THE ASSOCIATION AMONG ENVIRONMENTAL FACTORS, ECONOMIC FACTORS AND NCDs: IMPLICATIONS FOR THAILAND'S ECONOMY

BY

WIMONSIRI KACHANTORN

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER DEGREE OF ECONOMICS
FACULTY OF ECONOMICS
THAMMASAT UNIVERSITY
ACADEMIC YEAR 2023

THAMMASAT UNIVERSITY
FACULTY OF ECONOMICS

THESIS

BY

WIMONSIRI KACHANTORN

ENTITLED

THE ASSOCIATION AMONG ENVIRONMENTAL FACTORS, ECONOMIC FACTORS AND NCDs: IMPLICATIONS FOR THAILAND'S ECONOMY.

was approved as partial fulfillment of the requirements for
the degree of Master Degree of Economics

on August 9, 2024

Chairman Theepakorn Jithitikulchai
(Assistant Professor Theepakorn Jithitikulchai ,Ph.D.)

Member and Advisor Piyawong Punjatewakupt
(Assistant Professor Piyawong Punjatewakupt, Ph.D.)

Member Fei Gu
(Fei Gu, Ph.D)

Dean
(Assistant Professor Supachai Srisuchart, Ph.D)

Thesis Title	THE ASSOCIATION AMONG ENVIRONMENTAL FACTORS, ECONOMIC FACTORS AND NCDs: IMPLICATIONS FOR THAILAND'S ECONOMY
Author	Wimonsiri Kachantorn
Degree	Master Degree of Economics
Major Field/Faculty/University	Economics Faculty of Economics Thammasat University
Thesis Advisor	Assistant Professor Piyawong Punjatewakupt, Ph.D.
Academic Year	2023

ABSTRACT

Non-communicable diseases (NCDs) represent a critical public health and economic challenge in Thailand. This research investigates the impact of environmental and economic factors on NCDs mortality including diabetes, hypertension, ischemic heart disease, and chronic respiratory diseases. across Thai provinces between 2018 and 2021. By incorporating machine learning and spatial analysis, the study analyzes a range of satellite-derived environmental data, along with economic indicators.

The application of machine learning with Random Forest Model provides a high-accuracy prediction of NCDs mortality rates, achieving predictive accuracies between 98% and 99%. The analysis reveals that population density, urbanization levels, economic conditions, land use patterns, and climatic factors are significant predictors of NCDs mortality. Spatial analysis further uncovers distinct spatial patterns and clusters of NCDs mortality, emphasizing the influence of geographical location and neighboring effects.

These findings highlight the intricate linkages between environmental and economic determinants and their role in shaping NCDs mortality trends in Thailand. The results underscore the necessity for targeted public health policies and interventions that address specific environmental and economic risk factors, with the goal of reducing the NCDs burden and enhancing public health outcomes.

Keywords: Non-Communicable Diseases, Thailand, Spatial Analysis, Machine Learning, Environment, Economy

ACKNOWLEDGEMENTS

This thesis has been completed with the support of many individuals, to whom I wish to express my deepest gratitude.

First, I extend my heartfelt thanks to my advisor, Professor Piyawong Punjatewakupt. His unwavering patience and guidance throughout my study have been invaluable. His insightful comments and suggestions, from the inception of the topic to the completion of this work, along with his advice on developing effective analysis results, have significantly enhanced this thesis.

Second, I am profoundly grateful to the members of my thesis committee. Assistant Professor Theepakorn Jithitikulchai, whose research on this topic provided a foundational basis for my work, offered crucial suggestions on the health economics framework. Together with Assistant Professor Fei Gu, a specialist in econometrics, both provided essential suggestion that improved the quality of work.

I also wish to express my deepest appreciation to my parents for their patience and unwavering support throughout my long academic journey. Their encouragement has been a constant source of strength. In addition to the aforementioned individuals, I would like to acknowledge the many others who have contributed to this thesis in various ways. Although they are not mentioned by name, their support has been deeply appreciated.

I hope that this study fulfills its objectives and serves as a valuable contribution to the field, aiding both policymakers and researchers in their future endeavors. Any remaining errors or omissions are entirely my own responsibility.

Wimonsiri Kachantorn

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

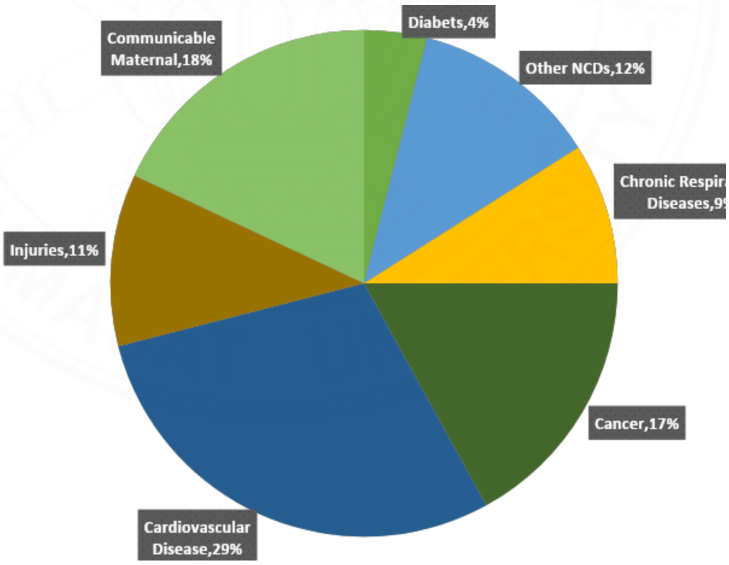
CHAPTER 1
INTRODUCTION

1.1 Motivation

Non-communicable diseases (NCDs) are significant threats to public health in Thailand, encompassing ailments like cardiovascular diseases, cancer, chronic respiratory diseases, and diabetes. The prevalence of NCDs has been increasing in Thailand due to a variety of factors, including urbanization, changes in dietary habits, and sedentary lifestyles (UNDP Thailand, 2021). The impact of NCDs on public health in Thailand and many other countries includes increased healthcare costs, reduced productivity, and a significant burden on the healthcare system (Office of Health Promotion, Department of Health, 2020).

Figure 1.1

Proportional mortality as percent (%) of total deaths in year 2020



Note: The author visualized data from World Health Organization- Noncommunicable Diseases country profiles.

Figure 1.1 provides an analysis of proportional mortality as a percentage of total deaths in Thailand for the year 2020. The pie chart highlights the significant impact of cardiovascular diseases, cancer, and other NCDs on overall mortality in Thailand based on the World Health Organization's Noncommunicable Diseases country profiles.

The dominant share of cardiovascular diseases (29%) as the leading cause of mortality indicates a significant burden on the healthcare system. The substantial mortality due to cancer (17%) and chronic respiratory diseases (9%) The relatively lower percentage of deaths due to diabetes (4%) compared to cardiovascular diseases and cancer suggests that while diabetes is a significant health concern. The category of "other NCDs" (12%) and injuries (11%) points to diverse health challenges that need a broad spectrum of healthcare services and interventions. The high percentage of deaths due to communicable maternal conditions (18%) indicates ongoing issues with maternal health and infectious diseases, requiring sustained efforts in maternal care and infectious disease control.

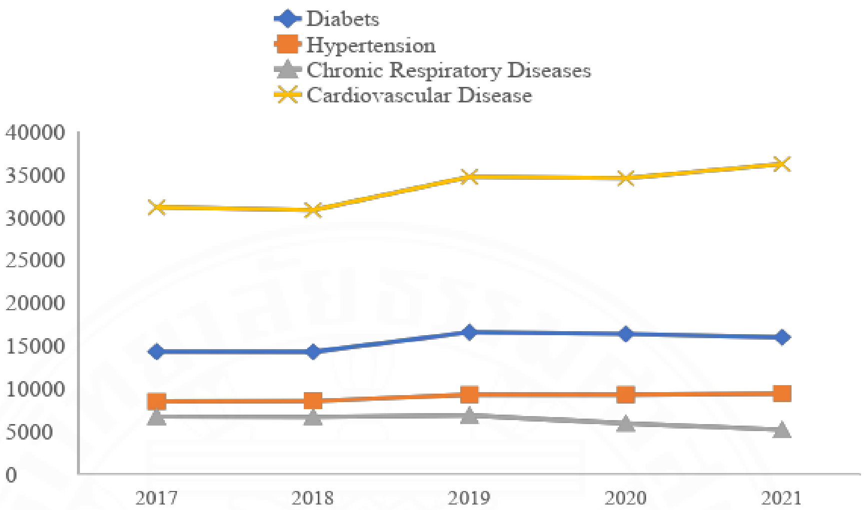
Overall, the data imply that NCDs, particularly cardiovascular diseases and cancer, are major health challenges in Thailand, necessitating focused healthcare strategies and resource allocation.

In 2020 Thai citizens have a 14% risk of dying prematurely (before the age of 70) from one of the four main NCDs (cardiovascular disease (CVD), diabetes, chronic respiratory disease, and cancer), with an 18% probability for men and 11% for women (WHO and UNDP Thailand, 2021).

Figure 1.2 which presents the number and mortality rates of four major non-communicable diseases (NCDs) per 100,000 people in Thailand from 2017 to 2021, offers several important implications regarding public health trends and challenges.

Figure 1.2

Number and mortality rate with 4 NCDs per 100,000 thousand people, Thailand, 2017-2021



Note: The author visualized data from the Strategy and Planning Division, Office of the Permanent Secretary, Ministry of Public Health.

The slight increase in the mortality rate for cardiovascular disease suggests a growing burden of this condition over the observed period. This trend may imply that risk factors associated with cardiovascular disease, such as obesity, hypertension, and unhealthy lifestyles, are becoming more prevalent or that the healthcare system faces challenges in effectively managing and treating this condition. The relatively stable mortality rates for diabetes and hypertension indicate that while these conditions remain significant public health concerns, their impact on mortality has not worsened. The slight decrease in the mortality rate for chronic respiratory diseases may reflect improvements in healthcare access, better management of respiratory conditions, or successful public health initiatives targeting risk factors such as smoking and air pollution.

The worldwide cost of NCDs is estimated between 2011 and 2030, it will be 46.7 trillion US dollars, or approximately 1,401 trillion Thai baht. NCDs have an economic impact not just on medical costs but also on lost population output. Production loss is indicated by the fact that 25% of the world's population has NCDs, which cause them to die before reaching age 60. (Office of Health Promotion Policy Research, International Health Policy Development Agency, 2014)

In addition to their significant financial expenses, NCDs have an adverse effect on individuals, and the government by reducing macroeconomic productivity through limited full labor force participation. The labor that people would have created in their remaining working years is lost when they pass away too soon or become too sick to work full-time. In addition, people with illnesses are more prone to miss work (absenteeism) or perform less well than they could at work (presenteeism). Chen, Simiao, et al. (2018) assessed the impact of NCDs on the United States' productivity, revealing a total loss of 94.9 trillion US dollars due to these diseases. The study proposed a dynamic production function model that posits three pathways through which NCDs affect the economy: direct mortality of working-age individuals reduces aggregate output; illness decreases productivity or leads to reduced working hours; and resources divert towards medical treatment and prevention. Addressing NCDs' economic implications is critical for mitigating their negative effects on individuals, society, and the national economy. Reduced illness correlates with higher productivity and economic growth at regional and national levels (Bukhman et al., 2015).

In the case of Thailand, the Ministry of Public Health of Thailand has shown that the annual economic losses from NCDs are estimated to be 1.6 trillion Thai baht (473 billion US dollars), or 9.7 % of GDP. These costs include 139 billion Thai baht to treat NCDs and 1.5 trillion Thai baht in lost productive capacity due to absenteeism, presenteeism or early withdrawal from the labor force due to premature death or disability (UNDP Thailand, 2021). The productivity losses due to NCDs account for 91% of all NCD-related costs.

The association between environmental factors, economic conditions, and the prevalence of NCDs are crucial for developing effective public health strategies and economic policies. By identifying and addressing the environmental and economic determinants of NCDs, policymakers can mitigate the health and economic burden of these diseases.

1.2 Statement of the problem

NCDs thus negatively affect socioeconomic development and the long-term fiscal sustainability of government and public services. Long-term management of NCDs system leads to increase healthcare expenditure. The costs associated with diagnosis, treatment, medications, hospitalizations, and rehabilitation for NCDs can place a substantial burden on individuals, families, and the healthcare system. The rising incidence of NCDs poses a significant challenge to the healthcare system, leading to increase medical expenses and demand for healthcare services. This trend risks overwhelming the healthcare budget, making it essential to address the root causes of NCDs.

Due to the enormous burden, many countries have made efforts to research the causes of NCDs occurring in the area. In addition, it was found that Environmental factors, including air pollution and climate change, significantly impact the incidence and mortality of NCDs. Environmental factors disproportionately affect vulnerable populations, altering their dietary habits and increasing the risk of diet-related NCDs. Moreover, geographic clusters with high NCDs mortality underscore the complex interaction between environmental and socioeconomic factors. (Savage et al.,2021).

Spatial analysis plays a crucial method for examining the distribution of NCDs by integrating diverse data sources. This methodology provides a comprehensive framework for analyzing how risk of the factors influences the prevalence and NCDs across different regions. By mapping and analyzing, spatial analysis provides valuable insights into the complex association among environmental and economic factor and their impact on health outcomes, thereby advancing our understanding of disease etiology and informing public health interventions (Pengpid & Peltzer, 2019).

The study aims to investigate the dependent factor that influences the prevalence of NCD mortality. The study uses machine learning and spatial econometric model to Figure out the pattern into the NCDs area density through spatial association, which changes how the effects work in different areas. It focuses on environmental and economic factors from satellite data. Understanding these dynamics is essential for shaping public health policies and fostering economic development. Furthermore, the study seeks to evaluate the economic losses associated with NCDs at the provincial level and assess their impact on Thailand's economy by province. By quantifying the economic burden of NCDs, this research aims to highlight the significant financial implications for local economies and provide a basis for targeted interventions to mitigate these losses. Understanding the economic impact of NCDs is crucial for developing strategies to reduce their prevalence and improve the overall health and economic well-being of the population.

1.3 Objective of the study

1) To study the environmental and economic factors influencing NCDs cases in Thailand by province.
2) To study the spatial among among environmental and economic factors and NCDs cases by province.
3) To evaluate economic losses at provincial level and study impact on Thailand's economy by province.

1.4 Scope of the study

This study focuses on the provincial environmental and economic factors and NCDs cases include diabetes, hypertension, ischemic and chronic in 76 provinces in the period 2018 to 2020 by applying satellite data and the prevalence and mortality rates of NCDs diseases from the Office of the Permanent Secretary, Ministry of Public Health Thailand. Data sources include satellite data for environmental variables such as land surface temperature, nighttime light density, and cropland area. The study considers economic factors, particularly Gross Provincial Product (GPP) and socioeconomic status are also analyzed.

1.5 The organization of the study

The study's organizational structure is designed for a clear exploration of the research problem, spread across five chapters. The first chapter introduces the study with an overview of the motivation, problem statement, research question, objectives, and scope. Chapter two reviews existing literature on NCDs and spatial analysis. Chapter three details the research methodology, including data collection, analysis, and spatial model. Chapter four presents and discusses the empirical findings on the relationship between environmental and economic factors, NCDs, and their economic impacts. Chapter five concludes the study by summarizing key findings, discussing limitations, and offering recommendations for future research.

CHAPTER 2
REVIEW OF LITERATURE

This chapter illustrates an overview of the review literature related to this study. The study begins with the association between NCDs and relevant factors. First, with a particular emphasis on environmental. Second, emphasis on economic factors. The Third, emphasize the dependent factor on NCDs mortality and prevalence rates, the study reviews machine learning to investigate the NCDs's area density through spatial association, which affects economy in different ways.

2.1 The association between economic and NCDs

The individual health might impact economic growth (Bloom & Canning, 2000; Jack & Lewis, 2009). Through its effect on the accumulation of human capital. This is because individuals who are healthier and in better health may be able to work more productively than those who are sick or poor circumstances. Consequently, those in good condition have the capacity to accumulate more income (Adda, Chandola, & Marmot, 2003; Wu, 2003).

By testing the Granger causality between wealth and health (Verma & Usmani, 2019) examined the association between economic growth and health in India. They utilized the vector autoregression (VAR) model with the gross state domestic product (GSDP) and the annual infant mortality rate (IMR) from 1985 to 2015. The study's primary conclusion is that GSDP and IMR have a bidirectional causal relationship. The study finds evidence of unilateral causation, which is either running from IMR to GSDP or from GSDP to IMR, while testing the null hypothesis for each state. The objective of establishing that kind of association is to increase the rate of economic growth in the healthcare sector. In another study, Erdil and Yetkiner (2009) investigate the relationships between real per capita GDP and real per capita expenditure on healthcare in nations with high and low incomes. Between health care expenditures and income, they identified both unidirectional and bidirectional causality. Accordingly, to bilateral causality, economic growth can also improve the health standards of the population through the purchase of healthcare (Grossman, 1972).

2.1.1 The impact of economic growth on NCDs

Premature deaths from NCDs curtail economic growth and trap populations in poverty. Families who live in poverty are more likely to be exposed to environmental and behavioral risk factors for NCDs. In developing countries, the burden of NCDs and risk factors is shifting toward the poor due to the long-term medical expenses associated with their care. For those who live in poverty or have recently escaped severe poverty, when faced with large, lifelong out-of-pocket expenses, impoverishment persists or can reoccur. (Engelgau et al., 2011)

Cohen et al. (2015) investigate the mechanisms through which economic growth impacts the epidemiology of NCDs. They develop a framework that divides the influence of economic growth on NCDs into three main effects: resource, behavior, and knowledge. Their findings reveal that these effects can be quantified through the consideration of two key elasticities. These elasticities pertain to the output and income factors influencing NCDs prevalence in populations, specifically healthcare, health-related behaviors and lifestyles, and medical knowledge. The study also highlighted that prolonged and severe economic recessions can negatively impact the epidemiology of NCDs by examining how reduced real per capita income affects future population health, particularly through decreased private and public health care resources and changes in health-related behaviors.

2.1.2 NCDs affect the economic growth

Erçelik (2018) examines the association between health expenditure and economic development (from 1980 to 2015) in Turkey by using the autoregressive distributed lag bound test model to analyze the association between healthcare expenditure (percent of GDP) and GDP per capita. The results of the association between GDP per capita, public and private healthcare spending, and investment in the long run with the coefficients of investment and total health expenditure are positive and statistically significant. Health issues, particularly those that impede economic growth, are known to be critical for maintaining economic stability. NCDs include chronic diseases (largely cardiovascular disease, diabetes cancer, and chronic respiratory disease), injuries, and mental health and nowadays NCDs are emerging health issues for developing countries. (Engelgau et al., 2011).

In 2019, NCDs were responsible for 41 million, or 71%, of the 55.4 million deaths globally. Among these, 15 million deaths occurred among individuals aged 30 to 69, with 77% of these deaths taking place in low- and middle-income countries (WHO, 2021). The economic burden of NCDs often falls on the informal economy, including individuals working in domestic settings, local markets, and the black market, which are typically excluded from economic indicators like GDP. On a macro level, reducing illness results in higher productivity and contributes to regional and national economic growth. The economic impact of NCDs extends beyond the health sector, affecting overall productivity and GDP (Bukhman G. et al., 2015).

Malkin et al. (2022) quantified the productivity loss associated with each NCDs. This was achieved by calculating the percentage increase in costs due to absenteeism and presenteeism. They then multiplied the productivity loss for each NCDs by the estimated number of affected workers and the GDP among the workforce to determine the disease-specific productivity losses. The study highlighted that NCDs significantly burden Saudi Arabia's economy, particularly through productivity losses and substantial direct medical costs.

Another study by Chen, Simiao, et al. (2018) evaluate the impact of NCDs on the productive capacity of the United States, create and calibrate a dynamic production function model. The results indicate a total loss of 94.9 trillion US dollars to all NCDs. In this framework, aggregate output is produced according to a human capital-augmented production function that NCDs influence the economy through the three pathways; firstly, the death of working-age people directly reduces aggregate output because physical capital is unable to make up part of what is lost during the production process when human capital is lost.

Secondly, if a working-age person suffers from disease but cannot get away from it, their productivity may decrease, they may work less time, or they may retire early, depending on how severe the illness is. And the last significant resource is needed for current NCDs initiatives like medical treatment and prevention. As an illustration, human capital as a health condition and economic growth are two things that are always interrelated and inseparable, as a higher stock of human capital tends to affect economic growth positively.

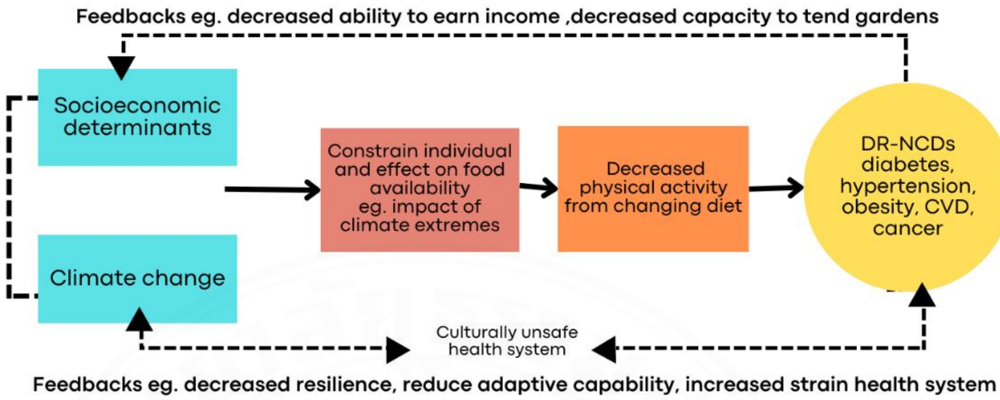
Thus, the overall health standard of a population can positively influences economic growth by enhancing human capital accumulation. Moreover, a healthier population is likely to accumulate physical capital, such as savings, more rapidly than a less healthy population, particularly within the working-age demographic. This increase in savings subsequently fuels economic growth through heightened investment (Romer, 1986; Solow, 1956). Consequently, health significantly contributes to economic growth by impacting physical capital accumulation. Improved well-being among the working-age population ultimately leads to greater productivity capacity (Isaksson, 2007).

2.2 The NCDs' environmental determinants

The significant determinants of NCDs are physical and biological environmental factors (Henke and Petropoulos 2013), including air pollution (Ross et al. 2013), climate change (Chen et al. 2010), hazards (Few et al. 2009), and water quality (Gu¨nther and Schipper 2013). Socioeconomic determinants strongly influence the distribution of risk factors, morbidity, and mortality of NCDs, and climate change disproportionately affects vulnerable populations who lack the resources to adequately respond (Savage et al., 2021). This study characterized by complexity and feedbacks in Figure 2.1.

Figure 2.1

Visual depiction of the associations between the NCDs and Climate



Note: The author visualized data from Savage et al. (2021).

The Figure shows the structural drivers of diet-related NCDs. The distribution of risk factors, morbidity, and mortality of NCDs are strongly influenced by socioeconomic determinants of health that give rise to health inequalities. The distribution of the impacts of the various structural drivers of NCDs is determined by the social determinants for instance, climate change disproportionately affects vulnerable populations who lack the resources to adequately respond (World Health Organisation, 2015). Climate change exacerbates the socioeconomic determinants leading to dietary change and increased DR-NCD risk. (Costello A et al., 2009). For example, climate change, by rendering subsistence gardening more difficult and less reliable, serves as a push factor for changing ways of life with a greater dependence on cash income and store-bought foods.

Moreover, Nkhasi, et al. (2016) identify towns with a high risk of fatalities from specific NCDs, as well as associated socioeconomic and demographic characteristics. Urban and rural areas exhibit clusters of high-risk mortality from specific NCDs. Therefore, this study will include factors related to environmental demographic characteristics.

This study utilized machine learning techniques to tackle the complexity of various factors. Machine learning proved invaluable in navigating through vast amounts of data, offering predictive model that are particularly effective in addressing intricate health issues with numerous determinants.

2.3 Distribution of study NCDs's impact

2.3.1 Machine learning and NCDs

Machine learning techniques offer a highly potent approach for constructing high-quality predictive model, especially in scenarios where the relationships between variables are intricate and cannot be adequately explained by conventional statistical methods. These techniques are particularly beneficial in investigating health issues characterized by complex interactions among numerous determinants and the outcomes of interest. They not only help uncover existing associations but also provide greater predictive accuracy compared to traditional methods. This expands the use of socio-demographic data beyond simply explaining health patterns to predicting their distribution within communities (Luo et al., 2015).

Yu (2020) also introduced machine learning to global health research by utilizing free software R program packages for geographic mapping. This approach utilizes country-specific data for population and density, demonstrating the potential of global mapping in addressing local medical and health issues of global significance. The mapping technique shows the value of global mapping in helping to understand and grab global health issues, establishing and using a global perspective in the effort to examine local medical and health issues with global significance, and in decision-making to deal with global health challenges.

The various combinations of machine learning algorithms have been employed to address the issue of missing values. Kumar et al. (2015) suggested a predictive analysis approach within the Hadoop/MapReduce framework to forecast the prevalence of specific types of diabetes, associated complications, and appropriate treatment options. Similarly, Ram et al. (2015) utilized this technique to identify risks accumulated by diabetic patients. They developed a sophisticated processing-based predictive model and explored several methods for the continuous collection of clinical data.

Nuttapong ea al. (2023) aim to introduce the integrated application of open data and machine learning methods. The contribution of this study is the development of an application on Google Earth Engine for extracting satellite data, and machine learning techniques are applied to the obtained satellite data to predict provincial GDP. The empirical validation shows that the random forest (RF) method achieves the highest predictive power, with 97.7 percent accuracy.

Jain, S. (2022) conducted a study to estimate the shared risk factors for NCDs within a population. This study adopted machine learning to arrange the factor affecting NCDs. The predictive analytics combined with big data analysis to preprocess unstructured data and convert it into a structured format. The structured data, stored in a database. When certain metrics exceeded predefined thresholds, the analysis signaled an increased likelihood of disease occurrence, effectively forecasting disease risk based on these analytical metrics. This application of machine learning facilitated the identification of specific variables influencing NCDs

Future opportunities for geographic research in global health lie in the theoretical exploration of globalization and the geopolitics of infectious and chronic diseases, as well as health-related behaviors (Tim and Graham, 2012). This includes the utilization of mapping techniques and disease spread modeling, along with an emphasis on global health perspectives and geopolitical considerations. In studying the geographical association of factors influencing NCDs, this research employs spatial analysis to examine the relationships between significant determinants in each region.

2.3.2 NCDs's determinant for spatial analysis

Spatial correlation analysis plays a vital role in linking NCDs with geographic environmental and economic factors by examining spatial patterns and associations within a geographical context. It investigates the spatial patterns of health indicators and their associations with influencing factors (Li et al., 2024). This analysis extends the scope of study to serve multidisciplinary between economic health and GIS data.

Onprasongk et al. (2023) investigated the spatial relationship between environmental factors and the incidence of melioidosis in Thailand. The study utilized public health data and satellite-derived environmental factors spanning from 2006 to 2020. To ensure consistency, all variables were aggregated at the provincial level to match the spatial resolution of melioidosis incidence. Through an analysis of bivariate local indicators of spatial association (LISA), it was revealed that all satellite-based environmental factors exhibited statistically significant associations with melioidosis incidence. Furthermore, the analysis identified statistically significant clusters predominantly located in the Northeast region.

Nkhasi and Eeden (2016) examined the environmental factors focusing on air pollution, created environment and nature, exposure to light, chemical water pollution, climate, and the global environmental crisis, which is the distribution of mortality from NCD among persons by identifying towns with a high risk of fatalities from specific NCDs in small-area geospatial maps by using Moran's Index, hot spot analysis, and regionally weighted regression, which were applied by using demographic data in municipalities and the public health facility database. The study can point out the association between environmental factors and the occurrence of NCD. Clusters of high-risk deaths from a few NCDs were found to tend to occur in urban municipalities in the country's midsection and west, whereas those tended to occur in urban and rural areas of the northeastern region, as well as in urban municipalities of the Northern Cape. The hot zones revealed distinct socioeconomic and demographic risk factors for NCDs, such as urbanization, the concentration of public hospitals, the frequency of hospital deaths, smoking, asbestos roofs, and mining and quarrying operations.

Krickovic, J., et al. (2022) focused on the geographic analysis of NCDs. This research employed hot spot analysis to identify areas with unusually high incidences of NCDs. The study highlighted the West Backa County as having a significantly higher mortality rate compared to other regions, particularly noting an elevated rate of diabetes-related deaths. Furthermore, while West Backa County reported a high number of new diabetes cases, the mortality rate for these cases was relatively low compared to other counties. Helena (2014) presents the variety of contributions that Health Geography can provide for understanding health issues in different scales and contexts by analysis of spatial variables and regression logistic multifactorial.

2.3.3 Estimation of the economic impact from NCD

The focus of this part is evaluating the effect of various indicators on productivity. These indicators included disability-adjusted life years (DALYs), productivity costs, and labor market engagement metrics such as unemployment rates, return to work statistics, and sick leave durations (WHO, 2000).

The burden of diseases, as measured by DALYs, can vary across different birth cohorts, indicating cohort disparities that significantly influence health trajectories. Studies have shown that individuals with higher socioeconomic status, such as those residing in urban areas and earning higher incomes, tend to report better health outcomes (Dowd & Zajacova, 2007; Debiasi & Dribe, 2020; Regidor et al., 2010). However, for chronic diseases, which are often linked to lifestyle factors like smoking, alcohol consumption, and poor dietary habits (Ng et al., 2020), higher socioeconomic status may have a detrimental effect. Moreover, the association between socioeconomic status and health evolves over the life course and is influenced by cohort effects. The impact of education and income levels accumulates over a person's lifetime, leading to disparities in the health status of the elderly population. Therefore, it is crucial to consider both age and life course when examining the relationship between socioeconomic status and health outcomes.

In the context of China's distinctive social and political landscape, cohort effects on chronic diseases, as measured by DALYs, are profoundly accumulated. These effects manifest in significant disparities in DALYs across different birth cohorts, particularly concerning early life experiences and socioeconomic status. Notably, there are notable disparities in DALYs based on residential location and education level, with gender differences observed: while the DALYs gaps for females have narrowed over cohorts, those for males have not shown similar trends.

To address these disparities effectively, targeted efforts are necessary, focusing on lower socioeconomic status individuals and residents of rural areas, who stand to benefit the most from interventions. By directing resources and interventions toward these vulnerable groups, initiatives aimed at reducing inequalities in DALYs can be more impactful and inclusive, particularly for recent cohorts.

Zhang et al. (2021) conducted a study to investigate the trajectories of DALYs among individuals aged 45–90 years, focusing on cohort, socioeconomic status, and gender disparities. They employed a growth curve model to forecast DALYs trajectories influenced by various birth cohorts and SES factors. The findings revealed significant differences in DALYs attributable to cohort and SES factors, including residential location, education level, and income. Specifically, individuals from earlier cohorts tended to develop DALYs at a later age but experienced rapid growth with advancing age, whereas this trend was reversed for more recent cohorts. Moreover, residing in urban areas and possessing higher SES levels were associated with a decrease in the growth rate of DALYs with age, with convergence observed among the most recent cohorts. The study related that NCDs impact is a rising problem worldwide and a cause of public health concern in India. These diseases result in high healthcare demand and are expensive, especially in low-resource settings.

Menon et al. (2022) studied the NCDs burden estimates for India and its major states and the associated costs in 2017. The disease burden was estimated using DALYs, and the years of life lost (YLL) and years lived in disability (YLD) by using health expenditure data from the National Sample Survey were used to calculate out-of-pocket expenditure (OOPE) and catastrophic health expenditure (CHE). The results of NCDs account for huge variations in the utilization of health facilities, disease burden, and cost of treatment across the states. Karnataka had the highest DALYs rate, indicating variation across states in both disease burden and economic burden.

Another investigation by Chaker et al. (2015) revealed significant regional disparities in DALYs attributable to NCDs, particularly for cervical and lung cancer. The study found substantial productivity losses in the United States of America, ranging from 88 million US dollars for Chronic Obstructive Pulmonary Disease (COPD) to 20.9 billion US dollars for colon cancer. Coronary Heart Disease (CHD) was estimated to cost the Australian economy 13.2 billion US dollars annually. Individuals with Diabetes Mellitus (DM), COPD, and survivors of breast and lung cancer, in particular, face elevated risks of reduced labor market participation. Overall, NCDs have a profound impact on macroeconomic productivity across various World Health Organization regions, regardless of continent or income level.

Normally, studying the effects of disease, always use clinical statistics. In this research, the environmental and economic factors included other spatial study variables that are more spatial than geographic and have a tendency to influence NCDs, such as population density, urbanization, and gas pollution. The research has studied and compared machine learning Model to select effective Model to predict the noise of disease areas in order to make the study more effective.

CHARPTER 3
RESEARCH METHODOLOGY

This chapter outlines the research framework and methodology employed in this study. The study investigates the influence of environmental and economic factors on NCDs in Thailand by examining data on the prevalence and mortality rates of conditions such as diabetes, hypertension, ischemic heart disease, and chronic respiratory diseases from 2018 to 2021. Utilizing machine learning techniques and spatial econometric analysis, the research examines satellite data, including air pollution and land surface temperature, and economic indicators like GDP and nighttime light intensity. The goal is to identify spatial patterns and area densities of NCDs and assess their economic impact on the Thai economy.

3.1 Data Sources and description

This section examines a dataset that investigate the impact of environmental and economic factors on NCDs mortality, utilizing satellite data.

3.1.1 The NCDs mortality rate

The prevalence of mortality rates of NCDs diseases have been officially collected by the Ministry of Interior. This data set reveals the number of NCDs mortality including diabetes, hypertension, ischemic, and chronic. The NCDs mortality analyzed the number of deaths per 100,000 people between 2018 and 2021, categorized by province and health district. The mortality rates of NCDs diseases have been officially collected by the Ministry of Interior.

3.1.2 The environmental and geographic data

The study utilizes environmental data, primarily sourced from satellites and surveys. Google Earth Engine is a cloud service that provides access to satellite data.

3.1.2.1 Sentinel-5P (SP5)

Sentinel-5P (SP5) measures a portion of the pollution gases includes nitrogen dioxide (NO_2), carbon monoxide (CO), sulfur dioxide (SO_2) and methane (CH_4), with high temporal and spatial resolution. Sentinel-5P (SP5) was aimed at reducing the uncertainties by thoroughly assessing all aspects of instrument performance, stability, accuracy, and suitability of the data processing by comparison with independent measurements and analyses. S5P will provide measurements of elements of atmospheric chemistry at high temporal and spatial resolution.

3.1.2.2 Land Surface Temperature

Land Surface Temperature data show an increase in both daytime and nighttime temperatures over the year by the moderate resolution imaging spectroradiometer (MODIS) satellite is equipped with several sensors, allowing it to compile various information on the Earth’s surface. The satellite measures land surface temperature during daytime and nighttime with a spatial resolution of 1 km. The data used as a time series of ground-level temperatures. All satellite-based indicators were transformed to the monthly averages of each province.

3.1.2.3 Global land cover types

The data of Urban Areas and Cropland Areas are from Global land cover types were obtained by applying techniques to Terra and Aqua MODIS reflectance data. This data generates an annual global map of land use, which represents cropland and urban areas with a spatial resolution of 500 meters.

3.1.2.4 Precipitation (Rainfall)

A global data set that provides a long-period gridded rainfall time series which is produced by using an interpolation technique that incorporates satellite information and station data, generating precipitation estimates with high accuracy and resolution. Data on monthly precipitation averages were obtained from the integration of the global Cold Cloud Duration (CCD) rainfall estimate and Multi-Satellite Precipitation Analysis (TMPA 3B43).

3.1.2.5 Particulate matter with diameter of less than 2.5 micron

Particulate matter with a diameter less than 2.5 microns (PM 2.5) is also considered and collected daily basis from the automatic air quality monitoring points of the Pollution Control Department, which the Pollution Control Department reports annually.)

3.1.3. The Economic and Socio demographic data

There are Nighttime-Light Data, which reflects economic activity and socioeconomic status which is index derived from nighttime images captured by VIIRS/DNB satellite. The data provides a series of monthly and daily NTL data. Additionally, Gross Provincial Product (GPP) and population statistics are obtained from the Office for National Statistics, allowing for control of socioeconomic factors. Moreover, the study incorporates socioeconomic status data on tobacco and alcohol consumption, as well as age and gender demographics.

3.1.3.1 Gross Regional and Provincial Product (GPP)

The data include production of goods and services which in turn leads to income payable to factors of that production, disposition of that income, saving and investment of economic institutions. This also includes economic activities dealing with the rest of the world, especially export and import of goods and services. The data is from Office of the National Economic and Social Development Council.

3.1.3.2 Nighttime Light Data

Nighttime-Light Data is an indicator that represents economic activity and socioeconomic status, the NTL index can be derived by using nighttime imageries captured by satellites. In addition, The Visible Infrared Imaging Radiometer Suite/Day-Night Band (VIIRS/DNB) satellite distributes a series of monthly and daily NTL data.

3.1.3.3 Socioeconomic Status Data

For this study, the important determinant of NCDs is the socioeconomic demographic and individual behavior in daily life. To enhance the accuracy of the model, the study employs socioeconomic status data related to NCDs, including variables such as portion expenditure on tobacco and alcohol, age, and gender. These variables are averaged using the geometric mean technique over the period from 2018 to 2021.

3.1.3.4 Population Statistics

Total Population is counted of total population by regional boundaries from Office for National Statistics.

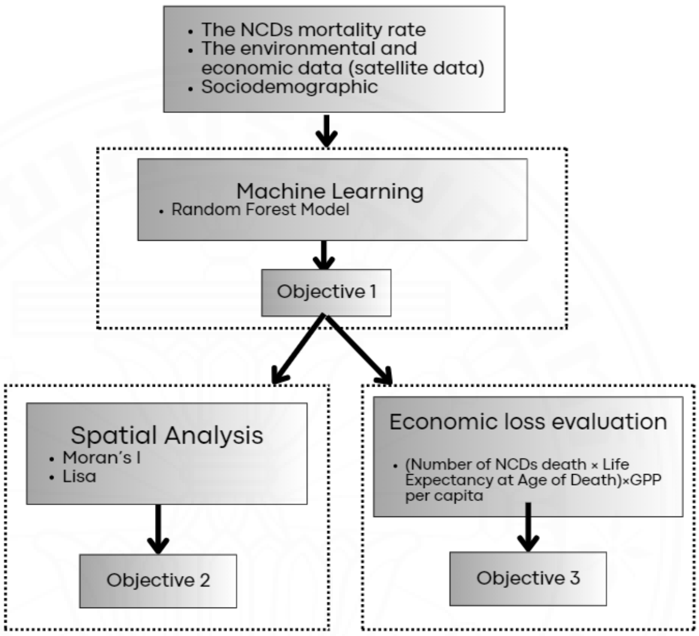
3.2 Framework of the study

This section illustrates a study framework that investigates the association between environmental and economic factors impacting NCDs in Thailand using the following processes. Firstly, all retrieved data will be normalized by taking natural logarithms. Then, machine learning techniques will be utilized to quantify the influential factors by setting mortality of NCDs as target variables, including diabetes, hypertension, ischemic, and chronic, during the years 2018–2021, while both environmental factors and economic factors were set as explanatory variables.

Then, spatial analysis e.g., Local Indicator of Spatial Association (LISA) will be applied for mapping the influence factors by colored map clustering. The mapping results reveal the NCDs density pattern and highlight the correlation of significant factor-related NCDs. Finally, to examines the economic impact of NCDs mortality in Thai provinces using QGIS and regression analysis. The study employs the concept of DALYs analysis, with identified NCDs mortality rates by age. To provide more insight, the study includes demographic and economic data such as population, gross domestic product (GDP), and household socio-economic status that are incorporated into the model to provide a holistic perspective on the economic impact of NCDs. And identifies affected provinces, estimates economic costs, conducts regression analysis to validate the model, and generates maps illustrating the impact as presented in Figure 3.1.

Figure 3.1

Conceptual framework of the correlation among related spatial factors NCDs and the Thailand economy



Note: Visualized by the author.

However, limited resource constraints have played significant roles in the availability of data limitations concerning the environmental indicators derived from satellite data, which are restricted to the years 2018–2021. Despite this constraint, the study asserts the sufficiency of the selected four-year period for conducting it, supported by evidence suggesting that the chosen timeframe adequately captures relevant environmental dynamics, ensures a more robust and reliable assessment of environmental influences on NCDs prevalence and mortality rates. This provides sufficient data for meaningful analysis and inference regarding the association between environmental factors and NCDs in Thailand's regions.

The following steps outline the detailed process of the study. It begins with identifying influencing factors through machine learning. Once these variables are obtained, they are utilized in a spatial analysis to examine the impact on specific areas. Subsequently, an analysis of the economic loss from non-communicable diseases (NCDs) at the regional level is conducted to assess the impact on the Thai economy.

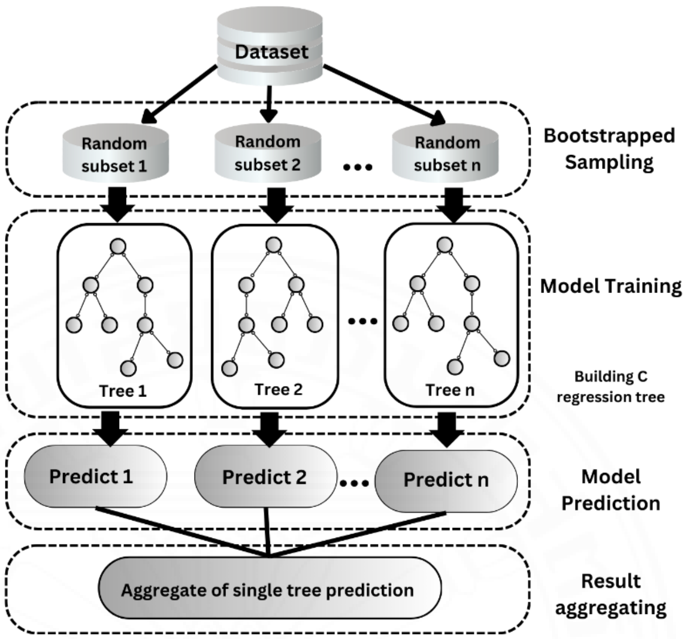
3.2.1 Identify influencing factors through machine learning

Along with the expanding public access to satellite data and machine learning techniques, Onprasonk et al. (2023) suggested that the crucial concern for estimating disease is that the clinical symptoms are similar to those of other diseases. Hence, the obtained regression coefficients might not precisely represent the actual association between environmental factors and disease. Machine learning, using open-source software packages like R Studio, drove this study. The order to identify a robust association and predict NCDs patterns between environmental and economic factors across Thai provinces, the most accurate machine learning that investigates a number of classification algorithms is the random Forest (RF) model.

One commonly employed supervised machine learning technique for discontinuity characterization is the artificial RF algorithm. The RF model, pioneered by Breiman in 2001, is an ensemble learning method that employs the bagging process to generate a lower-variance output through the coupling of deep trees. To evaluate feature importance, RF systematically varies one input variable while keeping others constant and measures the resultant decrease in the model's prediction accuracy. This process assigns a relative importance score to each input variable (Breiman, 2001). In making final decisions, the RF classifier aggregates the decisions of individual trees, resulting in robust generalization. Notably, RF classifier consistently outperforms many other classification methods in terms of accuracy while avoiding issues of overfitting. Refer to Figure 3.2 for an illustration of the RF algorithm's structure.

Figure 3.2

The structure of random forest algorithm



Note: The author visualized from Tao et al. (2022) and M.W. Ahmad et al. (2018).

3.2.1.1 Data Collection and Preprocessing

To investigate the impact of various factors on NCDs, the author collected data from multiple sources. The dataset comprised samples, each with 16 dimensions representing different variables such as demographic, environmental, and health-related factors. Data was sourced from national health surveys, economic reports, and environmental databases. Then the dataset was cleaned to handle missing values, outliers were identified and treated, and categorical variables were encoded. All variables were normalized to ensure comparability and to facilitate the machine learning process

3.2.1.2 Number of trees (ntree)

The number of trees to consider when building the model. The RF algorithm functions by building numerous trees (ntree) on a bootstrap sample of the training data, with replacements drawn from the original observations. It then outputs the results by considering the majority-voted prediction as the outcome or by calculating the average in regression tasks (Ali et al., 2021). RF consists of many decision trees grown in parallel, reducing both bias and variance (Breiman, 2001). The training process involves bootstrapping followed by aggregation, known as bagging. Bootstrapping means training individual decision trees on various subsets of the dataset with different features, ensuring each tree is unique and reducing overall variance. Instead of using all predictors, a fixed number of randomly sampled predictors are selected as split candidates. These two steps are then repeated until C such trees are grown, and new data is predicted by aggregating the prediction of the C trees. RF uses bagging to increase the diversity of the trees by growing them from different training data-sets, and hence reducing the overall variance of the model (Rodriguez-Galiano et al., 2015). A RF regression predictor can be expressed as

$$f_{RF}^{C}(x) = \frac{1}{C}\sum_{i=1}^{C} T_i(x) \tag{3.1}$$

where x is the vectored input variable, C is the number of trees, and $T_i\ (x)$ is a single regression tree constructed based on a subset of input variables and the bootstrapped samples.

3.2.1.3 Out-Of-Bag (OOB) error

As each tree is a subset of the data available, the “out-of-bag” samples to validate and quantify the quality of our trees. As the number of trees increases the OOB error will stabilize, evidence that our model might perform well on unseen data. The following Equation illustrates a high-level view of how the error, e, decreases for each i^{th} tree.

$$OOB_{error} = \frac{1}{n_{trees}} \sum_{i=1}^{n_{trees}} e_i \tag{3.2}$$

The RF algorithm offers native capability for out-of-bag error estimation during the forest construction process by utilizing samples not selected in training each tree. This subset, termed out-of-bag, enables unbiased estimation of generalization error without necessitating an external test dataset (Breiman, 2001). Moreover, RF assesses the importance of input features, facilitating dimensionality reduction for high-dimensional datasets (Ahmad et al., 2017).

3.2.1.4 Tree depth

In essence tree depth is the number of levels (nodes) a tree can have and unlike the previous example of adjusting the number of trees available, increasing tree depth can have a detrimental impact to model explainability, complexity and computational load. Unreasonably increasing the value of this parameter increases the risk of overfitting the model. The depth of a tree is proportional to its complexity, as deeper trees will have more nodes and therefore more decision boundaries. Whilst this might allow for a model to capture more intricate patterns within the data there is a greater risk of the model starting to overfit or becoming more difficult to interpret. Balancing tree depth with the tree count is important to reduce the likelihood of overfitting as the latter will, to a degree, mitigate it to some extent.

Also, Hastie et al. (2009) looked at the analyses of variable importance (VIMP) and minimal depth (MD) have been developed to quantify the influence degree of each variable on a model's predictive power. Hence, equipped with VIMP and MD, RF exhibits another advantage, providing a deeper understanding of the association between each variable and the predicted outcome. A RF operates by constructing an ensemble of decision trees, each trained on a subset of the data. This ensemble approach allows us to assess the relative importance of geographical and economic factors in predicting NCDs prevalence. The author conducts extensive feature selection and interpret the variable importance rankings to identify key contributors to the observed patterns.

3.2.2 Spatial analysis

After the author got the influencing factors by using machine learning to analyze the association among environmental and economic factors and NCDs in each area by spatial analysis. In the spatial analysis method, the research employed GIS and the Geoda program, which represent data in a map data format that integrates location data with its descriptive information and also aids in revealing patterns, associations, and geographic context to study spatial association. This study used GIS software to integrate various datasets in the spatial analysis method.

Additionally, the Geoda program facilitated data representation in a map format, merging location data with descriptive information and also aiding in revealing patterns of geographic context to study spatial association.

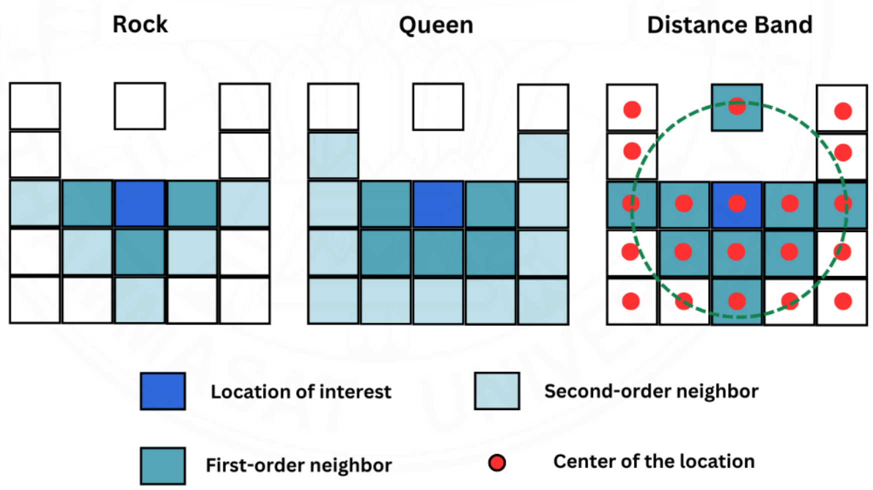
3.2.2.1 Spatial model

(1) Spatial weight matrix

The relative location or the neighbor relation is expressed in the spatial model via the spatial weight matrix. A spatial matrix is a non-negative matrix that represents the adjacent location of each individual in the cross-sectional space. The spatial weight matrix is mostly created by “Rook”, “Queen”, or “Distance band” contiguity criterion. Figure 3.3 illustrates the neighbor location created under each criterion. Rows and columns in the spatial weight matrix correspond to the space of the observations.

Figure 3.3

Spatial neighbor criterion



Note: The author visualized from Chitdecha, M. J., (2016).

The spatial weight matrix employed in this study is queen contiguity criterion. The neighbor’s location under the Queen contiguity criterion, the neighbor is every area that shares the common boundary line. The queen contiguity criterion was chosen considers all areas that share either a common boundary or a vertex as neighbors. This provides a more comprehensive representation of spatial relationships compared to the Rook criterion, which only considers shared boundaries.

By including areas that touch at corners, the Queen criterion captures a broader range of potential spatial interactions, reflecting the complexity of real-world spatial relationships more accurately. Moreover, the Queen contiguity criterion is particularly effective in capturing spatial dependencies that may occur through both direct and indirect contacts. In many empirical applications, such as environmental and economic studies, interactions are not limited to immediate neighbors but also occur through more intricate networks of connections. By encompassing both shared boundaries and vertices, the queen criterion ensures that these more nuanced spatial dependencies are adequately accounted for. Many geographic regions do not conform to regular grid patterns, often exhibiting irregular shapes and sizes. The queen contiguity criterion is well-suited to such irregular geographies because it captures neighboring relationships more flexibly. This flexibility is crucial for accurately modeling spatial processes in regions with complex, non-uniform spatial structures, where interactions may occur across multiple adjacent areas, not just through direct boundary sharing.

For Alignment with Theoretical Frameworks in spatial econometrics often assume that spatial interactions can occur through both direct and indirect pathways, including diagonal connections. The Queen contiguity criterion aligns with these frameworks by incorporating diagonal neighbors, thus ensuring that the spatial weight matrix reflects the underlying theoretical assumptions about spatial processes. This alignment enhances the interpretability and robustness of the empirical results. And the last the Queen contiguity criterion helps ensure the robustness of spatial analysis by providing a detailed and inclusive framework for defining spatial relationships. This robustness is crucial for producing reliable and accurate estimates of spatial effects, particularly in studies that examine the impact of spatial factors on outcomes across different regions. By using the Queen criterion, the study benefits from a robust analytical framework that enhances the validity of the findings.

(2) Spatial methods

The study employed two spatial methods:

1) Moran's I Statistic

The spatial autocorrelation statistic was one of the univariate computational techniques for quantifying the degree of spatial autocorrelation. Moran's, I value ranges between −1 and 1. Where the value was close to 1, it indicated a highly positive spatial autocorrelation; otherwise, it indicated an extremely negative one. While Moran's I has a zero value, there is no spatial autocorrelation.

$$\text{Moran's} \quad \mathrm{I} = \frac{\sum_i \sum_j w_{ij}\left(X_i - \overline{X}\right)\left(X_j - \overline{X}\right)}{\sum_i \sum_j w_{ij}\left(X_i - \overline{X}\right) \ / \ \mathrm{N}} \tag{3.3}$$

Where X_i is the variable used to study (NCDs, environmental and economic indicators, etc.). $\overline{X}$ is an arithmetic mean. n is the number of spatial unit indexed by and (in this case, the province).w_{ij} is weighted matrix which i and j are indices that denote spatial units or locations within a defined geographic area (i represents one spatial unit or province and j represents another spatial unit or province.).

2) Local Indicators of Spatial Association (LISA)

This method gauged data concentration at the spatial level and identified clustering patterns' locations. LISA generated outcomes, including a cluster map showcasing four categories of spatial correlations: high-high, low-low, low-high, and high-low (Anselin, 1995). LISA is a statistic used to measure the concentration of data at the spatial level. The outcome generated by LISA includes the cluster map showing four categories of spatial correlations, high–high, low–low, low–high and high–low (Anselin, 1995). The cases of high–high and low–low indicate the spatial cluster of positive correlation, and the areas of low–high and high–low define the spatial association of localized negative correlation. LISA method was applied to identify the clustering pattern's location. The mathematically defined as

$$I_i = Z_i \sum_j w_{ij} Z_j \tag{3.4}$$

Where Z_i and Z_j are observed values in the form of z-scores and w_{ij} as a row-standardized matrix which i represents values at a particular location and j represents clustered or dispersed relative to neighboring locations, both local Moran test statistics and Local Indicators of Spatial Association can be analyzed in the form of univariate spatial correlation and bivariate spatial correlation. If the variables of interest in study are X_i and Z_j.

3) Getis-Ord Statistics (G_i)

For local spatial autocorrelation was suggested by Getis and Ord (1992), and further elaborated upon in Ord and Getis (1995). It is derived from a point pattern analysis logic. In its earliest formulation the statistic consisted of a ratio of the number of observations within a given range of a point to the total count of points. In a more general form, the statistic is applied to the values at neighboring locations (as defined by the spatial weights). The $G_i$statistic consist of a ratio of the weighted average of the values in the neighboring locations, to the sum of all values, not including the value at the location (x_i).

$$G_i = \frac{\sum_{i=j} w_{ij} x_j}{\sum_{i=j} x_j} \tag{3.5}$$

The interpretation of the Getis-Ord statistics is very straightforward: a value larger than the mean (or, a positive value for a standardized z-value) suggests a High-High cluster or hot spot, a value smaller than the mean (or, negative for a z-value) indicates a Low-Low cluster or cold spot.

4) The significance map

Significance maps in GeoDa use permutation tests to determine the significance of the local statistics. This involves comparing the observed statistic to a distribution of statistics obtained under the null hypothesis The significance map shows the locations with a significant local statistic, with the degree of significance reflected in increasingly darker shades of green.

The map starts with $p < 0.05$ and shows all the categories of significance that are meaningful for the given number of permutations. The calculation of p-value is the proportion of permutations (Generate a large number of random permutations of the data and calculate the statistic for each permutation) that produce a statistic as extreme or more extreme than the observed statistic.

$$\rho = \frac{Number\ of\ permutation\ with\ I_i^{(perm)} \geq I_i^{(obs)}}{Total\ number\ of\ permutations} \tag{3.6}$$

The p-values are visualized on a map, with colors representing different significance levels (e.g., darker shades for lower p-values indicating higher significance).

3.2.3 Economic loss from NCDs

The approach aims to analyze the economic repercussions of NCDs mortality rates and their implications for regional economic stability and growth. Additionally, geographic information systems (GIS) software such as QGIS is utilized to create spatially explicit Model capable of analyzing economic data and mapping the distribution of economic impacts across different provinces. QGIS provides tools for importing, processing, and visualizing spatial data, allowing researchers to overlay economic indicators onto the Thailand map and generate thematic maps illustrating the economic impact of NCDs mortality.

3.2.3.1 Identify the economic loss from NCDs

The method, referred to as ECON-NCD (Economic Impact of NCDs), employs statistical and econometric techniques to assess the financial burden of NCDs mortality across different provinces in Thailand. In the initial step, data on NCDs mortality rates for each province are collected and organized into matrices. These matrices contain categorical and numerical data representing various economic indicators and NCDs mortality rates. Subsequently, diverse datasets encompassing economic variables and attributes of different provinces are utilized to identify regions with significant economic impacts due to NCDs mortality. Once provinces with significant economic impacts are identified, the method proceeds to estimate the economic cost of NCDs mortality in each province.

The calculation account for the disease burden indices were Years of Life lost due to premature mortality (YLL); Years lived in disability (YLD) and Disability Adjusted Life Years (DALYs). One DALY equates to one year of healthy life lost and is computed as the sum of YLL and YLD. (Menon et al., 2019) The methods, strengths, and limitations of the YLD therefore this study employs YLL and GPP to estimation economic loss from NCDs.

$$Economic\ loss = (Number\ of\ NCDs\ death \times Life\ Expectancy\ at\ Age\ of\ Death) \times GPP\ per\ capita \quad (3.7)$$

This estimation incorporates local economic parameters and indicators, including healthcare expenditure, loss of productivity, and impact on GDP growth, to quantify the economic burden of NCDs mortality.

3.2.3.2 Estimation of NCDs on Economic loss

This section investigates how changes in various economic factors impact economic losses associated with NCDs. The study integrates NCDs and different economic sectors to examine their combined effects on economic outcomes. This approach aims to account for variations in socio-economic status and environmental factors across different regions, which could introduce heterogeneity into the analysis. However, due to constraints in the available socio-economic data, this study focuses primarily on the direct economic impacts of NCDs.

(1) Data Collection and Aggregation

Initially, panel data at the provincial level is collected, including economic indicators such as GPP and economic loss due to NCDs. This data is then aggregated across different provinces and over time to construct a comprehensive dataset. The dataset includes relevant control variables that could influence economic outcomes, such as population density and environmental factors, ensuring a well-rounded analysis.

(2) Empirical Model

Given the constraints, this study suggested by Mincer (1974). The empirical model is non-linear, and aggregating data at the provincial level may introduce spatial autocorrelation issues (Goeij and Verbeek, 2000). To estimate the classification using 4 Model and test the robustness of the model.

For model 1 tests the association between economic loss and GPP due to economic loss from NCDs. We define k denoting each NCDs disease (Diabetes, Hypertension, Ischemic and Chronic), i representing the province, t representing the time period. The Equation is as follow

$$\ln(gpp_{kit}) = \beta_{k0} + \beta_{k1}\ln(econloss_{\,kit}) + \epsilon_{kit} \tag{3.8}$$

where $\ln(gpp_{kit})$ is the natural logarithm of the denotes the GPP for province i related to NCDs k at time t.

$\ln(econloss_{\,kit})$ is the natural logarithm of the economic loss due to NCDs k in province i at time t

β_{k0}, β_{k1} are the coefficients for NCDs k, β_{k1} representing the elasticity of GPP with respect to economic loss.

ϵ_{kit} is the error term capturing unobserved factors affecting gpp_{kit}

Model 2 incorporates additional control variables that have been identified through machine learning analysis. These variables are specific to each disease studied and are included in the model to adjust for external factors that could influence the outcomes being analyzed. The Equation is as follow

$$\ln(gpp_{kit}) = \beta_{k0} + \beta_{k1}\ln(econloss_{\,kit}) + \sum_j \gamma_{kj} z_{kjit} + \epsilon_{kit} \tag{3.9}$$

The additional variables are in the term of $\sum_j \gamma_{kj} z_{kjit}$ which is the sum of control variables z_{kjit} weighted by them respective coefficients γ_{kj} where j represents each control variable specific to NCDs k in province i at time t.

Model 3 incorporates both control variables and year dummies to capture time-variant effects. It examines how the number of patients increasing each year affects economic loss and GPP. The Equation is as follow

$$\ln\ gpp_{kit} = \beta_{k0} + \beta_{k1}\ln\ econloss_{\,kit} + \sum_j \gamma_{kj} z_{kjit} + \sum_t \delta_{kt} Year_t + \epsilon_{kit} \tag{3.10}$$

The additional variables are in the term of $\sum_t \delta_{kt} Year_t$ is the sum of dummy variables for each year t, weighted by Coefficients δ_{kt}, capturing the year-specific effects.

And model 4 which is includes control variables, year dummies, and the interaction effect between time-variant and economic loss from NCDs to capture the time-variant effect of economic loss. It assesses whether sensitivity changes with each year's variation. The Equation is as follows

$$\ln gpp_{kit} = \beta_{k0} + \beta_{k1} \ln econloss_{kit} + \sum_j \gamma_{kj} z_{kjit} + \sum_t \delta_{kt} Year_t + \sum_j \sum_t z_{kjit} \times Year_t + \epsilon_{kit} \quad (3.11)$$

The additional variables are in the term of $\sum_j \sum_t z_{kjit} \times Year_t$ is Interaction terms between control variables z_{kjit} and year dummies, capturing how the effect of each control variable changes over time.

In econometric analysis, it is crucial to ensure that the estimated model is valid and correctly specified. Therefore, this analysis employs four Model of specification to achieve a comprehensive understanding and robust results. By employing these four Model, this analysis ensures a thorough examination of the relationship between economic loss due to NCDs and GPP. Each model builds upon the previous one, adding layers of complexity and control to validate the findings. This step-by-step approach ensures that the estimated Model are valid and correctly specified, providing robust and reliable results for policy recommendations.

(3) Model Selection for Estimating the Economic Impact

To ensure the appropriateness of the model, conducted to determine whether a fixed effects model is more suitable than a random effects model. The results, presented in Appendix A, indicate significant differences, implying that unobserved heterogeneity is correlated with the independent variables. This supports the use of a fixed effects model, as recommended by Wooldridge, for controlling for unobserved heterogeneity across provinces. The fixed effects model is estimated using Ordinary Least Squares (OLS) to account for province-specific characteristics that may affect the relationship between economic loss and GPP.

Following the decision to use a fixed effects model, an endogeneity test is conducted to address potential biases from endogenous variables. This is crucial because unaccounted endogeneity can lead to biased and inconsistent estimations. The test results, included in Appendix B, confirm that the estimated relationships are not influenced by endogenous explanatory variables, thereby validating the model's robustness.

(4) Model Robustness and Variable Significance

Although the model demonstrates robust parameter estimates, it does not display the adjusted R-squared, which is typically used to select the most appropriate model. Generally, as more variables are added to the model, the R-squared value increases. However, this increase might not be significant. Therefore, it is essential to test whether the added variables are significant using a likelihood ratio (LR) test. The null hypothesis for the test must be clearly stated: testing if additional coefficients from the previous model are significant

$$H_0 : \beta_j = 0 \tag{3.12}$$

where j denotes all of additional coefficients from the model.

If the test results show significant changes, it indicates that the new model, with the additional variables, is more efficient. If not, adding the variables is unnecessary and may reduce the model's accuracy due to a loss of degrees of freedom. Thus, the goal is to select the most efficient model, ensuring significant improvements from one model to the next. The methodology detailed here reflects a structured and comprehensive approach to understanding the economic impacts of NCDs at the provincial level, ensuring that the chosen model is both robust and relevant for policy and economic analysis.

CHARPTER 4
RESULTS AND DISCUSSION

This chapter presents an in-depth analysis of the study's findings, exploring the intricate association among environmental, economic, and NCDs over the period from 2018 to 2021. The chapter begins with summary statistics, highlighting significant trends and variations in key variables. It then extends to spatial analysis, identifying patterns of geographical heterogeneity and their association with NCD mortality. The clustering analysis reveals high-density mortality clusters, which pinpoint critical regions for public health interventions. Furthermore, the chapter examines the influencing factors on NCDs, utilizing a RF model to determine optimal thresholds for significant variables. The spatial distribution of diseases is also explored, employing the LISA method to illustrate spatial correlations, with a focus on areas of severe mortality that may benefit from targeted policies. Finally, the economic impact of NCDs is assessed using QGIS mapping, revealing disparities in economic losses across different regions and emphasizing the need for effective strategies to mitigate these losses.

4.1 Summary statistics

The summary statistics for various environmental, economic, and health related variables from 2018 to 2021 reveal significant trends and variations.

Table 4.1

The summary statistics of data

Variable (Unit)	Abbreviations	2018		2019		2020		2021	
		Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
Sentinel-5P Nitrogen Dioxide (parts per million, ppm)	NO_2	0.031	0.033	0.063	0.011	0.060	0.008	0.065	0.008
Sentinel-5P Carbon Monoxide (parts per million, ppm)	CO	6.521	15.234	38.703	8.831	38.288	10.827	35.842	10.862
Sentinel-5P Sulphur Dioxide (parts per million, ppm)	SO_2	0.001	0.010	-0.023	0.039	-0.010	0.043	-0.015	0.054
Sentinel-5P Methane (parts per million, ppm))	CH_4	170,130.74	487,371.53	1,114,620.695	878,207.846	1,028,134.234	896,000.120	1,184,717.589	795,788.675
Land Surface Temperature Day time (degrees Celsius, °C)	lst_day	29.925	1.942	31.240	2.762	30.764	3.118	30.196	2.625
Land Surface Temperature Night time (degrees Celsius, °C)	lst_night	22.345	1.544	23.134	2.046	22.736	2.030	22.966	2.046
Urban Area Index (square kilometers, sq. km)	$urban\_area$	323.760	1.149	327.584	1.1490	331.408	1.149	337.933	0.977
Cropland Area Index (square kilometers, sq. km)	$cropland$	14.510	0.282	14.901	0.514	16.296	0.467	16.630	0.282
Nighttime Light Index (nanowatts per square centimeter per steradian ,nW/cm²/sr)	ntl_den	10,680.904	43.754	10,628.976	40.211	10,502.202	40.211	10,539.826	22.690
Precipitation (Rainfall) (millimeters, mm)	$Precipitation$	141.248	109.149	116.223	109.040	148.923	132.533	142.554	128.493

Table 4.2

The summary statistics of data (Cont.)

Variable (Unit)	Abbreviations	2018		2019		2020		2021	
		Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
Population (persons)	$Population$	825,136.785	5,060.625	837,645.772	5,247.370	850,523.402	5,247.369	86,6057.60	4,463
Gross Regional and Provincial Product (million baht)	GPP	21,5016.49	632,560.569	222,018.094	667,201.643	20,5461.5	612,868.949	212,341.158	626,933.303
Particulate matter with diameter of less than 2.5-micron micrograms per cubic meter, μg/m³)	$pm\ 2.5$	10.422	11.769	10.361	11.776	10.299	11.744	10.253	11.704
Portion expenditure of tobacco (percentage, %)	$tobaco$	0.354	0.151	0.330	0.165	0.373	0.173	0.382	0.187
Portion expenditure of alcohol (percentage, %)	$alcohal$	1.996	0.760	0.930	0.282	1.856	0.696	1.867	0.780
Age (Years)	age	41.052	2.929	42.007	2.901	41.910	2.937	42.485	2.987
Portion of Gender (percentage, %)	$male$	47.573	1.470	47.744	1.575	47.588	1.412	47.428	1.317
Diabetes disease (mortality cases per 100,000 people)	$diabetes$	186.540		216.211		213.671		207.440	
Hypertension disease (mortality cases per 100,000 people)	$hypertension$	112.618		122.092		121.697		119.467	
Ischemic disease (mortality cases per 100,000 people)	$ischemic$	272.118		268.842		278.750		286.853	
Chronic disease (mortality cases per 100,000 people)	$chronic$	88.421		90.526		78.263		68.787	

Note: The Author's summarized from Fehr, T. (2016). Sentinel-5 Precursor Scientific Validation Plan, and Puttanapong, N., et al. (2022).

Table 4.1 summarizes the data. The Sentinel-5P data shows fluctuations in air pollutants such as nitrogen dioxide (NO_2), carbon monoxide (CO), sulfur dioxide (SO_2) and methane (CH_4). For example, the mean value of NO_2 increased from 0.031 ppm in 2018 to 0.065 ppm in 2021. Similarly, CO levels increased significantly in 2019 before stabilizing in subsequent years. In contrast, SO_2 values showed negative fluctuations. Methane levels saw a considerable rise from 170,130.74 ppm in 2018 to 1,184,717.589 ppm in 2021, indicating substantial variability in methane concentrations. These changes imply a varying impact of air pollution over the years, which could correlate with changes in industrial activities, urbanization, and regulatory measures.

Land surface temperatures, both during the day and night, varied over the years, reflecting changes in climate patterns. Daytime temperatures increased from a mean of 29.925°C in 2018 to a peak of 31.240°C in 2019, then fluctuated slightly. Nighttime temperatures followed a similar trend, peaking in 2019. The increase in temperatures implies a potential influence of global warming and climate change on local temperature variations, which could impact agricultural productivity and human health. The Urban Area Index and Cropland Area Index indicate changes in land use, with urban areas expanding and cropland areas showing an increase, highlighting trends in urbanization and agricultural development. The Nighttime Light Index, which serves as a proxy for economic activity and urbanization levels, showed relatively stable values with minor variations, indicating consistent levels of urban activity. These trends imply ongoing urban expansion and stable economic activities in the region, affecting land use patterns and potentially contributing to environmental changes.

Precipitation data revealed variability, with rainfall peaking in 2020 and showing a slight decrease in 2021. The population size increased steadily from 825,136.785 persons in 2018 to 866,057.600 persons in 2021, reflecting demographic growth. The Gross Regional and Provincial Product (GPP) showed fluctuations, with notable changes in economic performance over the years. This data provides a measure of economic activity and highlights the impact of broader economic conditions on regional and provincial economies. The variations in precipitation and

GPP imply a dynamic interplay between economic activities and environmental conditions, which could influence resource availability and economic resilience.

Health variables provide insights into public health trends. The data on diabetes, hypertension, ischemic disease, and chronic disease show varying mortality rates over the years. For instance, the mean mortality rate for diabetes increased from 186.540 cases per 100,000 people in 2018 to 207.440 cases in 2021. Hypertension and ischemic disease rates showed similar trends, while chronic disease rates decreased over the observed period. These statistics highlight ongoing public health challenges and the burden of non-communicable diseases in the population. The trends in mortality rates imply the effectiveness of healthcare interventions and the need for targeted public health strategies to address the rising burden of NCDs.

Demographic variables such as age and gender distribution offer additional context for interpreting the health and economic data. The mean age of the population increased slightly over the years, while the portion of the male population remained relatively stable. The aging population implies potential challenges for healthcare services and economic productivity, necessitating adjustments in healthcare and social policies. Overall, this comprehensive data set serves as the foundation for further analysis in this study, providing essential insights into the interplay between environmental factors, economic conditions, and health outcomes.

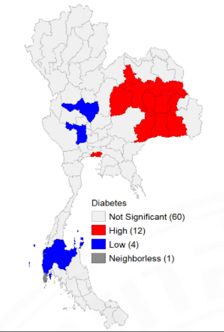
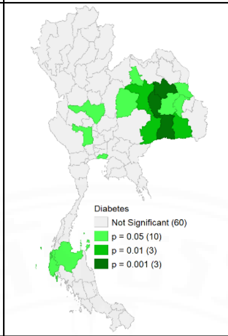
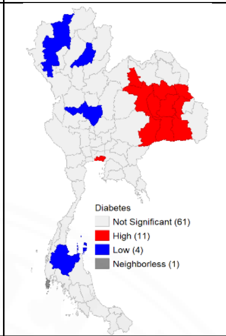
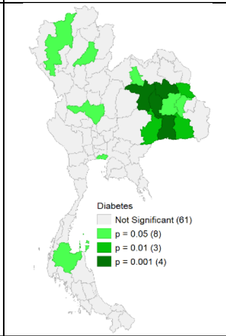
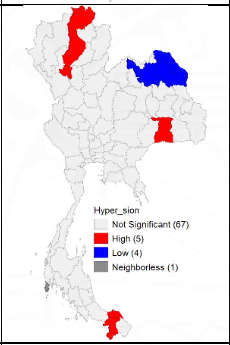
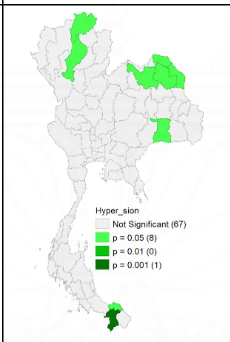
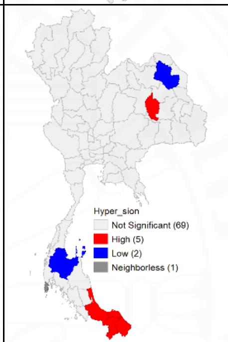
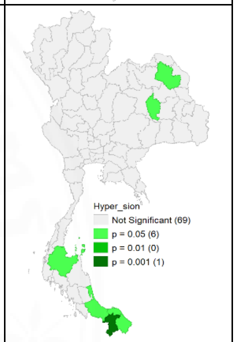
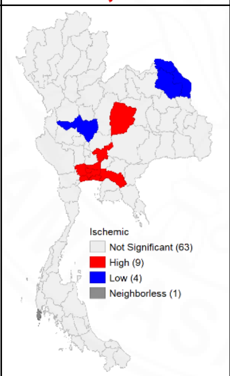
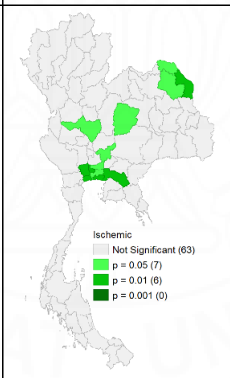
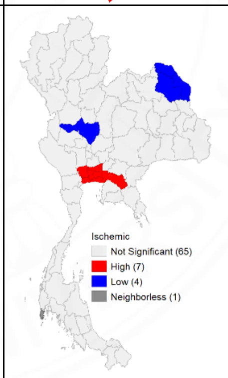
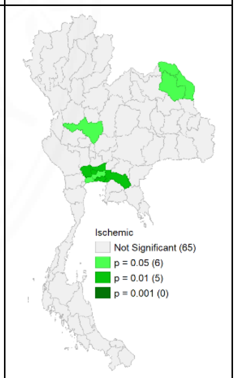
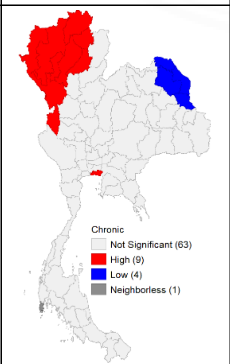
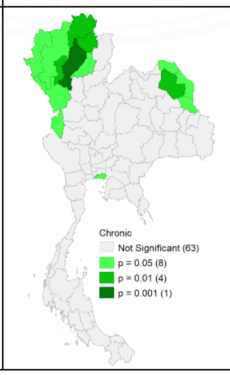
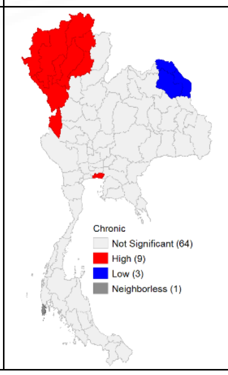
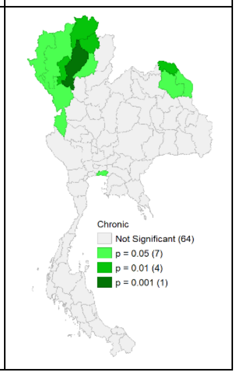
4.2 Spatial analysis by Local Spatial Clustering tool

This section extends spatial analysis to examine spatial autocorrelation of associations between geographical heterogeneity and NCDs mortality distribution. The aim is to gain deeper insights into the spatial distribution patterns of environmental and economic factors influencing NCD mortality. The results, summarized in Table 4.2, identify areas with high-density NCD mortality, represented in dark-red, and highlight significant spatial analysis.

Table 4.2 illustrates the spatial autocorrelation of NCDs mortality for the years 2018 and 2021. The Table provides a comparative analysis of the spatial density patterns and their statistical significance, showing regions where high or low mortality rates cluster. These patterns are visualized through the GI map and the significant map detailed in Chapter 3, which offer a broader context for understanding the spatial dynamics of NCD mortality. The GI map in Chapter 3 reveals overall spatial patterns, indicating areas of high and low NCD mortality. Significant clusters are marked on the significant map, showing regions with statistically significant local spatial autocorrelation.

Table 4.2

Illustrating the spatial autocorrelation of NCDs disease

Disease	Year 2018		Year 2021	
	GI Map	Significant Map	GI Map	Significant Map
Diabet				
Hypertension				
Ischemic				
Chronic				

Note: The Author visualized by using Geoda program.

Table 4.2 provides a detailed overview of the spatial autocorrelation of NCDs mortality across different regions, highlighting clusters of high and low mortality rates. The table reflects significant spatial associations between regions, as indicated by local Moran's I statistics. Dark-red areas in the table correspond to regions identified as high-high clusters, where both the region and its neighboring regions exhibit high NCD mortality rates. This suggests a strong positive spatial autocorrelation, indicating that high mortality in one region is associated with high mortality in adjacent regions.

Conversely, low-low clusters, represented in dark-blue, indicate regions with low NCD mortality surrounded by other low mortality areas, suggesting a negative spatial autocorrelation in these areas. The presence of high-low and low-high spatial outliers, represented by light-red and light-blue respectively, reveals regions where local NCD mortality rates are significantly different from those of neighboring regions, pointing to potential local anomalies or areas of interest for further investigation. The following interpretations connect the data from Table 4.2 to these maps

For diabetes, high mortality density is identified in most northeastern parts of the country, shown as dark-red areas on the map with statistical significance at the 0.1 level. The mortality of diabetes has been increasing in the northeastern region over the past five years, with adjacent areas also showing rising rates. Provinces with high diabetes mortality density include Khon Kaen, Nong Bua Lamphu, Chaiyaphum, Mukdahan, Yasothon, Sisaket, Surin, Buriram, Maha Sarakham, Roi Et, and Kalasin. Further analysis has mapped diabetes mortality density in Bangkok, providing additional insights into the distribution and impact of the disease.

The hypertension, mortality density is identified in most rural parts of the country, depicted as dark-red areas on the map with statistical significance at the 0.1 level. Provinces with high hypertension mortality density in 2018 include Chiang Rai, Lampang, Surin, Pattani, and Yala. Further analysis in 2021 has mapped hypertension mortality density in Songkhla, Pattani, Yala, Narathiwat, and Maha Sarakham.

For ischemic diseases, high mortality density is identified in most central parts of the country, depicted as dark-red areas on the map with statistical significance at the 0.1 level. Provinces with high ischemic mortality density include Pathum Thani, Nakhon Pathom, Samut Prakan, Bangkok, Nonthaburi, Samut Sakhon, and Chachoengsao.

For chronic diseases, high mortality density is identified in most northern parts, depicted as dark-red areas on the map with statistical significance at the 0.1 level. Provinces with high chronic mortality density include Chiang Rai, Chiang Mai, Mae Hong Son, Lamphun, Lampang, Phayao, Phrae, and Tak. Further analysis has mapped chronic mortality density in Bangkok, providing additional insights into the distribution and impact of the disease.

In summary, the integration of Table 4.2 with the GI map and significant map from Chapter 3 provides a comprehensive understanding of the spatial distribution of NCD mortality. The identified clusters and their statistical significance highlight regions with high mortality rates and suggest areas where targeted public health interventions may be needed. These spatial patterns reveal the complex interplay between geographic, environmental, and economic factors that shape the landscape of NCD mortality in the country.

4.3 Result of influencing factors through machine learning

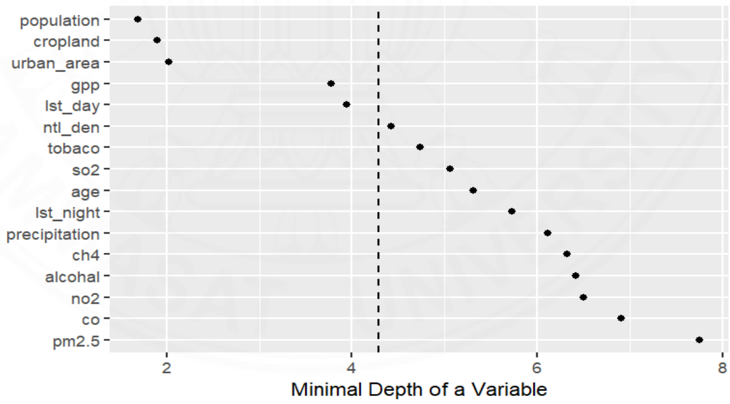
This section presents the findings of our machine learning analysis, focusing on the identification of key influencing factors for various non-communicable diseases (NCDs). Machine learning approach involved constructing a decision tree model, where the number of trees and the depth threshold were optimized to achieve robust results. The Out-of-Bag (OOB) error and accuracy metrics were calculated to assess the model's performance. The OOB error provides an unbiased estimate of the prediction error, while the accuracy metric indicates the proportion of correctly identified influencing factors.

In this analysis, we used a decision tree model to analyze the data, as detailed in Chapter 3. The number of trees, tree depth, and threshold values were systematically varied to identify the configuration that yielded the most accurate results. The dimension of the decision tree, indicated by the depth and number of trees, was critical in capturing the complexity of the influencing factors. The depth threshold determined the granularity of the model, with deeper trees capturing more nuanced interactions between variables.

Figures 4.1-4.4 illustrate the resulting model, highlighting the critical intersections and nodes that were identified as key determinants of NCDs. The results demonstrate the model's effectiveness in delineating the factors that contribute to disease prevalence and distribution, thereby providing valuable insights into the underlying causes of NCDs. The high accuracy of the model, validated by cross-validation techniques, underscores the reliability of these findings.

Figure 4.1

A visual depiction of the associations between the diabetes and the factor (Minimal depth analysis of diabetes)

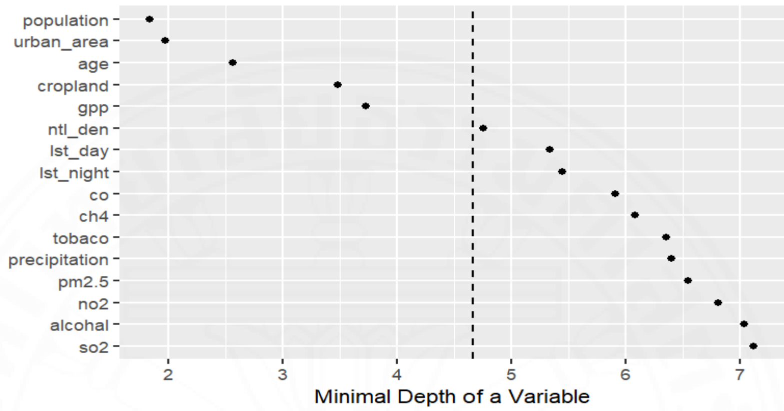


Note: The Author visualized by using R studio.

The diabetes cases, the model was built using a sample size of 213, with 16 dimensions and 500 trees (ntree). The depth threshold was set at 4.29. The model achieved a high accuracy of 99%, indicating its effectiveness in predicting mortality from diabetes based on the provided data, with an R-squared of 0.754, indicating that the model can explain 75.4% of the variation in mortality from diabetes. The minimal depth analysis exposed the most important factors for mortality from diabetes included population density, urban area, GPP, cropland area, nighttime light density (NTL), and land surface temperature (daytime), as presented in Figure 4.1.

Figure 4.2

A visual depiction of the associations between the hypertension and the factor (Minimal depth analysis of diabetes)

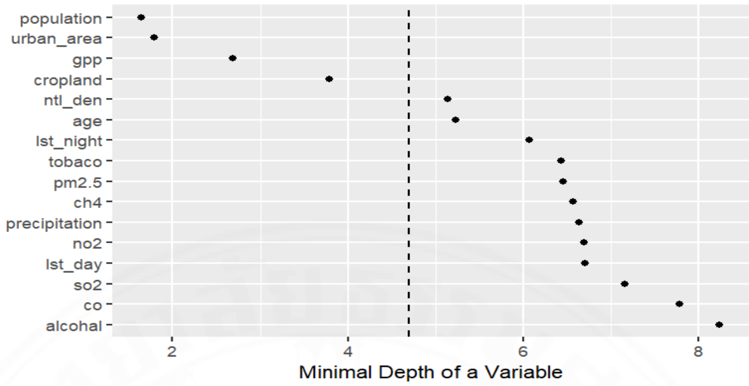


Note: The Author visualized by using R studio.

In the case of hypertension, the model was built using a sample size of 213, with 16 dimensions and 500 trees (ntree). The depth threshold was set at 4.66. The model achieved a high accuracy of 98.6%, indicating its effectiveness in predicting mortality based on the provided data, with an R-squared of 0.636, indicating that the model can explain 63.6% of the variation in mortality from hypertension. The minimal depth analysis revealed that the most influential factors for mortality from hypertension included population, urban area, GPP, cropland area, and NTL, as shown in Figure 4.2.

Figure 4.3

A visual depiction of the associations between the ischemic and the factor (Minimal depth analysis of ischemic)

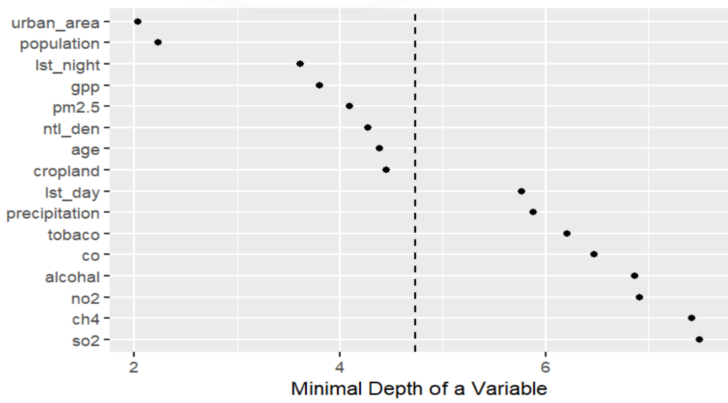


Note: The Author visualized by using R studio.

In the case of ischemic, the model was built using a sample size of 213, with 16 dimensions and 500 trees (ntree). The depth threshold was set at 4.726. The model had a high accuracy rate of 99.6%, which means it was good at predicting death from ischemic disease based on the data given. It also had an R-squared of 0.759, which means it can explain 75.9% of the variation in death from ischemic. The minimal depth analysis revealed that the most influential factors for mortality from ischemic included population, urban area, GPP, cropland area, and nighttime light density (NTL), as shown in Figure 4.3.

Figure 4.4

A visual depiction of the associations between the chronic and the factor (Minimal depth analysis of chronic)



Note: The Author visualized by using R studio.

The model was built using a sample size of 213, with 16 dimensions and 500 trees (ntree). The depth threshold was set at 4.732. Chronic disease with an R-squared of 0.627, the model can explain 62.7% of the variation in mortality from chronic. The model demonstrated a high accuracy rate of 98.81% in predicting the prevalence of death from chronic conditions, based on the provided data. Figure 4.4 shows that the minimal depth analysis showed that population, urban area, GPP, cropland area, nighttime light density (NTL), PM 2.5, land surface temperature (nighttime) and age were the most important factors for chronic mortality.

The machine learning analysis explored mortality patterns across various non-communicable diseases (NCDs), achieving high predictive accuracy rates for each condition. Key influencing factors identified included population density, urbanization, economic activity (GPP), land use (cropland area), nighttime light intensity (NTL), and environmental factors like temperature and air quality (PM 2.5).

4.4 Spatial Distribution of Disease

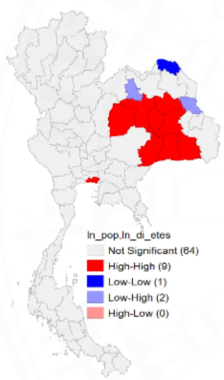
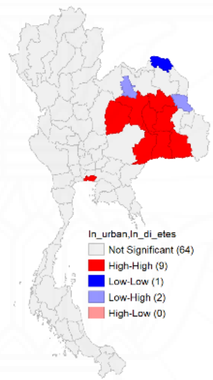
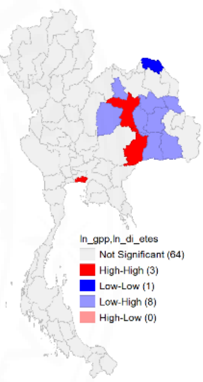
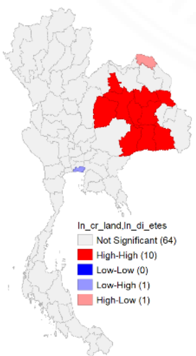
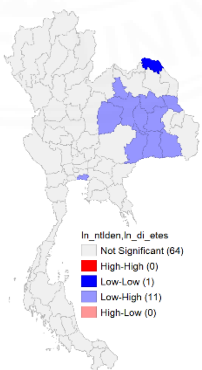
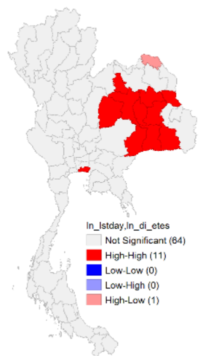
This part goes into more detail about the results of the RF algorithm by looking at both one-way and two-way connections between differences in geography and NCD deaths at the provincial level. We employed the LISA method to verify the spatial association refer in chapter 3. This statistical analysis used a color scheme on the map to illustrate the correlation between influencing factors and NCD mortality. The color scheme categorized correlations into four groups: dark-red, light-red, dark-blue, and light-blue, representing different levels of spatial autocorrelation. However, the study primarily focused on the dark-red scheme, highlighting areas with severe mortality from NCDs that could potentially benefit from public health policy interventions.

4.4.1 Diabetes disease

The outcomes showed the localized correlation between satellite-based indicators: population, urban area, GPP, cropland area, NTL, and daytime. The northeast region, including Nong Bua Lamphu, Khon Kaen, Chaiyaphum, Maha Sarakham, Buri Ram, Surin, Si Sa Ket, Roi Et, Kalasin, and Mukdahan, had the highest incidence rate of diabetes mortality as in Table 4.3

Table 4.3

Displays LISA's cluster map, which shows the localized association between influencing factors and diabetes.

Population (Moran's I=0.005)	Urban area (Moran's I=0.094)	GPP (Moran's I=-0.008)
		
Cropland area (Moran's I=0.095)	Nighttime light density (NTL) (Moran's I=0.006)	Land surface temperature (Daytime) (Moran's I=0.019)
		

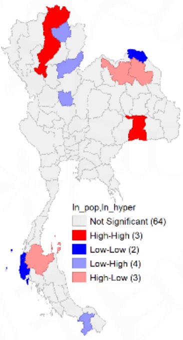
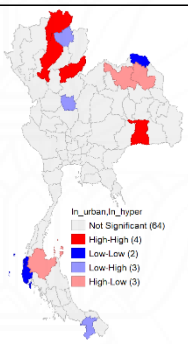
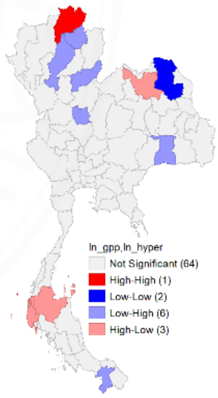
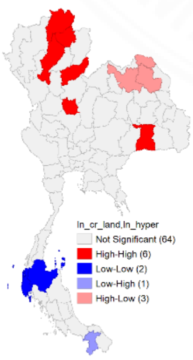
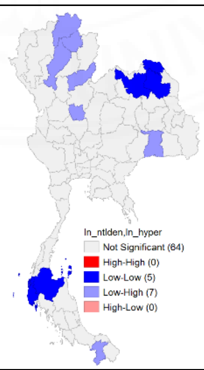
Note: The Author visualized by using Geoda program.

4.4.2 Hypertension disease

The outcomes showed a localized correlation between satellite-based indicators such as population, urban area, GPP, cropland area, and NTL. Some province regions, including Chiang Rai, Lampang, Uttaradit, and Phichit, showed a high incidence rate of hypertension mortality. Sakon Nakhon, Udon Thani, and Surin were in the northeast, while Surat Thani and Phang Nga were in the southeast. Table 4.4 displays the results.

Table 4.4

LISA's cluster map, which demonstrates the localized association between influencing factors and hypertension.

<table>
<tr><td>Population
(Moran’s I=0.041)</td><td>Urban area
(Moran’s I=0.086)</td><td>GPP
(Moran’s I=0.011)</td></tr>
<tr><td>
</td><td>
</td><td rowspan="3">
</td></tr>
<tr><td>Cropland area
(Moran’s I=0.029)</td><td>Nighttime light density
(NTL) (Moran’s I=0.031)</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

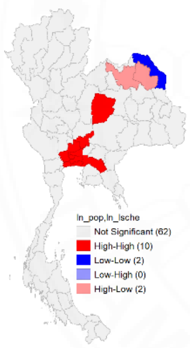
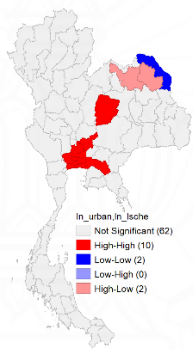
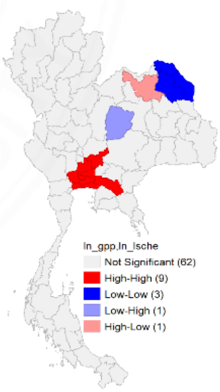
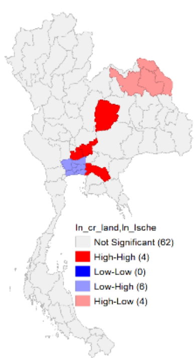
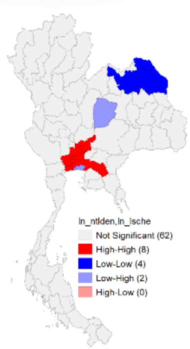
Note: The Author visualized by using Geoda program.

4.4.3 Ischemic disease

The outcomes showed a localized correlation between satellite-based indicators such as population, urban area, GPP, cropland area, and NTL. The center region, including Bangkok, Nonthaburi, Pathum Thani, Phra Nakhon Si Ayutthaya, Samut Prakan, Samut Sakhon, Chachoengsao, Nakhon Pathom, and Saraburi, had the highest incidence rate of ischemic mortality. Table 4.5 exhibits the results in several provinces in the northeast region, including Sakon Nakhon, Nakhon Phanom, and Udon Thani.

Table 4.5

LISA's cluster map showing the localized association between influencing factors and ischemic

<table>
<tr><td>Population
(Moran's I=0.094)</td><td>Urban area
(Moran's I=0.157)</td><td>GPP
(Moran's I=0.146)</td></tr>
<tr><td>
</td><td>
</td><td rowspan="3">
</td></tr>
<tr><td>Cropland area
(Moran's I=-0.081)</td><td>Nighttime light density
(NTL) (Moran's I=0.176)</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

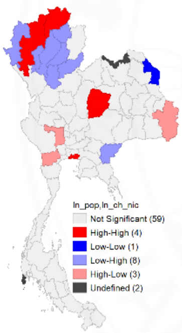
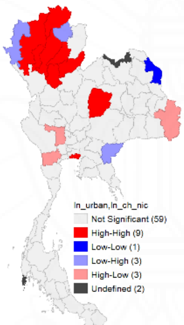
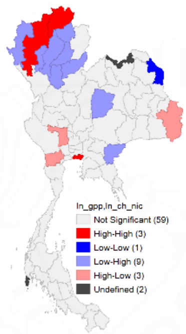
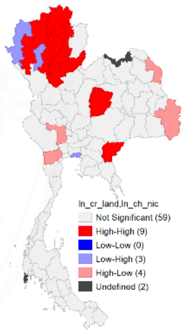
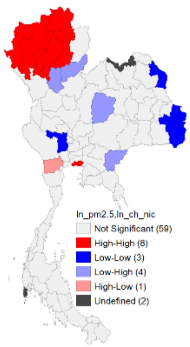
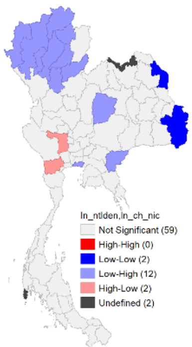
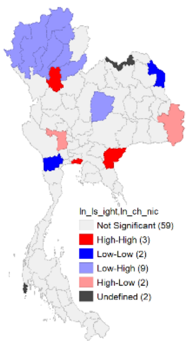
Note: The Author visualized by using Geoda program.

4.4.4 Chronic disease

The chronic mortality incidence rate was found in the mostly in northern region, including Mae Hong Son, Chiang Mai, Lamphun, Lampang, Uttaradit, Chiang Rai, Phayao, and some provinces from the center and northeast, such as Bangkok, Suphanburi, Ratchaburi Chaiyaphum, and Ubon Ratchathani, had the highest incidence rate of chronic mortality. The outcomes showed a localized correlation between satellite-based indicators: population, urban area, GPP, cropland area, NTL, PM 2.5, and nighttime. Table 4.6 exhibits the results.

Table 4.6

LISA's cluster map showing the localized association between influencing factors and ischemic

<table>
<tr><td>Population
(Moran's I=0.017)</td><td>Urban area
(Moran's I=0.118)</td><td>GPP
(Moran's I=0.029)</td><td>Cropland area
(Moran's I=0.011)</td></tr>
<tr><td>
</td><td>
</td><td>
</td><td>
</td></tr>
<tr><td>PM 2.5
(Moran's I=0.219)</td><td>Nighttime light density (NTL)
(Moran's I=-0.041)</td><td colspan="2">Land Surface Temperature (Nighttime)
(Moran's I=-0.001)</td></tr>
<tr><td>
</td><td>
</td><td colspan="2">
</td></tr>
</table>

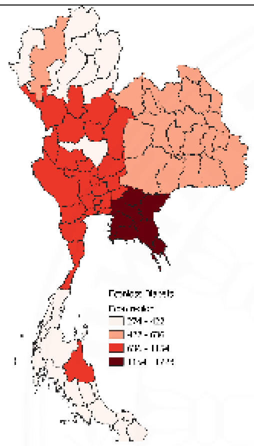
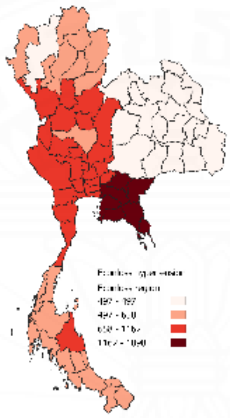
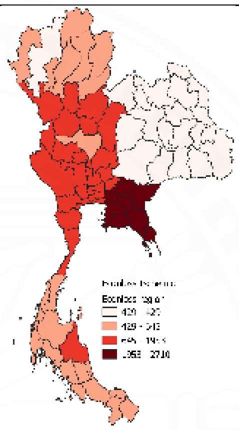
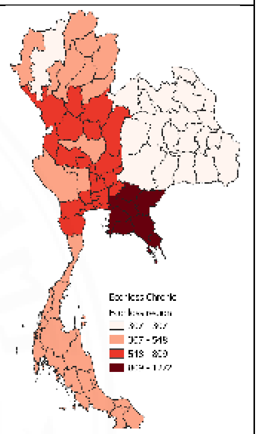
Note: The Author visualized by using Geoda program.

4.5 Economic loss from NCDs

The results derived from QGIS mapping highlight pronounced economic disparities linked to diabetes, hypertension, ischemic diseases, and chronic conditions across diverse regions within the study area, as illustrated in Table 4.7.

Table 4.7

The economic losses from NCDs across various regions 2018-2021

Diabetes	Hypertension	Ischemic	Chronic
			

Note: The Author visualized by using QGIS program.

In the Northern region, the economic loss attributed to diabetes is relatively low compared to other regions, while hypertension contributes significantly to the overall economic burden. Ischemic diseases and chronic conditions also impose notable economic losses in this region.

Moving to the Northeastern region, the economic burden of diabetes is higher compared to the Northern region, while ischemic diseases result in comparatively lower economic losses. Hypertension and chronic conditions also play significant roles in driving economic losses in this area.

In the Central region, all four types of diseases contribute substantially to the economic burden, with ischemic diseases causing the highest economic loss among the studied conditions. The Central region experiences the highest overall economic losses compared to other regions. In the Eastern region, economic losses attributed to all four types of diseases are notably high, with ischemic diseases causing the most substantial economic burden. Hypertension and chronic conditions also contribute significantly to the overall economic losses in this region.

In the Western region, economic losses due to hypertension are relatively lower compared to other regions, while ischemic diseases result in the highest economic losses. Diabetes and chronic conditions also contribute significantly to the economic burden in this area.

Lastly, in the Southern region, diabetes and hypertension contribute substantially to the economic losses, while ischemic diseases result in the highest economic loss among the studied conditions. Chronic conditions also impose notable economic burdens in this region.

Overall, the QGIS mapping provides valuable insights into the regional variations in economic losses associated with different types of NCDs, highlighting areas that may require targeted interventions and resource allocation to mitigate the economic impact of these health conditions.

4.5.1 Diabetes disease

The economic loss affecting the GPP due to diabetes was analyzed using the Mincer Equation (Equation (3.8)). Table 4.8 presents the regression results for the years 2018-2021.

Table 4.8

Regression result of economic loss from diabetes disease

Variable	Model			
	(1)	(2)	(3)	(4)
Economic loss	0.198***	0.179***	0.173***	0.192***
	(8.17)	(7.35)	(6.95)	(7.03)
Year dummies	Not Included	Not Included	Included	Included
Controls	Not Included	Included	Included	Included
Interaction Terms	Not Included	Not Included	Not Included	Included
Fixed effect test (F-stat)	135.50***	105.19***	111.28***	111.89***
R-squared	0.227	0.347	0.394	0.408
R squared change test (Chi2-stat)		49.51***	22.37***	7.09

Notes: standard error in parentheses; *p<0.1, **p<0.05, ***p<0.01

The results suggest that All Model show a positive and highly significant coefficient for economic loss, indicating a strong relationship between economic loss due to diabetes and the dependent variable in all Model. The coefficient estimates for the economic loss from diabetes across all Model are statistically significant at the 1% level, indicating that an increase in the economic loss from diabetes is associated with a significant change in GPP.

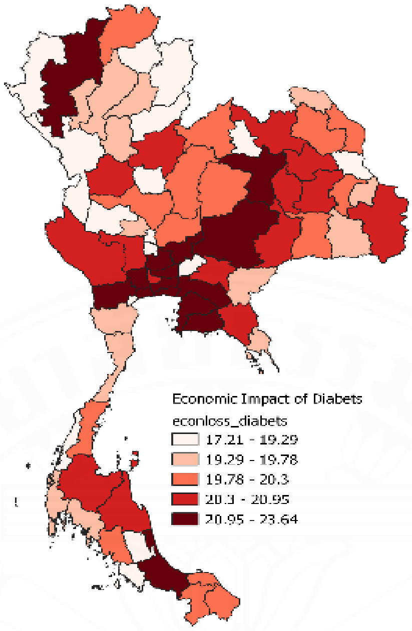
To analyze the economic loss due to diabetes and its impact on the GPP, the regression results from Table 4.8 are considered, with a particular focus on Model (3) as the most appropriate model. The selection of Model (3) is substantiated by the R-squared change test, which is significant and aligns with the likelihood ratio (LR) test as discussed in Chapter 3. n Model (3), the coefficient is 0.173 with a t-value of 6.95, indicating a strong and significant relationship between diabetes-related economic loss and the GPP.

This suggests that as the economic loss due to diabetes increases, there is a corresponding negative impact on the GPP. Model (3) yields an R-squared value of 0.394, indicating that approximately 39.4% of the variation in GPP can be explained by the variables included in the model. The R-squared change test is significant when moving from Model (2) to Model (3), highlighting the improved explanatory power of Model (3) compared to the previous Model. This result corroborates the selection of Model (3) as the most appropriate, as it provides a more accurate depiction of the economic impact of diabetes on GPP without overfitting or losing interpretability.

Aside from the regression results, Figure 4.5 illustrates the spatial distribution of economic losses, highlighting regional disparities and areas with significant economic impacts. To provide further support, Table 4.9 emphasizes the economic loss due to diabetes mortality across the top 10 provinces in Thailand, identifying which provinces are most urgent and important for targeted interventions.

Figure 4.5

Area of economic loss from diabetes



Notes: The author visualized using the QGIS program.

Figure 4.5 illustrates the economic loss from diabetes mortality across various provinces in Thailand. This analysis accounts for the economic loss value due to life expectancy at the age of death from diabetes and the GPP per capita to estimate the economic impact. The dark red areas on the map represent regions with the highest economic losses, predominantly located in the central and eastern parts of Thailand.

Table 4.9

Listed of economic loss from diabetes by province

Province	Economic loss (Million Bath)
Bangkok Metropolis	19,111.40
Samut Prakan	5,549.73
Chon buri	5,086.77
Rayong	4,092.60
Samut sakhon	2,236.16
Chachoengsao	2,186.64
Phra nakhon si ayutthaya	2,068.98
Nakhon pathom	1,980.63
Khon kaen	1,775.12
Pathum thani	1,510.34

Notes: The author's calculation from Equation (3.7).

Table 4.9 emphasizes the economic loss due to diabetes mortality across the top 10 provinces in Thailand. Bangkok Metropolis emerges as the province with the highest economic loss, totaling 19,111.40 million Baht. This staggering Figure underscores the significant financial impact of diabetes within the capital region.

Samut Prakan and Chon Buri follow closely behind, with economic losses of 5,549.73 million Baht and 5,086.77 million Baht, respectively. These provinces exhibit substantial economic burdens associated with diabetes-related healthcare costs and productivity losses.

Other notable provinces with considerable economic losses from diabetes includes Rayong, Samut Sakhon, and Chachoengsao, highlighting the widespread nature of the issue across different regions of Thailand.

4.5.2 Hypertension disease

The economic loss affecting the GPP due to hypertension was analyzed using the Mincer Equation (Equation (3.8)). Table 4.10 presents the regression results for the years 2018-2021.

Table 4.10

Regression result of economic loss from hypertension disease

Variable	Model			
	(1)	(2)	(3)	(4)
Economic loss	0.143***	0.133***	0.126***	0.124***
	(6.89)	(6.92)	(6.53)	(6.17)
Year dummies	Not Included	Not Included	Included	Included
Controls	Not Included	Included	Included	Included
Interaction Terms	Not Included	Not Included	Not Included	Included
Fixed effect test (F-stat)	256.26***	159.60***	165.64***	164.22***
R-squared	0.173	0.333	0.366	0.370
R squared change test (Chi2-stat)		64.46***	15.35***	2.17

Notes: standard error in parentheses; *p<0.1, **p<0.05, ***p<0.01

The results suggest that All Model show a positive and highly significant coefficient for economic loss, indicating a strong relationship between economic loss due to hypertension and the dependent variable in all Model. The coefficient estimates for the economic loss from hypertension across all Model are statistically significant at the 1% level, indicating that an increase in the economic loss from hypertension is associated with a significant change in GPP.

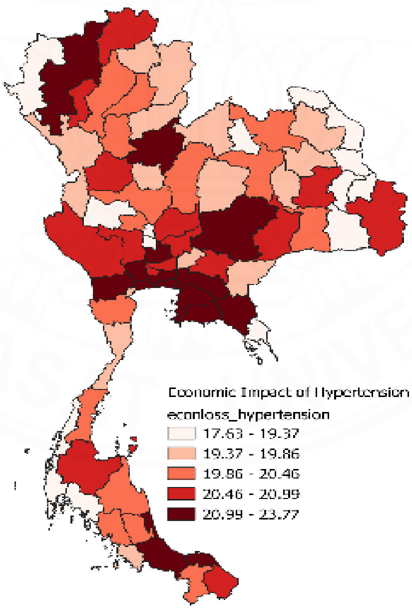
To analyze the economic loss due to hypertension and its impact on the GPP, the regression results from Table 4.10 are considered, with a particular focus on Model (3) as the most appropriate model. The selection of Model (3) is substantiated by the R-squared change test, which is significant and aligns with the likelihood ratio (LR) test as discussed in Chapter 3.

In Model (3), the coefficient for economic loss due to hypertension is 0.126 with a t-value of 6.53, indicating a strong and significant relationship between hypertension-related economic loss and GPP. This suggests that an increase in economic loss due to hypertension is associated with a corresponding negative impact on GPP, reflecting the economic burden that hypertension imposes on provincial economies. Model (3) yields an R-squared value of 0.366, indicating that approximately 36.6% of the variation in GPP can be explained by the variables included in the model. The R-squared change test is significant when moving from Model (2) to Model (3), highlighting the improved explanatory power of Model (3) compared to the previous Model. This improvement underscores the importance of accounting for year-specific effects and control variables in capturing the economic impact of hypertension on GPP.

Aside from the regression results, Figure 4.6 illustrates the spatial distribution of economic losses, highlighting regional disparities and areas with significant economic impacts. To provide further support, Table 4.11 emphasizes the economic loss due to hypertension mortality across the top 10 provinces in Thailand, identifying which provinces are most urgent and important for targeted interventions.

Figure 4.6

Area of economic loss from hypertension



Notes: The author visualized using the QGIS program.

Figure 4.6 illustrates the economic loss from hypertension mortality across various provinces in Thailand. This analysis accounts for the economic loss value due to life expectancy at the age of death from hypertension and the GPP per capita to estimate the economic impact. The dark red areas on the map represent regions with the highest economic losses, predominantly located in the rural province of Thailand.

Table 4.11

Listed of economic loss from hypertension by province

Province	Economic loss
Bangkok Metropolis	15,986.33
Chon buri	5,278.61
Rayong	4,354.37
Samut Prakan	4,319.76
Phra nakhon si ayutthaya	3,273.25
Chiang mai	2,948.21
Ratchaburi	2,394.41
Chachoengsao	2,097.79
Prachin buri	1,689.19
Pattani	1,573.87

Notes: The author's calculation from Equation (3.7).

Table 4.11 emphasizes the economic loss due to hypertension mortality across the top 10 provinces in Thailand. Once again, Bangkok Metropolis leads the list with an economic loss of 15,986.33 million Baht attributed to hypertension. This underscores the significant financial strain imposed by hypertension-related healthcare expenditures and productivity losses within the capital.

Chon Buri and Rayong follow suit, with economic losses of 5,278.61 million Baht and 4,354.37 million Baht, respectively. These provinces exhibit substantial economic burdens associated with hypertension, reflecting the need for targeted interventions to mitigate the impact.

Noteworthy provinces such as Samut Prakan, Phra Nakhon Si Ayutthaya, and Chiang Mai also feature prominently on the list, indicating the widespread prevalence and economic ramifications of hypertension across different regions.

4.5.3 Ischemic disease

The economic loss affecting the GPP due to hypertension was analyzed using the Mincer Equation (Equation 3.8). Table 4.12 presents the regression results for the years 2018-2021.

Table 4.12

Regression result of economic loss from ischemic disease.

Variable	Model			
	(1)	(2)	(3)	(4)
Economic loss	0.388***	0.352***	0.345***	0.342***
	(11.86)	(10.42)	(10.50)	(10.35)
Year dummies	Not Included	Not Included	Included	Included
Controls	Not Included	Included	Included	Included
Interaction Terms	Not Included	Not Included	Not Included	Included
Fixed effect test (F-stat)	85.16	55.21	58.70	58.45
R-squared	0.383	0.455***	0.494***	0.500
R squared change test (Chi2-stat)		29.86	22.87	3.54

Notes: standard error in parentheses; **p*<0.1, ***p*<0.05, ****p*<0.01

The results suggest that All Model show a positive and highly significant coefficient for economic loss, indicating a strong relationship between economic loss due to ischemic and the dependent variable in all Model. The coefficient estimates for the economic loss from ischemic across all Model are statistically significant at the 1% level, indicating that an increase in the economic loss from ischemic is associated with a significant change in GPP.

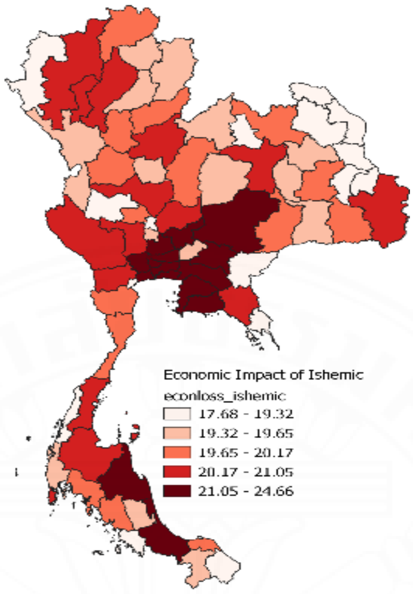
To analyze the economic loss due to ischemic disease and its impact on the GPP, the regression results from Table 4.12 are considered, with a particular focus on Model (3) as the most appropriate model. The selection of Model (3) is substantiated by the R-squared change test, which is significant and aligns with the likelihood ratio (LR) test as discussed in Chapter 3.

In Model (3), the coefficient for economic loss due to ischemic disease is 0.345 with a t-value of 10.50, indicating a strong and significant relationship between ischemic-related economic loss and GPP. This suggests that an increase in economic loss due to ischemic disease corresponds to a notable negative impact on GPP, reflecting the substantial economic burden that ischemic disease imposes on provincial economies.

Model (3) yields an R-squared value of 0.494, indicating that approximately 49.4% of the variation in GPP can be explained by the variables included in the model. The R-squared change test is significant when moving from Model (2) to Model (3), highlighting the enhanced explanatory power of Model (3) compared to previous Model. This significance underscores the importance of incorporating year-specific effects and control variables to capture the economic impact of ischemic disease on GPP effectively.

Figure 4.7

Area of economic loss from ischemic



Notes: The author visualized using the QGIS program.

Aside from the regression results, Figure 4.7 illustrates the spatial distribution of economic losses, highlighting regional disparities and areas with significant economic impacts. To provide further support, Table 4.13 emphasizes the economic loss due to ischemic mortality across the top 10 provinces in Thailand, identifying which provinces are most urgent and important for targeted interventions.

Figure 4.7 illustrates the economic loss from ischemic mortality across various provinces in Thailand. This analysis accounts for the economic loss value due to life expectancy at the age of death from ischemic and the GPP per capita to estimate the economic impact. The dark red areas on the map represent regions with the highest economic losses, predominantly located in the central parts of Thailand.

Table 4.13

Listed of economic loss from ischemic by province

Province	Economic loss
Bangkok Metropolis	52,039.38
Chon buri	7,765.17
Rayong	7,447.82
Samut Prakan	6,233.86
Phra nakhon si ayutthaya	4,108.43
Pathum thani	3,400.49
Nakhon pathom	3,041.46
Nonthaburi	2,666.76
Chachoengsao	2,524.01
Samut sakhon	2,512.59

Notes: The author's calculation from Equation (3.7).

Table 4.13 emphasizes the economic loss due to ischemic mortality across the top 10 provinces in Thailand. Bangkok Metropolis stands out once again, with an alarming economic loss of 52,039.38 million Baht attributed to ischemic diseases. This staggering Figure underscores the significant financial impact of ischemic conditions within the capital region. Chon Buri and Rayong follow closely behind, with economic losses of 7,765.17 million Baht and 7,447.82 million Baht, respectively. These provinces exhibit substantial economic burdens associated with ischemic diseases, highlighting the urgent need for targeted interventions to address the issue.

Other provinces such as Samut Prakan, Phra Nakhon Si Ayutthaya, and Pathum Thani also feature prominently on the list, indicating the widespread nature of ischemic diseases across different regions of Thailand.

4.5.4 Chronic disease

The economic loss affecting the GPP due to hypertension was analyzed using the Mincer Equation (Equation 3.8). Table 4.14 presents the regression results for the years 2018-2021.

Table 4.14

Regression result of economic loss from chronic disease

Variable	Model			
	(1)	(2)	(3)	(4)
Economic loss	0.0976***	0.0938***	0.0984***	0.0930**
	(4.19)	(3.91)	(3.64)	(3.31)
Year dummies	Not Included	Not Included	Included	Included
Controls	Not Included	Included	Included	Included
Interaction Terms	Not Included	Not Included	Not Included	Included
Fixed effect test (F-stat)	180.50***	74.93***	76.10***	75.80***
R-squared	0.072	0.281	0.308	0.316
R squared change test (Chi2-stat)		76.70***	11.58***	3.10

Notes: standard error in parentheses; **p*<0.1, ***p*<0.05, ****p*<0.01

The results suggest that All Model show a positive and highly significant coefficient for economic loss, indicating a strong relationship between economic loss due to chronic and the dependent variable in all Model. The coefficient estimates for the economic loss from chronic across all Model are statistically significant at the 1% level, indicating that an increase in the economic loss from chronic is associated with a significant change in GPP.

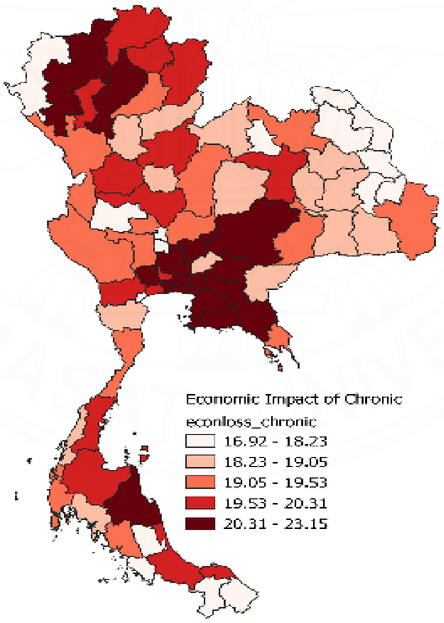
To analyze the economic loss due to chronic disease and its impact on the GPP, the regression results from Table 4.14 are considered, with a particular focus on Model (3) as the most appropriate model. The selection of Model (3) is substantiated by the R-squared change test, which is significant and aligns with the likelihood ratio (LR) test as discussed in Chapter 3.

In Model (3), the coefficient for economic loss due to chronic disease is 0.0984 with a t-value of 3.64, indicating a strong and significant relationship between chronic disease-related economic loss and the GPP. This suggests that an increase in economic loss due to chronic diseases is associated with a notable negative impact on GPP, underscoring the economic burden that chronic diseases place on provincial economies. Model (3) yields an R-squared value of 0.308, indicating that approximately 30.8% of the variation in GPP can be explained by the variables included in the model. The R-squared change test is significant when moving from Model (2) to Model (3), highlighting the improved explanatory power of Model (3) compared to previous Model. This significance underscores the value of including year-specific effects and control variables, which enhance the model's ability to accurately capture the economic impact of chronic diseases on GPP.

Aside from the regression results, Figure 4.8 illustrates the spatial distribution of economic losses, highlighting regional disparities and areas with significant economic impacts. To provide further support, Table 4.15 emphasizes the economic loss due to chronic mortality across the top 10 provinces in Thailand, identifying which provinces are most urgent and important for targeted interventions.

Figure 4.8

Area of economic loss from chronic



Notes: The author visualized using the QGIS program.

Figure 4.8 illustrates the economic loss from chronic mortality across various provinces in Thailand. This analysis accounts for the economic loss value due to life expectancy at the age of death from chronic and the GPP per capita to estimate the economic impact. The dark red areas on the map represent regions with the highest economic losses, predominantly located in the northern and central parts of Thailand

Table 4.15

Listed of economic loss from chronic by province

Province	Economic loss
Bangkok Metropolis	13,460.35
Rayong	3,073.21
Chon buri	2,828.66
Samut Prakan	2,365.90
Chiang mai	2,188.65
Chachoengsao	1,462.86
Pathum thani	1,361.74
Nakhon pathom	1,309.34
Saraburi	1,209.03
Samut sakhon	1,168.17

Notes: The author's calculation from Equation (3.7).

Table 4.15 emphasizes the economic loss due to chronic mortality across the top 10 provinces in Thailand. Bangkok Metropolis continues to dominate the list, with an economic loss of 13,460.35 million Baht attributed to chronic conditions. This underscores the significant financial strain imposed by chronic diseases within the capital city.

Rayong and Chon Buri follow suit, with economic losses of 3,073.21 million Baht and 2,828.66 million Baht, respectively. These provinces exhibit substantial economic burdens associated with chronic diseases, reflecting the need for targeted interventions to mitigate the impact.

Other noteworthy provinces such as Samut Prakan, Chiang Mai, and Chachoengsao also feature prominently on the list, indicating the widespread prevalence and economic ramifications of chronic diseases across different regions.

4.6 Discussion

Overall, the result reports that environmental and economic factors affected NCDs mortality. The RF minimal mortality was influenced by various factors such as population, urban area, GPP, cropland area, and nighttime light density (NTL). These factors varied depending on the specific type of NCDs disease, such as diabetes, which is associated with a hot zone. Therefore, the factor was related to the island surface temperature during the day.

Moreover, the chronic effects of pollution and urbanization make the PM and NTL significant. Similar to the main findings of Chakravorty and Heath and Hantrakun et al., the present work indicated that the provinces with hotspot clusters were primarily located in the northeast because their physical conditions and climate are agricultural land and drought. Environmental factors such as rainfall, vegetation, temperature, and climatic variations particularly influence the highest diabetes mortality incidence rate.

Additionally, we employed spatial analysis to statistically quantify the extent of NCDs mortality across Thailand. The results revealed variations in the NCDs mortality rate across different regions of the country. The northeastern region, including Nong Bua Lamphu, Khon Kaen, Chaiyaphum, Maha Sarakham, Buri Ram, Surin, Si Sa Ket, Roi Et, Kalasin, and Mukdahan, exhibited the highest concentration of diabetes mortality.

The outcomes of bivariate LISA showed the localized correlation between satellite-based indicators: population, urban area, GPP, cropland area, nighttime light density (NTL), and land surface temperature (daytime). The results suggested that the cultivation area, as indicated by the cropland index, played a significant role in this trend. Diabetes prevalence is severe in rural areas, particularly in the northeast region, where farmers often face low incomes and food insecurity. Conversely, areas with lower concentrations of nighttime light or lower socioeconomic development were also likely to experience higher rates of diabetes.

Savage et al. (2021) proposed the DR-NCDs risk framework, highlighting how structural drivers like climate change can exacerbate dietary-related NCDs. In contrast, Bangkok, with its sprawling urban areas and influx of migrant workers, emerged as a hotspot for NCDs mortality. This was consistent with findings by Moon (2021), indicating that heat waves increase mortality and morbidity risks for diabetes patients. During such events, mortality rates among diabetes patients may rise by approximately 18%, with a corresponding 10% increase in overall morbidity.

The incidence rate of hypertension-related mortality was higher in certain provinces, such as Chiang Rai, Lampang, Uttaradit, and Phichit in the northern region, and Sakon Nakhon, Udon Thani, and Surin in the northeast. Furthermore, in the southeast region, Surat Thani and Phang Nga showed elevated rates. Bivariate LISA analysis revealed a localized correlation between satellite-based indicators, including population density, urban area, GPP, cropland area, and nighttime light density (NTL).

Similarly, the incidence of ischemic mortality was highest in central regions such as Bangkok, Nonthaburi, Pathum Thani, Phra Nakhon Si Ayutthaya, Samut Prakan, Samut Sakhon, Chachoengsao, Nakhon Pathom, and Saraburi. Some provinces in the northeast, such as Sakon Nakhon and Nakhon Phanom, also exhibited high rates. The urban area index appeared to play a significant role, as rural areas showed lower prevalence, detection, and treatment rates of hypertension compared to urban areas. This aligns with findings by Ratovoson et al. (2015), indicating that hypertension was more prevalent in rural areas yet significantly less treated.

Consequently, there is a growing risk of a major epidemic of cardiovascular diseases in Madagascar's aging society. The incidence rate of chronic-related mortality was highest in the northern region, including provinces like Mae Hong Son, Chiang Mai, Lamphun, Lampang, Uttaradit, Chiang Rai, and Phayao. Additionally, some provinces in the central and northeastern regions, such as Bangkok, Suphanburi, Ratchaburi, Chaiyaphum, and Ubon Ratchathani, also showed elevated rates. Satellite-based indicators like population density, urban area, GPP, cropland area, nighttime light density (NTL), PM 2.5, and land surface temperature (Nighttime) were found to be locally correlated.

The presence of PM 2.5 suggests air pollution, particularly in the northern region, which often experiences pollution from forest fires.

Overall, the data across these Tables consistently highlight the significant economic burdens imposed by diabetes, hypertension, ischemic diseases, and chronic conditions in Bangkok Metropolis and other provinces. These findings underline the urgent need for targeted interventions and policies to address the widespread prevalence and financial ramifications of NCDs across Thailand. The analysis suggests that economic loss due to diabetes significantly impacts provincial economies in Thailand. As economic losses increase, there is a notable adverse effect on provincial economic performance. All Model show a positive and highly significant coefficient for economic loss, indicating a strong relationship between economic loss due to NCDs and the dependent variable in all Model.

This underscores the importance of addressing NCDs as a critical public health and economic issue, necessitating interventions to reduce its prevalence and mitigate its economic impact. Policymakers should consider the broader economic consequences of diabetes and allocate resources to healthcare initiatives that can reduce the prevalence and severity of diabetes-related complications.

CHARPTER 5
CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusion

The study identified high-risk clusters, as well as the spatial dynamics of NCDs mortality and its determinants. These findings highlight the complex interplay between environmental, economic, and demographic factors that influence the prevalence of NCDs mortality in Thailand. The results indicated significant implications for public health policy and intervention strategies aimed at addressing the risk factors associated with each disease. By understanding and targeting these factors, policymakers can develop more effective strategies to promote public health and well-being, ultimately reducing the burden of NCDs in the country.

This study has shown that both environmental and economic factors significantly impact NCDs mortality in Thailand. Key findings reveal that diabetes mortality is highly concentrated in the northeastern provinces such as Nong Bua Lamphu, Khon Kaen, and Chaiyaphum, where agricultural land and drought conditions significantly influence the prevalence of the disease. Environmental factors like rainfall, vegetation, temperature, and climatic variations are crucial in understanding the high incidence of diabetes mortality, with a notable correlation to land surface temperature during the day.

In terms of hypertension mortality, provinces in the north, such as Chiang Rai and Lampang, and in the northeast, such as Sakon Nakhon and Udon Thani, exhibit higher rates. Urbanization plays a significant role in these trends, with urban areas showing higher prevalence and detection rates. Ischemic mortality is most prominent in the central provinces, including Bangkok and Nonthaburi, which correlates with higher levels of economic activity and stress. Chronic disease mortality is particularly high in northern provinces like Chiang Mai and Lampang, where air pollution from forest fires significantly contributes to these mortality rates.

Overall, the data consistently demonstrate the significant economic impact of NCDs conditions on Bangkok Metropolis and other provinces. These findings emphasize the urgent need for targeted interventions and policies to address the widespread prevalence and financial burden of NCDs across Thailand.

The ongoing high economic losses due to NCDs indicate that current public health strategies may be inadequate, necessitating a reassessment and enhancement of health policies, prevention programs, and resource allocation to effectively combat the increasing prevalence of NCDs. Economic impacts are also evident, as diabetes-related economic losses adversely affect provincial economic performance. The correlation between economic loss and NCDs prevalence highlights the need for comprehensive public health and economic strategies.

5.2 Policy Recommendation

To mitigate the impact of NCDs and improve public health outcomes, several policy recommendations are proposed.

5.2.1 Integrated Health and Environmental Policies

Integrated health and environmental policies should address climate change and its health impacts by improving air quality, managing heat waves, and ensuring sustainable agricultural practices. Enhanced monitoring of environmental factors like air pollution and land surface temperature can guide targeted interventions in high-risk areas.

5.2.1.1 Develop and implement policies that address climate change impacts on health, focusing on improving air quality standards, managing heat waves, and promoting sustainable agricultural practices. These measures are crucial in reducing the environmental factors that contribute to NCDs.

5.2.1.2 Enhance monitoring systems for environmental factors such as air pollution (e.g., nitrogen dioxide, carbon monoxide) and land surface temperature. These data can guide targeted interventions in areas where environmental conditions exacerbate NCDs prevalence.

5.2.2 Preventive Healthcare Initiatives

Public health efforts should focus on preventive healthcare, particularly in rural and high-risk areas, through regular health screenings, public awareness campaigns, and lifestyle modification programs. Improving access to healthcare in rural areas is crucial, involving the construction of more healthcare facilities, training healthcare workers, and providing telemedicine services to remote regions. Economic and social support mechanisms are essential. Financial assistance should be provided to individuals and families affected by NCDs, including subsidies for medical expenses and nutritional support programs.

5.2.2.1 Prioritize preventive healthcare efforts in rural and high-risk areas through regular health screenings, comprehensive public awareness campaigns about NCDs risk factors, and initiatives promoting healthy lifestyle choices.

5.2.2.2 Improve access to healthcare services in rural regions by expanding infrastructure, training healthcare professionals to address NCDs prevention and management, and implementing telemedicine services to reach remote populations effectively.

5.2.3 Economic and Social Support Measures:

5.2.3.1 Establish financial assistance programs to support individuals and families affected by NCDs, including subsidies for medical expenses and nutritional support programs aimed at managing chronic conditions.

5.2.3.2 Implement comprehensive social protection policies addressing socio-economic determinants of health. These policies should include measures to improve education levels, create employment opportunities, and enhance living conditions in disadvantaged communities.

5.2.4 Urban and Rural Development Strategies

Urban planning should promote sustainable development that minimizes environmental degradation and health risks. This includes creating green spaces, improving public transportation, and reducing urban pollution. Investment in rural development programs can enhance quality of life and economic opportunities, reducing migration to urban areas and alleviating pressure on urban health services.

5.2.4.1 Promote sustainable urban development practices that minimize environmental degradation and health risks. This includes initiatives to develop green spaces, improve public transportation systems to reduce reliance on private vehicles, and mitigate urban pollution.

5.2.4.2 Invest in rural development programs to enhance quality of life and economic opportunities in rural areas. By improving living standards and economic prospects, these initiatives can reduce migration to urban centers and alleviate pressure on urban health services.

Ongoing research is vital to understand the complex relationship between environmental factors and NCDs. This includes studying the impact of climate change on health and developing new interventions. Utilizing data from environmental and health monitoring systems to inform policy decisions ensures evidence-based and targeted interventions. Support continuous research to deepen understanding of the complex interplay between environmental factors and NCDs prevalence. Focus research efforts on investigating the specific impacts of climate change on health outcomes and developing innovative interventions tailored to local contexts. Utilize data obtained from robust environmental and health monitoring systems to inform evidence-based policy decisions. This approach ensures that interventions are targeted, effective, and responsive to evolving health challenges posed by NCDs.

REFERENCES

Ahmad, M. W., Reynolds, J., & Rezgui, Y. (2018). Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. Journal of Cleaner Production, 203, 810-821. https://doi.org/10.1016/j.jclepro.2018.08.20.

Bloom, D., & Canning, D. (2003). Health as human capital and its impact on economic performance. The Geneva Papers on Risk and Insurance: Issues and Practice, 28(2), 304-315. https://www.jstor.org/stable/41952692.

Bukhman, G., Mocumbi, A. O., & Horton, R. (2015). Reframing NCDs and injuries for the poorest billion: A Lancet Commission. The Lancet, 386(10000), 1221-1222. https://doi.org/10.1016/S0140-6736(15)00278-0

Chen, S., Kuhn, M., Prettner, K., Bloom, D. E., & Husain, M. J. (2018). The macroeconomic burden of noncommunicable diseases in the United States: Estimates and projections. PLOS ONE, 13(11), e0206702. https://doi.org/10.1371/journal.pone.0206702.

Engelgau, M., Rosenhouse, S., El-Saharty, S., & Mahal, A. (2011). The economic effect of noncommunicable diseases on households and nations: A review of existing evidence. Journal of Health Communication, 16(Suppl 2), 75-81. https://doi.org/10.1080/10810730.2011.601394.

Erdil, E., & Yetkiner, I. H. (2009). The Granger-causality between health care expenditure and output: A panel data approach. Applied Economics, 41(4), 511-518. https://doi.org/10.1080/00036840601019083.

Jack, W., Lewis, M., & Aceso Global. (2009). Health investments and economic growth: Macroeconomic evidence and microeconomic foundations. Policy Research Working Paper. Retrieved from RePEc.

Jain, S., Jain, V., Jain, S., et al. (2019). Prevalence of modifiable risk factors for non-communicable diseases in urban slum: A cross-sectional study using WHO STEPS approach. International Journal of Community Medicine and Public Health, 6(4), 1565-1572. https://doi.org/10.18203/2394-6040.ijcmph20191385

Krishnan, G., Singh, S., Pathania, M., Gosavi, S., Abhishek, S., & Parchani, A. (2023). Artificial intelligence in clinical medicine: Catalyzing a sustainable global healthcare paradigm. Frontiers in Artificial Intelligence, 6, 1227091. https://doi.org/10.3389/frai.2023.1227091.

Li, J., Xu, Z., Wang, H., & Li, L. (2024). Geospatial analysis of spatial distribution, patterns, and relationships of health status in the Belt and Road Initiative. Scientific Reports, 14(1). https://doi.org/10.1038/s41598-023-50663-7.

Luo, W., Nguyen, T., Nichols, M., Tran, T., Rana, S., Gupta, S., et al. (2015). Is demography destiny? Application of machine learning techniques to accurately predict population health outcomes from a minimal demographic dataset. PLOS ONE, 10(5), e0125602. https://doi.org/10.1371/journal.pone.0125602.

Moon, J. (2021). The effect of the heatwave on the morbidity and mortality of diabetes patients: A meta-analysis for the era of the climate crisis. Environmental Research, 195(1), 110762. https://doi.org/10.1016/j.envres.2021.110762.

Onprasonk, S., Laohasiriwong, W., & Puttanapong, N. (2023, May 29). Spatial association between environmental factors and melioidosis in Thailand. https://doi.org/10.21203/rs.3.rs-2961968/v1.

Puttanapong, N., Luenam, A., & Jongwattanakul, P. (2022). Spatial analysis of inequality in Thailand: Applications of satellite data and spatial statistics/econometrics. Sustainability, 14, 3946. https://doi.org/10.3390/su14073946.

Ratovoson, R., Rasetarinera, O. R., Andrianantenaina, I., Rogier, C., Piola, P., & Pacaud, P. (2015). Hypertension, a neglected disease in rural and urban areas in Moramanga, Madagascar. PLOS ONE, 10(9), e0137408. https://doi.org/10.1371/journal.pone.0137408.

Savage, A., Bambrick, H., McIver, L., & Gallegos, D. (2021). Climate change and socioeconomic determinants are structural constraints to agency in diet-related non-communicable disease prevention in Vanuatu: A qualitative study. BMC Public Health, 21, 1231. https://doi.org/10.1186/s12889-021-11245-2.

Tuoane-Nkhasi, M., & van Eeden, A. (2016). Spatial patterns and correlates of mortality due to selected non-communicable diseases among adults in South Africa, 2011. GeoJournal, 82(5), 1005-1034.

Verma, C. S., & Usmani, G. (2019). Relationship between health and economic growth in India. Indian Journal of Human Development, 13(3), 344-356. https://doi.org/10.1177/09737030198876.

Yu, B. (2020). Statistical methods for global health and epidemiology. Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-030-35260-8_7.

Arsad, F. S., Hod, R., Ahmad, N., Ismail, R., Mohamed, N., Baharom, M., Osman, Y., Mohd Radi, M. F., & Tangang, F. (2022). The impact of heatwaves on mortality and morbidity and the associated vulnerability factors: A systematic review. International Journal of Environmental Research and Public Health, 19(23), 16356. doi: 10.3390/ijerph192316356

Henke and Petropoulos (2013). A GIS-based exploration of the relationships between human health, social deprivation and ecosystem services: The case of Wales, UK. Applied Geography,45,77–88. https://doi.org/10.1016/j.apgeog.2013.07.022

Tao, H., Salih, S., Oudah, A. Y., Abba, S. I., Ameen, A. M. S. A., Awadh, S. M., Alawi, O. A., Mostafa, R. R., Surendran, U. P., & Yaseen, Z. M. (2022). Development of new computational machine learning Model for longitudinal dispersion coefficient determination: Case study of natural streams, United States. Environmental Science and Pollution Research. https://doi.org/10.1007/s11356-022-18554-y

Ahmad, M. W., Reynolds, J., & Rezgui, Y. (2018). Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. Journal of Cleaner Production, 203, 810-821. https://doi.org/10.1016/j.jclepro.2018.08.20

APPENDIX

APPENDIX A

Result From Econometric Test

Estimated model with full criteria Hausman test and economic loss from NCDs to determine whether a fixed effects model is more appropriate than a random effects model. The Hausman test results are significant in all models, indicating that fixed effects are more appropriate than random effects for these models. Negative or large Hausman statistics suggest that fixed effects provide a better specification.

Table A.1

Diabetes

Variable	Model 1		Model 2		Model 3		Model 4	
	Fixed effect	Random effect	Fixed effect	Random effect	Fixed effect	Random effect	Fixed effect	Random effect
Economic loss	0.0976***	0.379***	0.0938***	0.298***	0.0984***	0.302***	0.0930**	0.339***
	(4.19)	(13.76)	(3.91)	(11.65)	(3.64)	(11.87)	(3.31)	(12.64)
Year dummies	Not include	Not include	Not include	Not include	included	included	included	included
Control Variable	Not include	Not include	included	included	included	included	included	included
constant	included	included	included	included	included	included	included	included
F statistic	17.58***		10.67***		8.718***		6.980***	
Chi2		189.4		496.3		520.0		608.9
R-squared	0.227		0.347		0.394		0.408	
Hausman Test	-189.94		81.47		73.08		48.56	
Misspecification Test (Prob.)	0.4697		0.0000		0.0000		0.0000	

Notes: standard error in parentheses; * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table A.2

Hypertension

Variable	Model 1		Model 2		Model 3		Model 4	
	Fixed effect	Random effect	Fixed effect	Random effect	Fixed effect	Random effect	Fixed effect	Random effect
Economic loss	0.143***	0.216***	0.133***	0.183***	0.126***	0.178***	0.124***	0.185***
	(6.89)	(9.33)	(6.92)	(8.52)	(6.53)	(8.28)	(6.17)	(8.30)
Year dummies	Not include	Not include	Not include	Not include	included	included	included	included
Control Variable	Not include	Not include	included	included	included	included	included	included
constant	included	included	included	included	included	included	included	included
F statistic	47.50***		22.03***		15.71***		11.49***	
Chi2		86.99***		293.0		303.1		321.7
R-squared	0.173		0.333		0.366		0.370	
Hausman Test	-50.25		120.50		118.96		115.02	
Misspecification Test (Prob.)	0.2356		0.0000		0.0000		0.0000	

Notes: standard error in parentheses; * p<0.1, **p<0.05, ***p<0.01

Table A.3

Ischemic

Variable	Model 1		Model 2		Model 3		Model 4	
	Fixed effect	Random effect	Fixed effect	Random effect	Fixed effect	Random effect	Fixed effect	Random effect
Economic loss	0.388***	0.665***	0.352***	0.544***	0.345***	0.536***	0.342***	0.533***
	(11.86)	(27.04)	(10.42)	(20.80)	(10.50)	(20.83)	(10.35)	(20.87)
Year dummies	Not include	Not include	Not include	Not include	included	included	included	included
Control Variable	Not include	Not include	included	included	included	included	included	included
constant	included	included	included	included	included	included	included	included
F statistic	140.7***		46.26***		30.59***		21.62***	
Chi2		730.9		1377.9		1387.1		1419.30
R-squared	0.383		0.455		0.494		0.500	
Hausman Test	166.07		100.23		102.14		88.35	
Misspecification Test (Prob.)	0.0000		0.2108		0.2360		0.1581	

Notes: standard error in parentheses; * *p*<0.1, ***p*<0.05, ****p*<0.01

Table A.4

Chronic

Variable	Model 1		Model 2		Model 3		Model 4	
	Fixed effect	Random effect	Fixed effect	Random effect	Fixed effect	Random effect	Fixed effect	Random effect
Economic loss	0.0976***	0.225***	0.0938***	0.228***	0.0984***	0.237***	0.0930**	0.239***
	(4.19)	(8.45)	(3.91)	(9.61)	(3.64)	(8.77)	(3.31)	(8.67)
Year dummies	Not include	Not include	Not include	Not include	included	included	included	included
Control Variable	Not include	Not include	included	included	included	included	included	included
constant	included	included	included	included	included	included	included	included
F statistic	17.58***		10.67***		8.718***		6.980***	
Chi2		71.45		565.6		589.5		588.6
R-squared	0.072		0.281		0.308		0.316	
Hausman Test	-98.13		1225.42		74.78		48.18	
Misspecification Test (Prob.)	0.0478		0.0000		0.0000		0.0000	

Notes: standard error in parentheses; * *p*<0.1, ***p*<0.05, ****p*<0.01

APPENDIX B
Endogeneity and Overidentification Tests

In our main analysis, we tested for endogeneity and overidentification across four cases (diabetes, hypertension, ischemic heart disease, and chronic respiratory diseases). This testing is crucial to ensure the reliability of our econometric models and to validate the use of instrumental variables (IVs). Although no significant issues were detected, this section explains the methodology and results of these tests in detail.

1) Instrumental Variable Selection

We began by identifying instrumental variables (IVs) that exhibit a high correlation with the economic losses associated with each NCDs and a low correlation with the Gross Provincial Product (GPP) (Gui, R., et al. ,2023). The chosen instruments for each NCDs were determined based on their influential factors as identified through machine learning algorithms. These IVs are crucial for addressing endogeneity issues by serving as proxies that are correlated with the endogenous explanatory variables but not with the error term in the regression model. The following instruments were selected for each NCDs

1.1) Diabetes: Precipitation (ln precip) and the portion of alcohol expenditure (ln alcohol).

1.2) Hypertension: Precipitation (ln precip) and the portion of alcohol expenditure (ln alcohol).

1.3) Ischemic Disease: Nighttime temperature (ln lstnight) and the portion of tobacco expenditure (ln tobacco).

1.4) Chronic Diseases: Carbon dioxide (ln co2), sulfur dioxide (ln so2), and the portion of alcohol expenditure (ln alcohol).

2) Sargan-Hansen statistic for Overidentification

The Sargan-Hansen test evaluates whether the instruments are valid and uncorrelated with the error terms. A high p-value suggests that the instruments are appropriate and the model is correctly specified (Jann, M., 2024).

Table B.1

Sargan-Hansen Test for Overidentification

Model	**Sargan-Hansen statistic (Prob.)**
Diabetes	0.308 (0.5787)
Hypertension	0.063(0.8024)
ischemic	0.181 (0.6703)
chronic	3.928 (0.1403)

Notes: standard error in parentheses; * p<0.1, **p<0.05, ***p<0.01

The Sargan-Hansen test results for all diseases indicate that the p-values are greater than 0.05, suggesting no significant overidentification issues. This implies that the instruments used are valid and there is no evidence that they are correlated with the error terms in the model by Roodman (2006).

3) Hausman Test for Endogeneity

The Hausman test (Bettinger, E.P., 2010). compares the fixed effects model and the instrumental variable model to determine if there is endogeneity. A significant p-value indicates the presence of endogeneity, which justifies the use of IVs.

Table B.2

Hausman Test for Endogeneity

Model	Hausman (Prob.)
Diabetes	2.30 (0.8060)
Hypertension	8.64 (0.1243)
ischemic	0.50 (0.9738)
chronic	0.62 (0.9997)

Notes: standard error in parentheses; * $p<0.1$, ** $p<0.05$, *** $p<0.01$

The Hausman test results for all diseases show p-values greater than 0.05, indicating no significant endogeneity issues. This suggests that the fixed effects models do not suffer from bias due to omitted variables that could be correlated with the predictors (J. Hahn et al., 2011).

There are limitations in identifying suitable instrumental variables that meet the conditions of relevance and exogeneity. The selected IVs represent the best available options based on their correlation with the control variables and their minimal correlation with the economic losses due to NCDs. Despite these limitations, the IVs chosen are the best possible under the given constraints. The results from both the Sargan-Hansen and Hausman tests support their validity and the robustness of the models. These findings should be noted as part of the study's limitations. Future research could explore alternative instruments or additional methods to further validate the models and address any potential endogeneity issues.