Utilizing Machine Learning Approaches in Predicting the Impact of Climate Change Factors to Common Seasonal Disease Outbreak

Riah E. Encarnacion
Caraga State University

https://doi.org/10.5109/7395640

出版情報：Proceedings of International Exchange and Innovation Conference on Engineering & Sciences (IEICES). 11, pp.1036-1043, 2025-10-30. International Exchange and Innovation Conference on Engineering & Sciences
バージョン：

Utilizing Machine Learning Approaches in Predicting the Impact of Climate Change Factors to Common Seasonal Disease Outbreak

Riah E. Encarnacion, DIT[1]
[1]Caraga State University
reencarnacion@carsu.edu.ph

Abstract: *Climate change impacts individuals in various ways, particularly in terms of public health and the spread of common illnesses. To address the effects and manage the risks of disease transmission and outbreaks, this study aimed to develop a prediction model using Machine Learning approaches based on seasonal diseases and climate change data. The research utilized the Knowledge Discovery from Data framework in conducting the study and analyzed 15 years' worth of climate-related and disease-specific data from credible sources in the Philippines. The research findings revealed that Machine Learning, specifically the Decision Tree classifiers, a type of supervised Machine Learning in data mining, were effective in uncovering hidden patterns and rule dependencies within the data sets. These discovered rules can be translated into programmable code, which can be applied in developing a functional climate-driven early warning system to forecast the occurrence of diseases such as tuberculosis, malaria, dengue, typhoid, and diarrhea.*

Keywords: machine learning; decision tree model; climate change; common diseases

1. INTRODUCTION

Climate change is a pressing global issue that impacts multiple facets of human life. It refers to the recurring shifts in Earth's climate, driven by changes in the atmosphere and its interactions with the planet's geological, chemical, biological, and geographical systems [1]. These changes may stem from natural phenomena like volcanic eruptions, typhoons, or solar activity, or be human induced, such as the burning of fossil fuels that release carbon dioxide and other greenhouse gases [2]. In addition, urban transportation is regarded as a significant contributor to energy use and environmental pollution [3]. Drastic shifts in climate can influence agriculture, threaten food security, and contribute to the spread of diseases that pose serious risks to human health.

Climate change significantly impacts human health [3] by influencing the spread and transmission of various diseases. Extreme weather conditions such as increased humidity, heavy rainfall, and rising temperatures [4][5] have been linked to outbreaks of vector-borne (e.g., dengue, malaria), water-borne (e.g., cholera, typhoid), air-borne (e.g., tuberculosis, influenza), and food-borne diseases (e.g., diarrhea) [6]. Warmer climates create favorable environments for disease vectors like mosquitoes [7], while heavy rains and flooding can damage sanitation infrastructure [8], increasing exposure to contaminated water [9]. Microorganisms like bacteria and fungi can thrive indoors when there is enough moisture, leading to indoor air pollution [10]. High temperatures may inhibit some pathogens, such as the malaria parasite, but also prolong the survival of others in humid conditions [11]. Additionally, climate-related disasters may displace populations and disrupt healthcare access, indirectly affecting diseases like tuberculosis and those associated with weakened immune systems, such as HIV/AIDS and diabetes [12].

In the Philippines, where climate change significantly affects public health, understanding its impact on disease transmission is essential. Seasonal variations in disease cases are expected to intensify with climate change. Machine learning (ML) and big data are increasingly recognized as effective tools for predicting and managing disease outbreaks [13]. Algorithms like Decision Trees, Random Forests, and Artificial Neural Networks have proven useful in assessing climate risks [14] and forecasting diseases [15]. Data mining and statistical methods help predict weather variables and identify disease patterns across time and regions [16]. Overall, ML-driven models offer strong potential for enhancing climate-related disease prediction and response efforts.

As a result, various studies have been conducted to predict climate patterns to reduce climate-related risks [17]. However, no existing model has been developed to forecast disease transmission based on climate data within the Philippine context. This gap drives the researcher to develop a predictive model using machine learning, specifically decision tree algorithms, and extract rules to forecast the potential impact of climatic factors on the outbreak of infectious diseases. Specifically, the study seeks to answer the following questions:

1. What decision tree models can be proposed to predict the outbreak of common seasonal diseases from climate change datasets?

2.What are the generated rules derived from the decision tree models to predict the outbreak of common seasonal diseases?

3.What is the accuracy result and performance level of the disease outbreak prediction models based on the validated metrics?

The results of the study can be translated into programmable code, which can be applied in developing a functional climate-driven early warning system to forecast the occurrence of diseases such as tuberculosis, malaria, dengue, typhoid, and diarrhea.

2. MATERIALS AND METHODS

2.1 Knowledge Discovery in Database (KDD) Framework

The Knowledge Discovery from Data (KDD), a concept first introduced at the inaugural KDD Workshop in 1989 [18], is a comprehensive, iterative process, with data mining being just one phase within it. KDD encompasses the entire procedure of deriving useful knowledge from data, whereas data mining specifically focuses on using algorithms to detect patterns and insights [19].

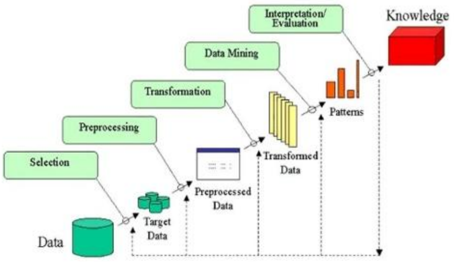


Fig. 1: Iterative Steps of the Knowledge Discovery from Data (KDD) (Fayyad et a., 1996)

2.1.1 Data Selection and Pre-processing

This study utilized 15 years of data on identified climate change impacts and disease cases in the Philippines from 2009-2023. The selection of datasets used for data mining plays a pivotal role in knowledge discovery. The climate change impacts are narrowed in specific indicators which include carbon dioxide emission (F1), greenhouse gas emission (F2), surface temperature (F3), typhoons (F3), rainfall (F4), number of typhoons entering the Philippine Area of Responsibility (X5). Meanwhile, five common seasonal diseases occur in the Philippines, such as Tuberculosis (D1), Influenza (D2), Malaria (D3), Dengue (D4), Cholera (D5), Typhoid (D6), and Diarrhea (D7). The datasets included herein are secondary data that were collected from reliable websites and studies specified in the table below.

Table 1. Climate Change Factors and Common Seasonal Disease Datasets

Factors	Data	Code	Attribute	Unit	Sources
Climate Change Factors	**Carbon dioxide emission**	F1	Numeric	Million tons	Global Carbon Budget (2024) Our World in Data
	Green Gas Emission	F2	Numeric	Tons/person	Jones et al. (2024); Population based on various sources (2024)
	Surface Temperature	F3	Numeric	Degree Celsius	World Bank Group (2024)
	Rainfall	F4	Numeric	Thousand millimeters	World Bank Group (2024)
	Typhoons	F5	Numeric	Annual number	PAG-ASA as cited by Malawani (2024)
Common Seasonal Diseases	**Tuberculosis cases**	D1	Numeric	Per thousand	Our World in Data (2024), Noriega (2024)
	Influenza Cases	D2	*Numeric*	*Per thousand*	*FluNet by the World Health Organization (2023)*
	Malaria cases	D3	Numeric	Per thousand	Maru (2023)
	Dengue cases	D4	Numeric	Per thousand	DOH (2023), Edillio et al. (2015), Aumentado et al. (2015)
	Cholera cases	D5	Numeric	Per thousand	DOH (2023), Our World in Data (2023)
	Typhoid cases	D6	Numeric	Per thousand	Statista (2023), Philippine Statistics Authority (2023)
	Diarrhea Cases	D7	*Numeric*	*Per thousand*	*IHME, Global Burden of Disease (2024). Our World in Data (2021)*

Before datasets were analyzed, the researcher ensured the removal of "noisy" and irrelevant data. Data preprocessing must be properly considered before mining since poor data quality will result to erroneously analysis and rules generation. Missing data and outliers were checked using Jamovi version 2.3.21 software and were dealt with accordingly.

The data for each common disease and climatic factors were aggregated from monthly data to generate the annual data. Table 2.0 shows the Average values of the climate change factors.

Table 2. The Average Values of Climate Change Factors in the Philippines Per Year

Year	F1	F2	F3	F4	F5
2009	75.91	2.40	25.99	22	3.16
2010	82.60	2.70	26.27	11	2.82
2011	83.41	2.70	26.03	19	2.60
2012	88.12	2.40	26.37	17	3.36
2013	95.52	2.40	26.37	25	2.96
2014	100.78	2.40	26.24	19	2.73
2015	111.85	2.40	26.46	15	2.52
2016	120.99	2.40	26.79	13	2.37
2017	134.45	2.50	26.35	22	2.49
2018	140.79	2.50	26.55	21	3.18
2019	143.53	2.50	26.62	21	2.77
2020	132.23	2.30	26.73	22	2.40
2021	141.17	2.40	26.72	15	2.53
2022	144.90	2.40	26.61	18	2.81
2023	154.58	2.50	26.65	11	3.07

After the datasets were cleaned, the researcher then computed the mean per item set of each common disease that occurred in the Philippines for the past 15 years from 2009 to 2023 as presented in Table 3.0. For the disease datasets, the mean values for each item were: Tuberculosis (553), Influenza (15,009), Malaria (9.15), Dengue (158.43), Cholera (9.77), Typhoid (7.91), and Diarrhea (50.81).

Table 3. The Average Cases of Common Seasonal Diseases in the Philippines Per Year

Year	D1	D2	D3	D4	D5	D6	D7
2009	524.00	7688.00	24.14	48.23	2.98	10.28	70.01
2010	528.00	1724.00	20.12	54.94	5.65	6.30	67.40
2011	499.00	894.00	19.33	171.13	5.75	2.82	64.74
2012	511.00	651.00	9.62	125.97	9.77	10.33	61.89
2013	524.00	975.00	8.15	187.03	12.85	12.57	58.30
2014	537.00	521.00	7.73	204.90	5.08	4.57	54.00
2015	549.00	164.00	4.60	113.48	4.55	8.11	49.70
2016	562.00	259.00	8.27	108.26	4.76	11.37	46.08
2017	574.00	268.00	6.96	180.47	14.48	13.06	43.85
2018	583.00	236.00	4.01	131.82	3.91	8.96	42.58
2019	591.00	592.00	4.90	216.19	2.28	7.55	41.51
2020	599.00	0.00	5.78	437.53	3.88	3.75	40.72
2021	591.00	105.00	6.12	90.13	1.95	2.04	40.30
2022	650.00	244.00	4.30	79.82	2.02	4.84	40.60
2023	470.00	688.00	3.19	226.49	66.61	12.10	40.43
Average	**553**	**15009**	**9.15**	**158.43**	**9.77**	**7.91**	**50.81**

2.1.2 Data Transformation

The process of data transformation in data mining involves converting data into a suitable format or structure for analysis to ensure that the data is cleaned, consistent, and ready for mining. In this study the data transformation process utilized the normalization method by scaling data to fall within a specific range. The binary digits are commonly used for the common seasonal disease cases. Hence, 1 is assigned for every item above the mean, while 0 for items equal or below the mean. Table 4 shows the transformed data of the 15-year records of the diseases. These transformations improve the quality and effectiveness of mining algorithms by ensuring consistent, interpretable input data.

Table 4.0 The Numerical Transformed Data of Common Seasonal Disease Cases in the Philippines

Year	D1	D2	D3	D4	D5	D6	D7
2009	0	1	1	0	0	1	1
2010	0	1	1	0	0	0	1
2011	0	1	1	1	0	0	1
2012	0	1	1	0	1	1	1
2013	0	0	0	1	1	1	1
2014	0	0	0	1	0	0	1
2015	0	0	0	0	0	1	0
2016	1	0	0	0	0	1	0
2017	1	0	0	1	1	1	0
2018	1	1	0	0	0	1	0
2019	1	0	0	1	0	0	0
2020	1	0	0	1	0	0	0
2021	1	0	0	0	0	0	0
2022	1	1	0	0	0	0	0
2023	0	1	0	1	1	1	0

After which, the values were discretized to convert the numerical values into discrete categories. Discretization is commonly used in data mining and machine learning when algorithms require categorical input or when simplifying data improves interpretability, simplifies model input, reduces noise and overfitting, and enhances rule-based systems. Table 5, presents the clustering-based to form bins, hence numerical data were discretized as 1 or "HI" mean High Impact and 0 or "LI" mean Low Impact.

Table 5. The Discretized Data of Common Seasonal Disease Cases in the Philippines

Year	D1	D2	D3	D4	D5	D6	D7
2009	LI	HI	HI	LI	LI	HI	HI
2010	LI	HI	HI	LI	LI	LI	HI
2011	LI	HI	HI	HI	LI	LI	HI
2012	LI	HI	HI	LI	HI	HI	HI
2013	LI	LI	LI	HI	HI	HI	HI
2014	LI	LI	LI	HI	LI	LI	HI
2015	LI	LI	LI	LI	LI	HI	LI
2016	HI	LI	LI	LI	LI	HI	LI
2017	HI	LI	LI	HI	HI	HI	LI
2018	HI	HI	LI	LI	LI	HI	LI
2019	HI	LI	LI	HI	LI	LI	LI
2020	HI	LI	LI	HI	LI	LI	LI
2021	HI	LI	LI	LI	LI	LI	LI
2022	HI	HI	LI	LI	LI	LI	LI
2023	LI	HI	LI	HI	HI	HI	LI

Table 6. The Final Datasets for Data Mining

F1	F2	F3	F4	F5	D1	D2	D3	D4	D5	D6	D7
75.91	2.40	25.99	22	3.16	LI	HI	HI	LI	LI	HI	HI
82.60	2.70	26.27	11	2.82	LI	HI	HI	LI	LI	LI	HI
83.41	2.70	26.03	19	2.60	LI	HI	HI	HI	LI	LI	HI
88.12	2.40	26.37	17	3.36	LI	HI	HI	LI	HI	HI	HI
95.52	2.40	26.37	25	2.96	LI	LI	LI	HI	HI	HI	HI
100.78	2.40	26.24	19	2.73	LI	LI	LI	HI	LI	LI	HI
111.85	2.40	26.46	15	2.52	LI	LI	LI	LI	LI	HI	LI
120.99	2.40	26.79	13	2.37	HI	LI	LI	LI	LI	HI	LI
134.45	2.50	26.35	22	2.49	HI	LI	LI	HI	HI	HI	LI
140.79	2.50	26.55	21	3.18	HI	HI	LI	LI	LI	HI	LI
143.53	2.50	26.62	21	2.77	HI	LI	LI	HI	LI	LI	LI
132.23	2.30	26.73	22	2.40	HI	LI	LI	HI	LI	LI	LI
141.17	2.40	26.72	15	2.53	HI	LI	LI	LI	LI	LI	LI
144.90	2.40	26.61	18	2.81	HI	HI	LI	LI	LI	LI	LI
154.58	2.50	26.65	11	3.07	LI	HI	LI	HI	HI	HI	LI

The final datasets are composed of the combined data of the common seasonal diseases and the climate change factors as can be seen on Table 6. These datasets are readily available for data mining using machine learning method.

2.1.3 Data Mining using Machine Learning and Decision Tree Approaches

Many people believed that knowledge is power, and this is what makes Data Mining useful. It has the ability to bring relevant discovered information that is useful in making strategic decisions in order to succeed. Han, J. et al. briefly define Data Mining as "searching for knowledge in data". The authors agree that, analogous to "gold mining", data mining should have been termed as "knowledge mining from data" [20].

Machine learning techniques, a subset of data mining within the realm of artificial intelligence, involve self-directed computational processes that utilize logical or binary operations to acquire knowledge from diverse datasets [21]. Mahesh identifies various machine learning tasks, including supervised, unsupervised, and semi-supervised learning, as well as reinforcement learning, multi-task learning, ensemble methods, neural networks, and instance-based learning [22]. This research focuses on using a decision tree model to predict antecedents by applying numerical rules and identifying potential predictors from frequently co-occurring item sets.

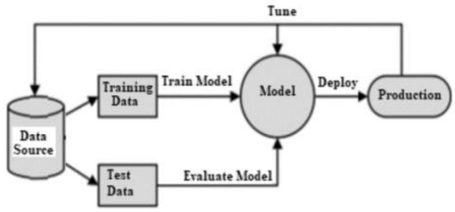
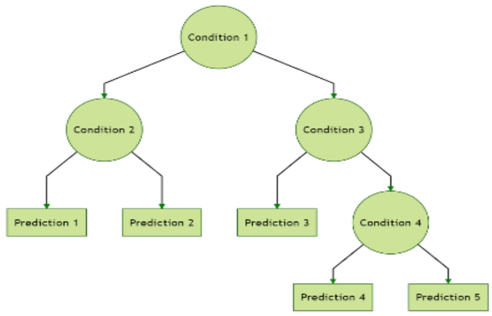


Fig 2. Workflow of Supervised Machine Learning
Source: Mahesh (2019)

The decision tree method is inherently predictive and derives rules using information gain. As a supervised machine learning technique, it learns to associate inputs with corresponding outputs based on sample input-output pairs [21]. Figure 3 below illustrates the visual



representation of the decision tree model.

Fig 3. Decision Tree Model and Its Components
Source: International Business Machines (IBM) Corporation (n.d.)

The input variables of the study were climate change factors data in numeric format and disease cases, which were also tagged as the target or dependent variable. These binary digits were discretized into 1 as "High Impact" and 0 as "Low Impact". As mentioned previously, the method of discretization, which involves converting continuous variables into discrete variables, aids in handling outliers by allocating these values to the lower or higher intervals with the distribution's remaining inlier values [23].

2.1.4 Data Interpretation and Evaluation

Validating predictive models is crucial to ensure their dependability and accuracy, typically done through well-established evaluation metrics. Model effectiveness is measured using indicators like precision, recall (sensitivity), the F1 score, and the Receiver Operating Characteristic (ROC) curve. The F1 score reflects a model's ability to accurately classify outcomes, balancing both precision and recall. Precision represents the percentage of predicted positive cases that are truly positive, while recall measures how well the model identifies actual positive cases. A high recall indicates few false negatives, whereas high precision points to fewer false positives. Bajaj [24] notes that strong model performance is demonstrated by precision, recall, and F1 scores exceeding 0.5. For binary classification tasks, the ROC curve is widely used to evaluate model performance. Additionally, classification accuracy (CA)—the ratio of correct predictions to total predictions—is another metric, with scores above 50% regarded as acceptable [24][25].

2.2 Use of AI Software for Machine Learning Approaches

This study employs the Decision Tree Approach using Orange version 3.36.2, an open-source machine learning platform built on Python. The tool supports visual programming, enabling interactive data exploration and visualization. The workflow for the Decision Tree Approach is illustrated in the figure below.

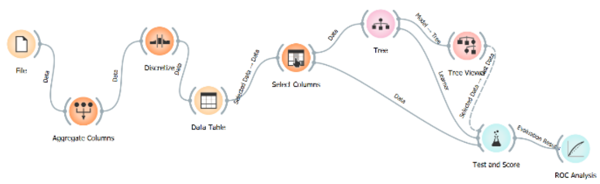


Fig. 4. The Workflow of the Decision Tree Analysis in Orange version 3.36.2

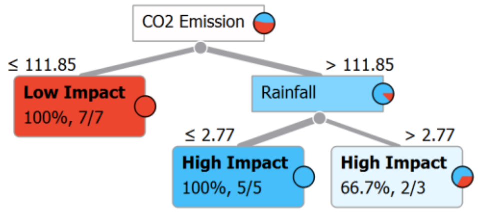
3. RESULTS AND DISCUSSION

The Decision Tree Models are generated using the Orange App software to help predict the occurrences of Common Seasonal Diseases Outbreaks driven by Climate Change Datasets. The target variables were the disease datasets, while the predictor variables were CO2 emission, greenhouse gas emission, temperature, typhoon, and rainfall.

3.1 Tuberculosis Transmission

As illustrated in Figure 5, the key predictors for tuberculosis were carbon dioxide emissions and rainfall levels. For example, when annual carbon dioxide emissions are less than or equal to 111.85 million tons, the likelihood of tuberculosis transmission is 100%, though with low impact. In contrast, if emissions exceed 111.85 million tons and rainfall is at or below 2.77 units, there is a 100% probability of a high-impact tuberculosis outbreak. However, when rainfall exceeds 2.77, the probability drops to 66.7%. These findings suggest that exposure to elevated carbon dioxide levels, combined with specific amounts of rainfall, may increase the risk of tuberculosis incidence.

An increase in pollutant emissions is a major contributor to declining health conditions. According to research by Yahya and Rafiq [26], rising carbon dioxide emissions are linked to a higher incidence of tuberculosis. The onset of tuberculosis is strongly associated with pollutants like CO_2, which can weaken the immune system, disrupt macrophage function, affect epithelial barrier integrity, and alter airway resistance [27]. Tremblay [28] explains that carbon dioxide may elevate the risk of M. tuberculosis reactivation by initiating the interleukin-10 pathway in lung macrophages. In microbiology, Park et al. [29] noted that mycobacteria can utilize carbon as an energy source under aerobic conditions. Additionally, acid rain—resulting from pollution—can worsen TB infections, as the pollutants that contribute to acid rain also generate fine particles that may exacerbate existing

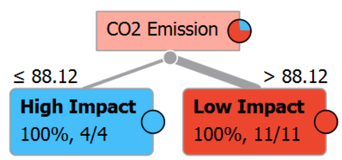


health problems or lead to respiratory issues [30].

Fig. 5. Predictors of Tuberculosis Outbreak

3.1.2 Influenza Outbreak

As depicted in Figure 6, the key factors influencing tuberculosis transmission were rainfall, typhoon frequency, and greenhouse gas emissions. For instance, when rainfall exceeds 2.77 thousand millimeters and typhoon occurrences are around or beyond 18 times, there is a 100% likelihood of disease transmission with a high impact influenza outbreak. Conversely, if rainfall falls below 2.77 thousand millimeters and greenhouse gas emissions are under 2.4 tons, transmission still occurs at a 100% rate but with lower impact. These findings suggest that increased rainfall, coupled with intensified typhoon activity, may significantly elevate the risk of

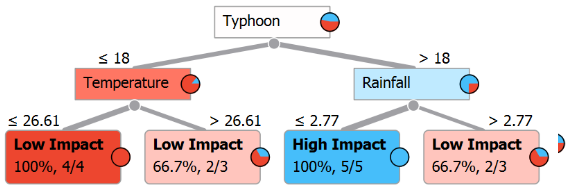


influenza outbreaks.

Fig. 6. Predictors of Malaria Outbreak

Numerous studies have shown that the transmission of influenza and coronavirus tends to increase during colder seasons, particularly in periods of heavy rainfall and typhoon activity, as these viruses thrive in cool and dry environments [31]. Research by Tian et al. [32] also revealed that different pathogens require varying temperature conditions for survival—some thrive at higher temperatures, while others persist in lower ones. Additionally, the study by Bukhari and Jameel [33] highlights that humidity levels between 22 °C and 25 °C and high temperatures below 38 °C influence the stability of influenza and coronavirus. In Asian countries like the Philippines, extreme weather patterns may also contribute to the spread of insect-borne diseases and the movement of birds, potentially increasing the risk of outbreaks such as Avian Influenza H5N1.

3.1.4 Dengue Outbreak
Figure 7 highlights typhoon frequency, rainfall, and temperature as the primary factors affecting dengue transmission. Specifically, when annual typhoon occurrences exceed 18 and rainfall is at or below 2.77 thousand millimeters, the likelihood of a high-impact dengue outbreak reaches 100%. Conversely, if there are fewer than 18 typhoons in a year and the temperature is at or below 26.61°C, dengue transmission still occurs with 100% probability, though the impact is lower. These findings indicate that the interplay between typhoon events, rainfall volume, and temperature levels plays a critical role in shaping the risk and severity of dengue outbreaks.



Fig, 7. Predictors of Dengue Outbreak

Typhoons often lead to heavy rainfall and flooding, which in turn create more stagnant water, especially in swampy regions [34]. These stagnant waters serve as ideal breeding grounds for mosquitoes (Williams et al., 2016). According to Manogaram [35], dengue outbreaks tend to rise during both summer and winter seasons. Moreover, as temperatures increase and mosquito breeding sites become drier, the likelihood of dengue epidemics may decline due to climate warming. Conversely, a drop in the temperature could lead to more favorable mosquito breeding environments, thereby increasing the risk of dengue transmission.

3.1.5 Cholera Outbreak
As illustrated in Figure 8, rainfall emerged as the sole predictor for cholera transmission. For example, when annual rainfall exceeds 2.82 thousand millimeters, there is a 60% chance of a high-impact cholera outbreak. In contrast, if yearly rainfall falls between 2.49 and 2.82 thousand millimeters, the probability of a cholera outbreak rises to 100%, albeit with low impact. This association suggests that specific levels of rainfall may influence the likelihood and severity of cholera transmission.

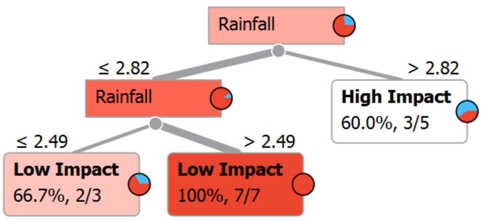


Fig. 8. Predictors of Cholera Outbreak

Climate change has also been linked to waterborne infectious diseases. For example, rainfall precipitate may also cause an increase in cholera transmission and increase the cholera pathogens. Waterborne diseases may increase due to the rainfall. The rainy season can increase the development of fecal pathogens [36]. Fredrick et al. [37] discovered that heavy rainfall can affect the water supply system, which was then contaminated by the drainage caused by rain during the cyclone, resulting in an outbreak of cholera due to the consumption of contaminated water. Cholera and giardiasis may experience increased outbreaks due to extreme temperatures and flooding caused by heavy rainfall [8]. Righetto et al. [38] indicated that infections result from exposure to high pathogen concentrations in the water. These, in turn, are driven by rainfall-induced washout of outdoor dumping areas or cesspool overflows, hydrologic transport through waterways, and by movement of susceptible and infected individuals.

3.1.6 Typhoid Outbreak
As shown in Figure 9, the predictor for typhoid transmission was only rainfall. For instance, if the total rainfall for a year is more than 2.82 thousand millimeters, there is a high impact of typhoid transmission with 100% probability. It can be inferred from this association that rain precipitate at certain amount may increase typhoid transmission.

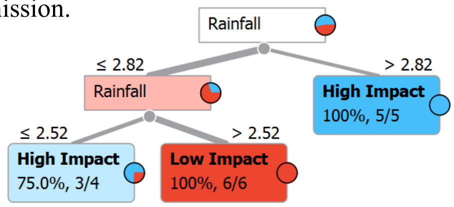


Fig. 9. Predictors of Typhoid Outbreak

The study by [39] suggests that climate change can alter both the timing and volume of rainfall, leading to unexpected precipitation events at irregular times. Such intense rainfall can trigger outbreaks of various disease-causing pathogens. Climate change may also enhance the reproduction, survival, and life cycle of pathogens, either directly or indirectly [40], contributing to the spread of unexpected waterborne diseases in the environment [39]. Additionally, Gao et al. [41] observed that the relative risk of typhoid fever increased shortly after the onset of light rainfall, with a delayed effect occurring within 13 to 26 days. However, as rainfall continued to rise, the risk plateaued, with the highest cumulative risk recorded at 100mm of rainfall (relative risk 24.96 [95% CI: 4.54–87.21]).

3.1.7 Diarrhea Outbreak
As presented in Figure 10, the only predictor of Diarrhea outbreak is Carbon Dioxide emission. For instance, if the carbon dioxide emission is less than equal to 100.78 million tons, there is a high impact of diarrhea outbreak with 100% probability. Whereas, there is a low impact of diarrhea outbreak with 100% probability if the carbon dioxide emission is greater than 100.78 million tons.

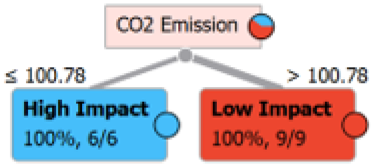


Fig. 10. Predictors of Diarrhea Transmission

Carbon emissions may lead to plenty of health issues, and may result in an increase in temperature, which in turn may lead to a change in the climate. Jacobson et al. [42] suggest that prolonged exposure to elevated amounts of ambient carbon dioxide may lead to plenty of health risks, such as, inflammation, reductions in higher-level cognitive abilities, bone demineralization, kidney calcification, oxidative stress, and endothelial dysfunction. The carbon dioxide emission also affects the temperature, which is commonly related to common seasonal infectious diseases such as Diarrhea [43].

3.2 The Generated Rules to Predict the Transmission of Common Seasonal Diseases

Based on the decision tree models, the following rules are generated to predict the transmission of common seasonal diseases, as can be seen in Table 7. These rules could be converted to programming codes, which will then be used in future works, creating a program for a prediction model that applies the aforementioned rules to predict the outcome.

Table 7. Generated Rules for Climate Change-Disease Transmission Prediction Model

Disease	Climate Change Risk Factor(s)	Generated Rules	Probability of Transmission	Impact Level
Tuberculosis	Carbon Dioxide Emission (CO2 Emission)	{CO2 Emission ≤ 111.85}	100%	Low Impact
	Carbon Dioxide Emission (CO2 Emission) and Rainfall	{CO2 Emission > 111.85, Rainfall >2.77}	66.7%	High Impact
	Carbon Dioxide Emission (CO2 Emission) and Rainfall	{CO2 Emission > 111.85, Rainfall ≤ 2.77}	100%	High Impact
Influenza	Rainfall and Typhoons	{Rainfall > 2.77, Typhoon >18}	66.7%	High Impact
	Rainfall and Typhoons	{Rainfall > 2.77, Typhoon ≤18}	100%	High Impact
	Rainfall and Green House Gas Emission	{Rainfall ≤ 2.77, Green House Gas Emission ≤ 2.4)	100%	Low Impact
	Rainfall and Green House Gas Emission	{Rainfall ≤ 2.77, Green House Gas Emission > 2.4)	66.7%	Low Impact
Malaria	Carbon Dioxide Emission (CO2 Emission)	{CO2 Emission > 88.12	100%	Low Impact
	Carbon Dioxide Emission (CO2 Emission)	{CO2 Emission ≤ 88.12	100%	High Impact
Dengue	Typhoon and Rainfall	{Typhoon > 18, Rainfall >2.77}	66.7%	Low Impact
	Typhoon and Rainfall	{Typhoon > 18, Rainfall <2.77}	100%	High Impact
	Typhoon and Temperature	{Typhoon ≤ 18, Temperature ≤ 26.61}	100%	Low Impact
	Typhoon and Temperature	{Typhoon ≤ 18, Temperature > 26.61}	66.7%	Low Impact
Cholera	Rainfall	{Rainfall > 2.82}	60.0%	High Impact
	Rainfall	{Rainfall ≤ 2.82, Rainfall ≤ 2.49}	66.7%	Low Impact
	Rainfall	{Rainfall ≤ 2.82, Rainfall >2.49}	100%	Low Impact
Typhoid	Rainfall	{Rainfall >2.82}	100%	High Impact
	Rainfall	{Rainfall ≤ 2.82, Rainfall ≤ 2.52}	75%	High Impact
	Rainfall	{Rainfall ≤ 2.82, Rainfall > 2.52}	100%	Low Impact
Diarrhea	Carbon Dioxide Emission (CO2 Emission)	{CO2 Emission > 100.78}	100%	Low Impact
	Carbon Dioxide Emission (CO2 Emission)	{CO2 Emission ≤ 100.78}	100%	High Impact

3.3 The Accuracy Results and Performance Level of the Disease Outbreak Prediction Models

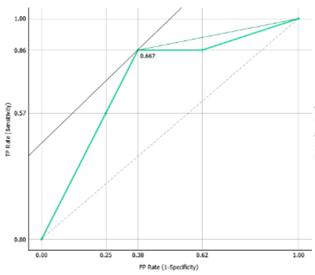
Table 8 shows the summary of performance and accuracy levels of the climate change-disease transmission prediction models using the cross-validation widget in Orange 3.36.2. In this study, the predetermined training set size was 75%, and there were 100 randomly selected possible scenarios during model testing.

Based on the results, all precision, sensitivity, and F1 score values exceeded the 0.5 threshold, except for the models predicting influenza and cholera outbreaks. This indicates that the prediction models for tuberculosis, malaria, dengue, typhoid, and diarrhea demonstrated high precision and sensitivity, further supported by their elevated F1 scores. Consequently, the classification accuracy of these five models also surpassed 50%. For example, the diarrhea prediction model achieved an accuracy of 86.7%, followed by 80% for tuberculosis, and 73.3% for malaria, dengue, and typhoid. These findings suggest that the five validated models are reliable and exhibit strong predictive performance.

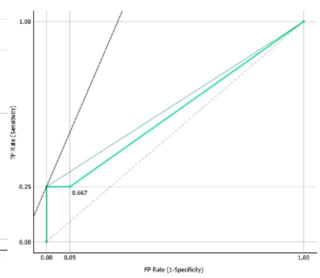
Table 8. Summary of Model Performance and Accuracy Levels

Disease Type	Precision	Sensitivity	F1 Score	Classification Accuracy (CA)
Tuberculosis	0.750	0.857	0.800	0.800
Influenza	0.400	0.429	0.462	0.533
Malaria	0.600	0.750	0.500	0.733
Dengue	0.667	0.857	0.750	0.733
Cholera	0.400	0.500	0.444	0.667
Typhoid	0.700	0.875	0.778	0.733
Diarrhea	0.833	0.833	0.833	0.867

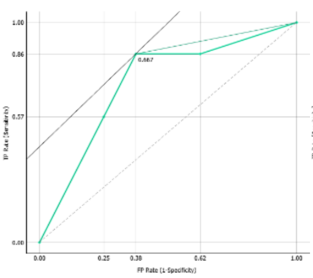
However, the influenza outbreak prediction model falls below the 0.5 threshold for both precision and sensitivity. As shown in the table, the model's precision is 0.400 and its sensitivity is 0.429, indicating low performance in correctly identifying true positives. This is further supported by its low F1 score. Although the model achieved a classification accuracy of 53.3% in predicting influenza outbreaks, its low precision and sensitivity suggest it may not be reliable. Therefore, additional cross-validation is necessary to improve the model's performance.



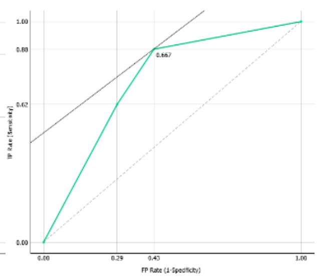
ROC Curve for Tuberculosis (80.0%)



ROC Curve for Malaria (73.3%)



ROC Curve for Dengue (73.3%)



ROC Curve for Typhoid (73.3%)

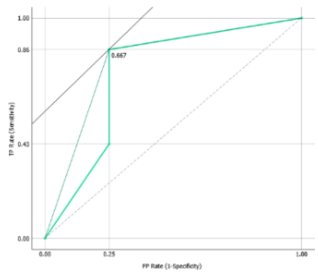


Fig. 12. ROC Curve for Diarrhea (73.3%)

Figure 12 further elaborates the discussion for tuberculosis, malaria, dengue, typhoid, and diarrhea outbreaks. To determine the accuracy of the generated prediction model, the study applied the Receiving Operating Characteristic Curve (ROC) and the Area under ROC Curve (AUC as show in Figure 8). The study revealed that following significant findings: TUBERCULOSIS has 80.0% accuracy; MALARIA has 73.3% accuracy; DENGUE has 73.3% accuracy; TYPHOID has 73.3% accuracy; and DIARRHEA has accuracy 86.7%. The model has gained a high acceptability and accuracy rate in predicting the students' performance in their internship training.

A precision value above 0.500 indicates that the model is highly reliable in accurately identifying high-impact cases as truly high impact, and low-impact cases as truly low impact. A high precision score reflects a low occurrence of false positives. Similarly, a sensitivity score above 0.500 shows that the model is effective in minimizing false negatives. The combination of high precision and sensitivity is further validated by an elevated F1 score. An F1 score closer to 1 signifies that the model excels in both precision and sensitivity, making it effective at identifying prediction errors. The ROC curve also provides a visual representation of the tuberculosis prediction model's performance.

4. CONCLUSION AND RECOMMENDATIONS

The findings suggest that numerous rules were identified through machine learning-based data mining. As a supervised learning model, the decision tree classifier successfully uncovered hidden patterns and dependency rules within the dataset. Using 15 years of climate change indicator data and its relationship to disease transmission, five predictive models demonstrated strong performance and high accuracy. Carbon dioxide emissions emerged as key predictors for outbreaks of tuberculosis, malaria, and diarrhea. Rainfall levels were found to influence the transmission of influenza, cholera, and typhoid, while typhoon activity appeared to intensify dengue cases. The dengue prediction model included rainfall, typhoon frequency, and greenhouse gas emissions as predictors; however, it exhibited lower performance and accuracy based on evaluation results. Therefore, future research is recommended to investigate additional predictors beyond those analyzed in this study.

Nonetheless, the rules derived from machine learning can be translated into programmable code, which can be used to develop software applications for predicting outbreaks of tuberculosis, malaria, dengue, typhoid, and diarrhea. These computer-based predictions can assist the government, particularly the Department of Health, in formulating policies and strategies aimed at preventing the spread of these diseases, especially given the unavoidable impact of climate change. Such predictive tools can also be valuable for health authorities and healthcare providers in protecting the public from infectious diseases.

5. REFERENCES

[1] Jackson, S. T. (2023, December 10). climate change. Encyclopedia Britannica. https://www.britannica.com/science/climate-change

[2] United Nations (2020). What is Climate Change, https://www.un.org/en/climatechange/what-is-climate-change

[3] Maharjan, S., Tsurusaki, N., Divigalpitiya, P., The Influence of Urban Form on Travel Energy Consumption in Kathmandu. Proceedings of International Exchange and Innovation Conference on Engineering & Sciences (IEICES). 4, pp.90-95, 2018-10-18. Interdisciplinary Graduate School of Engineering Sciences, Kyushu University.

[4] Altizer S., Ostfeld R., Johnson P.T.J., Kutz S., Harvell C.D. Climate change and infectious diseases: from evidence to a predictive framework. Science. 2013;341:514–519. doi: 10.1126/science.1239401. [DOI] [PubMed] [Google Scholar]

[5] Ostfeld R.S., Brunner J.L. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.; 2015. Climate change and Ixodes tick-borne diseases of humans; p. 370. [DOI] [PMC free article] [PubMed] [Google Scholar][Ref list]

[6] Rodo X, Pascual M, Doblas-Reyes F.J, Gershunov A, Stone D.A, Giorgi F, Hundson P.J, Kinter J, Rodrigue-Arias M.A, Stenseth N.C (2013) Climate change and infectiouos diseases: can we meet the needs for better prediction? clim.chang 118: 625-640. [Ref list]

[7] Lipp, E., Huq, A., and Cowell, R. (2002). Effects of Global Climate on Infectious Disease: the Cholera Model, https://doi.org/10.1128/cmr.15.4.757-770.2002

[8] Centers for Disease Control and Prevention (2023). Climate change and infectious diseases, https://www.cdc.gov/ncezid/what-we-do/climate-change-and-infectious-diseases/index.html

[9] Indhumathi K, Sathesh Kumar K. A review on prediction of seasonal diseases based on climate change using big data. Mater Today Proc. 2021;37:2648-2652. doi:10.1016/j.matpr.2020.08.517

[10] Coalition Against Typhoid (2019). Typhoid + Climate Change, https://www.coalitionagainsttyphoid.org/wp-content/uploads/2020/07/Typhoid-Climate-Change-10-2019-final-4.pdf

[11] Al Madhoun, W.A., Abed, E.Y., Elmanama, A.A., Kim, H. Assessment of Indoor Microbial Environment of Labs and Faculty Offices at a University in Gaza, Palestine. Proceedings of International Exchange and Innovation Conference on Engineering & Sciences (IEICES). 3, pp.11-14, 2017-10-19. Interdisciplinary Graduate School of Engineering Sciences, Kyushu University. https://doi.org/10.15017/1906117

[10] Ekundayo, F. Using machine learning to predict disease outbreaks and enhance public health surveillance. World Journal of Advanced Research and Reviews, 2024, 24(03), 794–811.

[12] Ravelo, J. (2023). https://www.devex.com/news/is-tuberculosis-being-left-out-of-climate-health-debate-105193

[13] T. Abdullahi and G. Nitschke, "Predicting Disease Outbreaks with Climate Data," 2021 IEEE Congress on Evolutionary Computation (CEC), Kraków, Poland, 2021, pp. 989-996, doi: 10.1109/CEC45853.2021.9504740.

[14] Zennaro, F., Furlan, E., Simeoni, C., Torresan, S., Aslan, S., Critto, A., and Marcomini, A. (2021). Exploring machine learning potential for climate change risk assessment, https://doi.org/10.1016/j.earscirev.2021.103752

[15] Huntingford, C., Jeffers, E., Bonsall, M., Christensen, H., Lees, T., and Yang, H. (2019). Machine learning and artificial intelligence to aid climate change research and preparedness, DOI 10.1088/1748-9326/ab4e55
[16] Olaiya, F and Adeyemo, A B. (2012). Application of Data Mining Techniques in Weather Prediction and Climate Change Studies, https://doi.org/10.5815/ijieeb.2012.01.07
[17] Beer, T. Climate variability and change: a perspective from the oceania region. Geosci. Lett. 1, 5 (2014). https://doi.org/10.1186/2196-4092-1-5
[18] Frawley, W. J. , Piatetsky-Shapiro, G., and Matheus, C. J. (1991). Knowledge Discovery in Databases: An Overview. AAAI/MIT Press
[19] Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R. (1996). Advances in Knowledge Discovery and Data Mining. AAAI/MIT Press.
[20] Han, J., Kamber, M., Pei, J., Kaufmann, M. (2011). Data Mining: Concepts and Techniques, 3rd Editon. United States of America, The Morgan Kaufman Publishers.
[21] Ayodele, T. (2010). Machine Learning Overview. https://www.researchgate.net/publication/221907649_Machine_Learning_Overview
[22] Mahesh, B. (2019). Machine Learning Algorithms –A Review. https://www.researchgate.net/publication/344717762_Machine_Learning_Algorithms_-A_Review
[23] Dubey, A. (2018). Discretization using Decision Trees. https://towardsdatascience.com/discretisation-using-decision-trees-21910483fa4b
[24] Bajaj, A. (2023). Performance Metrics in Machine Learning [Complete Guide]. https://neptune.ai/blog/performance-metrics-in-machine-learning-complete-guide
[25] Deepchecks Community Blog (2023). Understanding F1 Score, Accuracy, ROC-AUC, and PR-AUC Metrics for Models.
[26] Yahya, F. and Rafiq, M. (2020). The Influence of Air Pollution and Clean Energy on Tuberculosis: The Moderating Role of Urbanization. Iranian Journal of Public Health. 49(6), p. 1101, 10.18502/ijph.v49i6.3362
[27] Cohen, A., and Mehta, S. (2007). Pollution and Tuberculosis: Outdoor Sources. https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0040142
[28] Tremblay, G.A. (2007). Historical Statistics support a hypothesis linking Tuberculosis and Air Pollution caused by Coal. The International Journal of Tuberculosis and Lung Disease. 11(7), pp. 722-732, https://www.ingentaconnect.com/content/iuatld/ijtld/2007/00000011/00000007/art00003
[29] Park, S., Hwang, E., Park, H., Kim, J., Heo, J., Lee, K., Song, T., Kim, E., Ro, Y., Kim, S., and Kim, Y. (2003). Growth of Mycobacteria on Cabon Monoxide and Methanol. Journal of Bacteriology. 185(1), pp. 142-147,
[30] US Environmental Protection Agency (n.d.). Why acid rain is harmful? https://www3.epa.gov/acidrain/education/site_students/whyharmful.html
[31] E. Kudo, E. Song, L.J. Yockey Low ambient humidity impairs barrier function and innate resistance against influenza infection Proc. Natl. Acad. Sci. U.S.A., 116 (2019), pp. 10905-10910
[32] Tian H.Y., Zhou S., Dong L., Van T.P., Boeckel, Cui Y.j., Wu Y.R., Cazelles B., Huang S.Q., Yang R.F., Grenfell B.T., Xu B. Avian influenza H5N1 viral and bird midration networks in Asia. Proc. Natl. Acad. Sci. U. S. A. 2015;112:172–177. doi: 10.1073/pnas.1405216112. [33] Q. Bukhari, Y. Jameel, Will coronavirus pandemic diminish by summer? SSRN [Preprint] Available at https://papers.ssrn.com/sol3/papers 2020 cfm?abstract_id=3556998. Accessed 4 May 2025.
[34] Fredrick, T., Ponnaiah, M., Murhekar, M. V., J Jayaraman, Y., David, J. K., Vadivoo, S., & Joshua, V. (2015). Cholera outbreak linked with lack of safe water supply following a tropical cyclone in Pondicherry, India, 2012. Journal of health, population, and nutrition, 33(1), 31–38.
[35] Manogaran G., Lopez D. spatial cumulative sum algorithm with big data analytics for climate change detection. J. Comput. Elect. Eng. 2018:207–221.
[36] Asadgol Z., Mohammadi H., Kermani M., Badirzadeh A., Gholami M. The effect of climate change on cholera disease: The road ahead using artificial neural network. PLoS ONE. 2019;14(11) doi: 10.1371/journal.pone.0224813.
[37] Fredrick, T., Ponnaiah, M., Murhekar, M. V., Jayaraman, Y., David, J. K., Vadivoo, S., & Joshua, V. (2015). Cholera outbreak linked with lack of safe water supply following a tropical cyclone in Pondicherry, India, 2012. Journal of health, population, and nutrition, 33(1), 31–38.
[38] Righetto, L., Bertuzzo, E., Mari, L., Schild, E., Casagrandi, R., Gatto, M., Rogriguez-Iturbe, I., and Rinaldo, A. (2013). Rainfall mediations in the spreading of epidemic cholera, https://doi.org/10.1016/j.advwatres.2013.07.006
[39] J. Jofre, A.R. Blanch, F. Lucena Water-borne infectious disease outbreaks associated with water scarcity and rainfall events
[40] Wu X, Lu Y, Zhou S, Chen L, Xu B (2016) Impact of climate change on human infectious disease: Empirical evidence and human adaption. Journal of Environment International
[41] Gao, Q., Liu, Z., Xiang, J., Zhang, Y., Tong, M. X., Wang, S., Zhang, Y., Liu, Q., Jiang, B., & Bi, P. (2021). Impact of Temperature and Rainfall on Typhoid/Paratyphoid Fever in Taizhou, China: Effect Estimation and Vulnerable Group Identification. The American journal of tropical medicine and hygiene, 106(2), 532–542. https://doi.org/10.4269/ajtmh.20-1457
[42] Jacobson, T.A., Kler, J.S., Hernke, M.T. et al. Direct human health risks of increased atmospheric carbon dioxide. Nat Sustain 2, 691–701 (2019). https://doi.org/10.1038/s41893-019-0323-1