set size* increases, the model learns more from more data while reducing the training error. From the plot, the training error remains constant along the origin on the horizontal axis of the curve, indicating poor performance. Moreover, the model cannot adjust to new data due to overfitting the training data resulting from the constancy of its value along the curve. Hence, the model is simplistic and cannot capture further complex data patterns. Compared to the *random forest model* in Figure 30, the *decision tree model* did not perform well, but it performs better than the *gradient boosting model* in Figure 29.

4.3.6 Evaluation Metrics Tabulation

In this research, various evaluation metrics are considered to analyze the performance of the three regression models in the machine learning process for the simulated output data in Table 13 and Figure32. From the *Mean Absolute Error (MAE)*, it was observed that the Random Forest model performed the best with the smallest error of 39.59, compared to 93.76 for the Decision

Tree and 452.18 for the Gradient Boosting model. Furthermore, based on the other metrics, the better performance of the Random Forest model was also observed when considering the *Mean Squared Error (MSE)* and the *Root Mean Squared Error (RMSE)*. The Random Forest model achieved an MSE of 18091.15 and an RMSE of 134.50, notably smaller than the corresponding values for the other two models. Also, considering the R-squared and Explained Variance metrics indicate the goodness of fit of a set of predictions. Again, the Random Forest model outperformed the others, having a value of 0.999997 (*approximately close to 1*) both in R-squared and Explained Variance, indicating that it explains the majority of the Variance in the target variable. Moreover, in the *Median Absolute Error (MedAE)*, the Random Forest model (14.83) has the smallest value, which is 14.83, suggesting it also has the lowest median error. The *Mean Squared Logarithmic Error*(MSLE) is a valuable metric for the target variable experiencing exponential growth. In this case, the Random Forest model also depicts a smallest value of 0.00001, indicating that it is the best model if the target variable experiences exponential growth.

Model	**MAE**	**MSE**	**RMSE**	**R-Squared**	**Explained Variance**	**Median AE**	**MSLE**
Decision Tree	93.76	72017.51	268.36	0.99999	0.99999	41.61	0.00011
Random Forest	39.59	18091.15	134.50	0.999997	0.999997	14.83	0.00001
Gradient Boosting	452.18	401176.45	633.38	0.99994	0.99994	325.79	0.00008

Table 13: Comparison of Regression Models

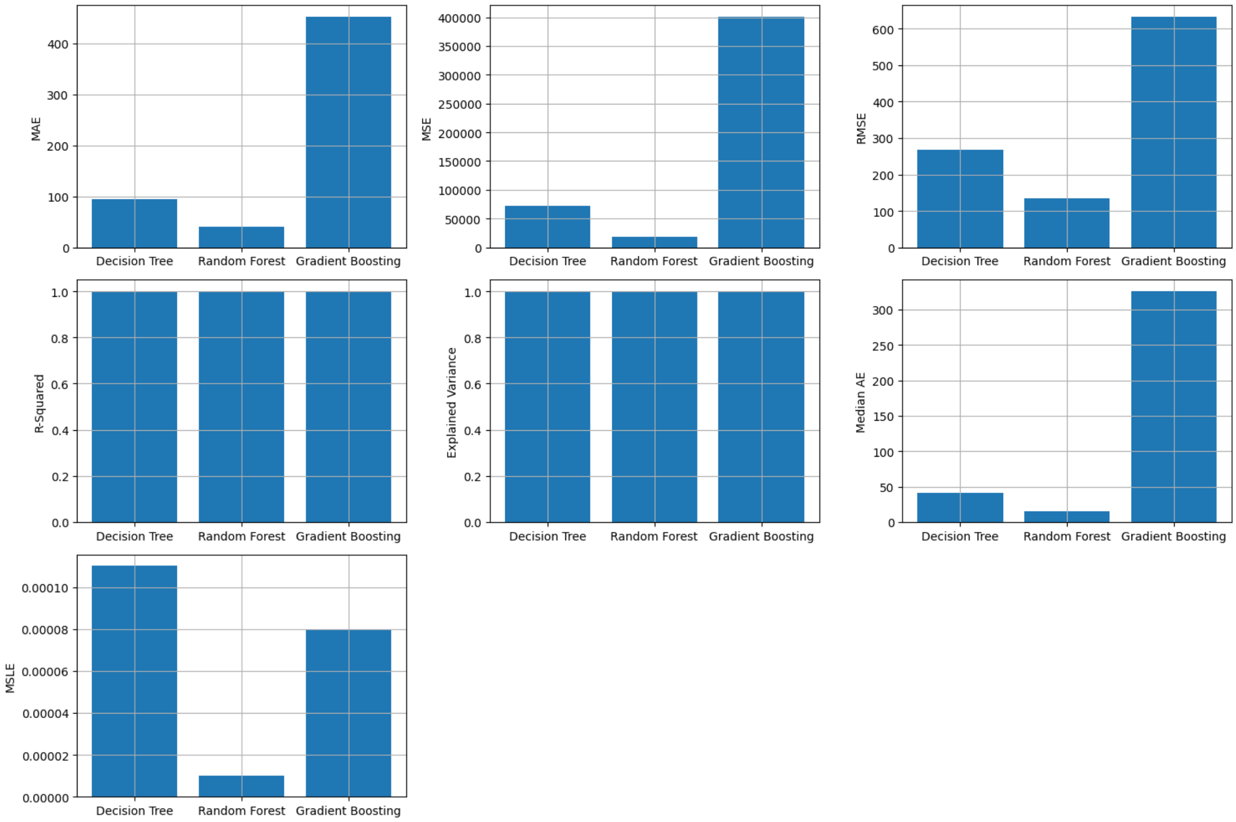

Figure 32: Evaluation Metrics of the Machine Learning Models

Figure 32 above displays the plots of each evaluation metric to present the visibility of the performance of each algorithm. This visualization process makes it easy to understand how the algorithms are plotted against the values of the metrics. From the analysis of the evaluation metrics, the *Random Forest model* consistently outperformed the *Decision Tree and Gradient Boosting models* across all metrics, making it the most suitable choice for predicting the '*Minority Population with Two or More Chronic Diseases*' using the given other variables as features in the dataset.

5 Conclusions

Addressing chronic diseases in minority populations in the US is challenging due to the complex and dynamic nature of the factors surrounding the onsets of these diseases. Chronic disease incidence and prevalence in the US have accounted for the highest allocation of medical funds, amounting to trillions of dollars annually. These diseases often lead to prolonged acute health conditions, resulting in adverse health outcomes such as disability and death. The burden of these conditions disproportionately affects people of color the most due to the nature of their built environment and socioeconomic status. The incidence of chronic diseases is usually associated with and influenced by the interaction of social determinants of health and risk factors. Despite continuous spending and efforts to address these health challenges, health and healthcare disparities still exist in the US, primarily affecting minority populations (*people of color*). Moreover, a gap in understanding between the various demographic groups persists. As a result, some low-income Americans and racial minorities have a much higher incidence of diseases. The various treatment options, healthcare interventions, and access to healthcare therapy are still severely restricted in these regions. In these minority-populated regions, since the interplays of SDOH and chronic disease risk factors are complex and dynamic and significantly influence the incidence of prolonged ailments, there is a need to address these health outcome challenges. Therefore, this study addresses these issues from the root cause by employing system dynamics modeling approach to explore the individual and collective interacting factors in these minority populations. Afterward, machine learning techniques were utilized to predict the possible health outcomes to develop proactive measures and interventions in these regions. The system dynamics model projected the trends of the influence of the interacting factors. Furthermore, in the machine learning process, the Random Forest model outperformed the decision tree, gradient boosting models with the lowest evaluating metrics scores with MAE of 39.59 and MSE of 18091.15, R-squared 0.999997, Median AE of 14.83, and 0.00001, offering precision and predictive prowess. These combined techniques provide a holistic view of chronic disease risk factors and SDOH interventions and how they impact health outcomes for preparing for possible interventions and policy considerations in minority populations.

5.1 Research Limitations

This research contributes uniquely to the body of knowledge in public health, health and healthcare disparities, chronic disease management, and health interventions by considering the complexity and dynamics involved in the surrounding factors adversely influencing the health outcomes of minority and underserved communities. Despite this novel contribution, several limitations must be acknowledged, such as *data availability and quality, model complexity, model generalizations, impact interventions, and subjectivity in system dynamics modeling.*

Moreover, by identifying these limitations, the research emphasizes the contextual and provisional nature of the findings. The future direction of the research will address some of these limitations and still advance in understanding the complex and dynamic nature of these interacting factors in minority populations.

5.2 Future Work

This research has presented a novel multifaceted technique, which integrates system dynamics modeling and machine learning for understanding the dynamic and complex interactions of chronic disease risk factors and SDOH and addressing the influence of these interactions on chronic disease incidence, prevalence, and health outcomes in minority populations. While the study has contributed valuable insights, several promising avenues for future work have been identified, such as *expansion to diverse populations, integration of additional variables, ensuring quality data availability, collaboration with stakeholders, ethical considerations, interdisciplinary collaborations, and policy impact assessment.*

References

[1] Holman, Halsted. "The Relation of the Chronic Disease Epidemic to the Health Care Crisis." *ACR Open Rheumatology*, vol. 2, no. 3, 2020, pp. 167-173, https://doi.org/10.1002/acr2.11114.

[2] Baciu, A., Negussie, Y., Geller, A., & Weinstein, J. N. (2019). The Root Causes of Health Inequity. Nih.gov; National Academies Press (US). https://www.ncbi.nlm.nih.gov/books/NBK425845/

[3] Daniel, Hilary, et al. "Addressing Social Determinants to Improve Patient Care and Promote Health Equity: An American College of Physicians Position Paper." *Annals of Internal Medicine*, vol. 168, no. 8, 2018, p. 577, https://doi.org/10.7326/m17-2441.

[4] Kwame, A., & Petrucka, P. M. (2021). A literature-based study of patient-centered care and communication in nurse-patient interactions: Barriers, facilitators, and the way forward. *BMC Nursing*, 20(1), 1–10. https://doi.org/10.1186/s12912-021-00684-2.

[5] Smedley, B. D., Stith, A. Y., & Nelson, A. R. (2003). *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington (DC): National Academies Press (US).

[6] Crook, E. D., & Peters, M. (2008). Health disparities in chronic diseases: where the money is. *American Journal of Medical Sciences*, 335, 266-270. doi: 10.1097/maj.0b013e31816902f1.

[7] Price, J. H., Khubchandani, J., McKinney, M., & Braun, R. (2013). Racial/ethnic disparities in chronic diseases of youths and access to health care in the United States. *BioMed Research International*, 2013, 787616. doi: 10.1155/2013/787616.

[8] Centers for Disease Control and Prevention. (2022). Health and Economic Costs of Chronic Diseases | CDC. Retrieved from https://www.cdc.gov/chronicdisease/about/costs/index.htm.

[9] Peter G. Peterson Foundation. (2022, July 19). How does the U.S. healthcare system compare to other countries? *Www.pgpf.org; Peter G. Peterson Foundation*. Retrieved from https://www.pgpf.org/blog/2022/07/how-does-the-us-healthcare-system-compare-to-other-countries

[10] Baciu, A., Negussie, Y., Geller, A., Weinstein, J. N. (2018). The State of Health Disparities in the United States. *Nih.gov; National Academies Press (US).* Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK425844/

[11] Galvani, A., Bauch, C. T., Singer, B. H., & Fitzpatrick, M. C. (2020). Improving the Prognosis of Healthcare in the United States. *Lancet (London, England)*, 395(10223), 524. https://doi.org/10.1016/S0140-6736(19)33019-3.

[12] Halle, M., Lewis, C. B., & Seshamani, M. Health Disparities: A Case for Closing the Gap. Retrieved from http://healthreform.gov/.

[13] Hill, L., Ndugga, N., & Artiga, S. (2023, March 15). Key data on health and health care by race and ethnicity. KFF. https://www.kff.org/racial-equity-and-health-policy/report/key-data-on-health-and-health-care-by-race-and-ethnicity/

[14] World Health Organization. (2021). Social determinants of health. World Health Organisation. https://www.who.int/health-topics/social-determinants-of-health

[15] Healthy People 2030. (2023). Healthy People 2030. Health.Gov. https://health.gov/healthypeople

[16] CDC. (2020, October 23). CDC - Social Determinants of Health - STLT Gateway. *www.cdc.gov.* https://www.cdc.gov/publichealthgateway/sdoh/index.html

[17] American Psychological Association. "Ethnic and Racial Minorities & Socioeconomic Status." American Psychological Association, 2017, www.apa.org/pi/ses/resources/publications/minorities.

[18] Healthy People 2030. "Neighborhood and Built Environment." Health.gov, 2021, health.gov/healthypeople/objectives-and-data/browse-objectives/neighborhood-and-built-environment.

[19] Tulane University. "Social Determinant of Health: Education Is Crucial." Publichealth.tulane.edu, 27 Jan. 2021, publichealth.tulane.edu/blog/social-determinant-of-health-education-is-crucial/.

[20] Moloughney, BW. "Social Determinants of Health: What Can Public Health Do to Address Inequities in Infectious Disease?" *Canada Communicable Disease Report*, vol. 42, no. Suppl 1, 2016, p. S1-14, https://doi.org/10.14745/ccdr.v42is1a03.

[21] Gelman, Andrew, et al. "An Analysis of the New York City Police Department's 'Stop-And-Frisk' Policy in the Context of Claims of Racial Bias." *Journal of the American Statistical Association*, vol. 102, no. 479, Sept. 2007, pp. 813–823, https://doi.org/10.1198/016214506000001040.

[22] Behera, Basanta, and Prasad, Ram. "Primary Health-Care Goal and Principles." *Healthcare Strategies and Planning for Social Inclusion and Development*, 2021, pp. 221-239, https://doi.org/10.1016/B978-0-323-90446-9.00008-3.

[23] McWilliams, J. "Health Consequences of Uninsurance among Adults in the United States: Recent Evidence and Implications." *The Milbank Quarterly*, vol. 87, no. 2, 2009, pp. 443-494, https://doi.org/10.1111/j.1468-0009.2009.00564.x.

[24] Coughlin, Steven, et al. "Health Literacy, Social Determinants of Health, and Disease Prevention and Control." *Journal of Environment and Health Sciences*, vol. 6, no. 1, 2019, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7889072/.

[25] "What Are the Social Determinants of Health?" United Way Halifax, 25 Nov. 2021, www.unitedwayhalifax.ca/blog/what-are-the-social-determinants-of-health/.

[26] Bharmal, Nazleen, et al. "Understanding the upstream social determinants of health." California: RAND (2015): 1-18.

[27] Cockerham, William, et al. "The Social Determinants of Chronic Disease." *American Journal of Preventive Medicine*, vol. 52, no. 1, 2017, pp. S5-S12, https://doi.org/10.1016/j.amepre.2016.09.010.

[28] Federico, Monica, et al. "The Impact of Social Determinants of Health on Children with Asthma." *The Journal of Allergy and Clinical Immunology: In Practice*, vol. 8, no. 6, 2020, pp. 1808-1814, https://doi.org/10.1016/j.jaip.2020.03.028.

[29] Raphael, Dennis. "A Discourse Analysis of the Social Determinants of Health." *Critical Public Health*, vol. 21, no. 2, June 2011, pp. 221-236, https://doi.org/10.1080/09581596.2010.485606.

[30] Singu, Sravani, et al. "Impact of Social Determinants of Health on the Emerging COVID-19 Pandemic in the United States." *Frontiers in Public*

Health, vol. 8, no. 406, July 2020, https://doi.org/10.3389/fpubh.2020.00406.

[31] Asare, Matthew, et al. "Social Determinants of Health: A Framework for Studying Cancer Health Disparities and Minority Participation in Research." *Oncology Nursing Forum*, vol. 44, no. 1, 2017, p. 20, https://doi.org/10.1188/17.ONF.20-23.

[32] Alcendor, Donald. "Racial Disparities-Associated COVID-19 Mortality among Minority Populations in the US." *Journal of Clinical Medicine*, vol. 9, no. 8, 2020, p. 2442, https://doi.org/10.3390/jcm9082442.

[33] Russo, R. G., et al. "COVID-19, Social Determinants of Health, and Opportunities for Preventing Cardiovascular Disease: A Conceptual Framework." *Journal of the American Heart Association*, vol. 10, no. 24, 2021, https://doi.org/10.1161/jaha.121.022721.

[34] Fraiman, Yarden S., et al. "The Influence of Social Determinants of Health on the Genetic Diagnostic Odyssey: Who Remains Undiagnosed, Why, and to What Effect?" *Pediatric Research*, vol. 89, no. 2, Jan. 2021, pp. 295-300, https://doi.org/10.1038/s41390-020-01151-5.

[35] Lantz, Paula M., et al. "Socioeconomic Factors, Health Behaviors, and Mortality." *JAMA*, vol. 279, no. 21, June 1998, p. 1703, https://doi.org/10.1001/jama.279.21.1703.

[36] Clarke, Janice, et al. "An Innovative Approach to Health Care Delivery for Patients with Chronic Conditions." *Population Health Management*, vol. 20, no. 1, 2017, pp. 23-30, https://doi.org/10.1089/pop.2016.0076.

[37] HealthITAnalytics. "Social Determinants Strongly Linked to Onset of Chronic Disease." *HealthITAnalytics*, 23 May 2019, https://healthitanalytics.com/news/social-determinants-strongly-linked-to-onset-of-chronic-disease.

[38] Brakefield, W., Olusanya, O., White, B., & Shaban-Nejad, A. (2023). Social Determinants and Indicators of COVID-19 Among Marginalized Communities: A Scientific Review and Call to Action for Pandemic Response and Recovery. *Disaster Medicine and Public Health Preparedness*, 17, E193. doi:10.1017/dmp.2022.104

[39] Emeny, Rebecca, et al. "Inclusion of Social and Structural Determinants of Health to Advance Understanding of Their Influence on the Biology

of Chronic Disease." *Current Protocols*, vol. 2, no. 10, 2022, p. e556, https://doi.org/10.1002/cpz1.556.

[40] Gallant, Mary P., et al. "Chronic Illness Self-Care and the Family Lives of Older Adults: A Synthetic Review across Four Ethnic Groups." *Journal of Cross-Cultural Gerontology*, vol. 25, no. 1, Feb. 2010, pp. 21-43, https://doi.org/10.1007/s10823-010-9112-z.

[41] Battineni, Gopi, et al. "Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis." *Journal of Personalized Medicine*, vol. 10, no. 2, 2020, p. 21, https://doi.org/10.3390/jpm10020021.

[42] Nusinovici, Simon, et al. "Logistic Regression Was as Good as Machine Learning for Predicting Major Chronic Diseases." *Journal of Clinical Epidemiology*, vol. 122, 2020, pp. 56-69, https://doi.org/10.1016/j.jclinepi.2020.03.002.

[43] Chien Hsiang Chang, et al. "A Health Care Platform Design: Applying Novel Machine Learning Methods to Predict Chronic Cardiac Disease." Sept. 2020, https://doi.org/10.21606/drs.2020.351.

[44] Alanazi, Rayan. "Identification and Prediction of Chronic Diseases Using Machine Learning Approach." *Journal of Healthcare Engineering*, vol. 2022, Feb. 2022, p. e2826127, https://doi.org/10.1155/2022/2826127.

[45] Khalid, Hira, et al. "Machine Learning Hybrid Model for the Prediction of Chronic Kidney Disease." *Computational Intelligence and Neuroscience*, edited by Farman Ali, vol. 2023, Mar. 2023, pp. 1-14, https://doi.org/10.1155/2023/9266889.

[46] Chen, Min, et al. "Disease Prediction by Machine Learning over Big Data from Healthcare Communities." *IEEE Access*, vol. 5, 2017, pp. 8869-8879, https://doi.org/10.1109/access.2017.2694446.

[47] Yinusa, Ahmeed, et al. "A Study on CKD Progression and Health Disparities Using System Dynamics Modeling." *Healthcare*, vol. 10, no. 9, 2022, p. 1628, https://doi.org/10.3390/healthcare10091628.

[48] Loyo, Heather Karina, et al. "From Model to Action." *Health Promotion Practice*, vol. 14, no. 1, Apr. 2012, pp. 53-61, https://doi.org/10.1177/1524839910390305.

[49] Ciplak, N., & Barton, J. R. (2012). A system dynamics approach for healthcare waste management: a case study in Istanbul Metropolitan City, Turkey. *Waste Management & Research*, 30(6), 576-586. https://doi.org/10.1177/0734242x12443405.

[50] "Stella Architect." *www.iseesystems.com.*, https://www.iseesystems.com/store/products/stella-architect.aspx.

[51] "Software – Ventana Systems." *vensim.com.*, Retrieved July 25, 2023, from https://vensim.com/software/.

[52] Martin-Martinez, E., et al. "PySD: System Dynamics Modeling in Python." *Journal of Open Source Software*, vol. 7, no. 78, 2022, p. 4329, https://doi.org/10.21105/joss.04329.

[53] Adane, Tigist Fetene, et al. "Application of System Dynamics for Analysis of Performance of Manufacturing Systems." *Journal of Manufacturing Systems*, vol. 53, 2019, pp. 212-233, https://doi.org/10.1016/j.jmsy.2019.10.004.

[54] United States Census Bureau. (2022). *U.S. Census Bureau quickfacts: United States.* https://www.census.gov/quickfacts/fact/table/US/PST045222.

[55] McKinney, W. (2010). Data structures for statistical computing in Python. In Proceedings of the 9th Python in Science Conference (Vol. 445, pp. 51-56).

[56] Matplotlib. (2012). *Matplotlib: Python plotting — Matplotlib 3.1.1 documentation.* https://matplotlib.org/.

[57] seaborn. (2012). *seaborn: statistical data visualization — seaborn 0.9.0 documentation.* https://seaborn.pydata.org/.

[58] Airhihenbuwa, C. O., et al. (2021). *Global Perspectives on Improving Chronic Disease Prevention and Management in Diverse Settings. Preventing Chronic Disease*, 18. https://doi.org/10.5888/pcd18.210055.

[59] Clarke, G. M., et al. "Evaluating the Impact of Healthcare Interventions Using Routine Data." *BMJ*, vol. 365, June 2019, p. l2239. https://doi.org/10.1136/bmj.l2239.

[60] "Disease Management | AMCP.org." *Amcp.org*, 18 July 2019. https://www.amcp.org/about/managed-care-pharmacy-101/concepts-managed-care-pharmacy/disease-management.

[61] "Multi-Component Obesity Prevention Interventions." *County Health Rankings & Roadmaps.* https://www.countyhealthrankings.org/take-action-to-improve-health/what-works-for-health/strategies/multi-component-obesity-prevention-interventions.

[62] Breiman, L. "Random Forests." *Machine Learning*, vol. 45, no. 1, 2001, pp. 5–32. https://doi.org/10.1023/a:1010933404324.

[63] Quinlan, J. R. "Induction of Decision Trees." *Machine Learning*, vol. 1, no. 1, 1986, pp. 81–106. https://doi.org/10.1007/bf00116251.

[64] Friedman, J. H. "Greedy Function Approximation: A Gradient Boosting Machine." *The Annals of Statistics*, vol. 29, no. 5, 2001, pp. 1189–1232. https://www.jstor.org/stable/2699986.

[65] Richardson, G. P., & Pugh, A. L. *Introduction to System Dynamics Modeling with DYNAMO.* MIT Press, 1981.