Effects of heatwave features on machine-learning-based heat-related ambulance calls prediction models in Japan

Deng Ke[a,*], Kiyoshi Takahashi[b], Jun'ya Takakura[b], Kaoru Takara[c], Bahareh Kamranzad[d]

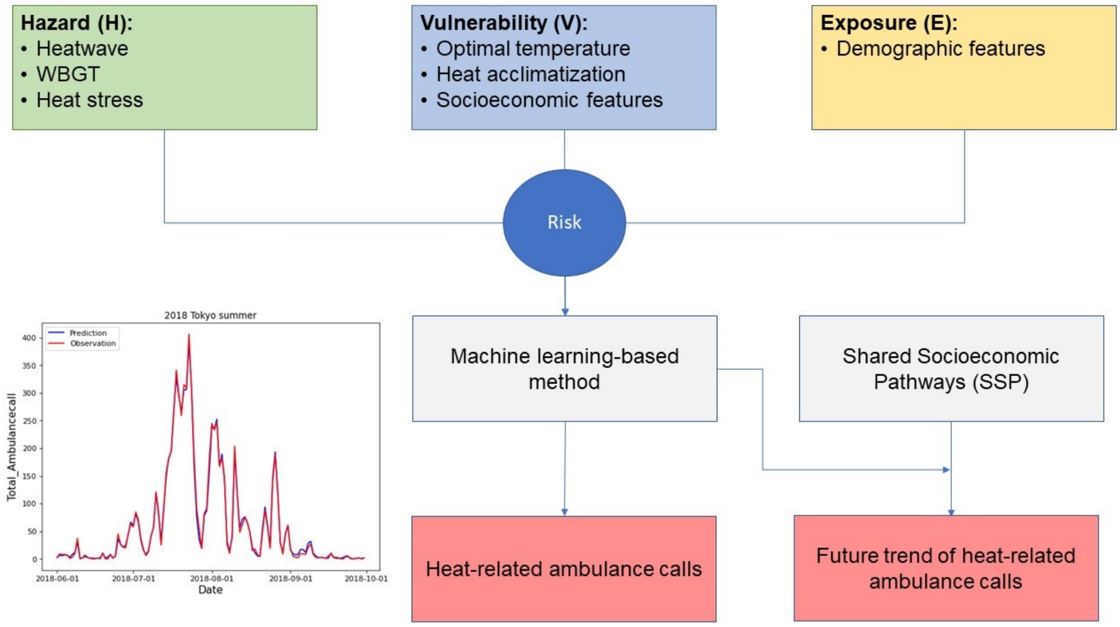
a. Graduate School of Advanced Integrated Studies in Human Survivability, Kyoto University, Yoshida-Nakaadachi 1, Sakyo-ku, Kyoto 606-8306, Japan
b. Center for Social & Environmental Systems Research, National Institute for Environmental Studies, 16-2, Onogawa, Tsukuba, Ibaraki, 305–8506, Japan
c. Disaster Prevention Research Institute (DPRI), Kyoto University, Uji, Kyoto 611-0011, Japan
d. Department of Civil and Environmental Engineering, University of Strathclyde, Glasgow, G11XJ, United Kingdom

*Corresponding author.
E-mail address: ke.deng.57z@st.kyoto-u.ac.jp

Highlights:

- Including heatwave features can significantly improve prediction accuracy.
- At the end of the 21st century, the total number of heat-related ambulance calls in Japan will increase four-fold under SSP-5.85.
- At the end of the 21st century, Shikoku, Tokai, and Kinki regions will have the highest ambulance call growth under SSP-1.26, SSP-2.45, SSP-5.85.
- Only the regional-level model with heatwave features can provide a desirable prediction accuracy in Okinawa.

Graphical Abstract:



Abstract

Researchers agree that there is substantial evidence of an increasing trend in both the frequency and duration of extreme temperature events. Increasing extreme temperature events will place more pressure on public health and emergency medical resources, and societies will need to find effective and reliable solutions to adapt to hotter summers. This study developed an effective method to predict the number of daily heat-related ambulance calls. Both national- and regional-level models were developed to evaluate the performance of machine-learning–based methods on heat-related ambulance call prediction. The national model showed a high prediction accuracy and can be applied over most regions, while the regional model showed extremely high prediction accuracy in each corresponding region and reliable accuracy in special cases. We found that the introduction of heatwave features, including accumulated heat stress, heat acclimatization, and optimal temperature, significantly improved prediction accuracy. The adjusted coefficient of determination (adjusted R^2) of the national model improved from 0.9061 to 0.9659 by including these features, and the adjusted R^2 of the regional model also improved from 0.9102 to 0.9860. Furthermore, we used five bias-corrected global climate models (GCMs) to forecast the total number of summer heat-related ambulance calls under three different future climate scenarios nationally and regionally. Our analysis demonstrated that, at the end of the 21st century, the total number of heat-related ambulance calls in Japan will reach approximately 250,000 per year (nearly four times the current amount) under SSP-5.85. Our results suggest that disaster management agencies can use this highly accurate model to forecast potential high emergency medical resource burden caused by extreme heat events, allowing them to raise and improve public awareness and prepare countermeasures in advance. The method proposed in Japan in this paper can be applied to other countries that have relevant data and weather information systems.

Keywords: climate change, heat-related impacts, emergency ambulance calls, Extreme Gradient Boosting, Shared Socioeconomic Pathways (SSP).

1. Introduction

The number of extreme weather events related to climate change has increased dramatically in recent years (Martínez & Bandala. 2018). Heatwaves are one type of extreme weather event and represent a major threat to public health and human society (Campbell et al., 2018). Several fatal heatwaves have already occurred in the 21st century, causing a large number of deaths in different regions. According to the World Meteorological Organization (WMO), the total number of deaths caused by heatwaves grew by 2300% between 1991–2000 and 2001–2010 (Robine et al., 2008; Vos et al., 2010). Moreover, continuing global warming and climate change have resulted in higher average temperatures. Previous research shows that even small changes in average temperature can result in considerable changes in the intensity and frequency of extreme weather events (Mearns et al., 1984). Analyses of long-term climate records show evidence of an increasing trend in both the frequency and duration of extreme temperature events (Fischer et al., 2010). Russo et al. (2014) proposed a model that projected heatwave events that have the same intensity as the 2010 heatwave in Russia will become common in the near future. Under these conditions, there is an urgent need to investigate the associations between meteorological information and heat-related health impacts. High-accuracy heat-related health prediction models are required to assist governments and individuals in taking proper countermeasures in advance.

Evaluating the impacts of severe heat-related extreme weather events on both mortality and morbidity has attracted increasing interest. Multiple analyses about the associations between temperature and mortality have indicated that region-specific risks change according to climate characteristics, geographic information, and socio-economic conditions (Chen et al., 2018; Nairn et al., 2018). A multi-country, multi-community study showed that heatwaves had significant cumulative associations with health outcomes in all countries, but results varied by community (Guo et al., 2017). Morbidity data, including ambulance calls and hospital admissions, are good candidates to examine the daily health impacts of heat because they are sensitive to changes in temperature (Campbell et al., 2018).

Some efforts have been made to predict heat-related morbidity (heat stroke) and the number of ambulance calls by using various types of meteorological data. Previous efforts to investigate the associations between high temperature and its impacts (including mortality, morbidity, or ambulance calls) can be broadly classified into two types. The first type of research uses numerical computational methodsor statistical methods to evaluate the heat-related health outcomes based on physical laws of weather. For example, Guo evaluated the hourly associations between heat and ambulance calls in Australia by using a distributed lag non-linear model (Guo et al., 2017). Kaisai et al. (2017) used a polynomial regression method to predict summer heatstroke risk in Tokyo. Nairn et al. (2018) proposed excess heat factors considering both long- and short-term effects of heat to calculate the number of deaths caused by heatwaves. Kodera et al. (2019) utilized an integrated computational technique to estimate daily peak core temperature elevation and daily water loss, and found that these metrics are important for predicting heat-related ambulance calls. Oka et al. (2021) proposed random forest-based prediction model with a unique metric named relative temperature to represent heat acclimatization when evaluate the number of heat stroke patients. However, differences between geographic information and socio-economic conditions are rarely considered in these studies, which results in model predictions that are only accurate in the target regions. This type of research also has been shown to have some limitations in properly estimating the effects of heat when the absolute temperature is not very high in early summer. The second type of research incorporates machine-learning (ML) or deep-learning methods. Typical examples include a random forest model developed by Wang et al. (2019) and a hybrid ML model developed by Ogata et al. (2021). Wang's model sacrificed prediction accuracy to study cities in different climate zones, although socio-economic features were considered. Ogata's model obtained high levels of prediction accuracy by applying under-sampling and bagging methods in their hybrid model, but their research only focused on the Kinki region and cannot be applied to other regions without training additional models. A high accuracy prediction model that can evaluate daily heat-related health impacts in different regions has yet to be developed.

Thus, we developed national- and regional-level models to evaluate the performance of ML-based methods on the prediction of daily heat-related ambulance calls in Japan. Heatwave features, demographic data, and socio-economic conditions are included in the models to properly evaluate health risks in different regions and on a national scale. Furthermore, we estimated future heat-related risks for different regions in Japan under different socio-economic scenarios. Such an approach will enable future projections of heatwaves and their impact on public health. The paper is organized as follows. We describe the datasets and methods in Section 2. The performances of the national- and regional-level models are presented Section 3. We discuss the limitations of the current model and future research directions in section 4 and summarize the research in Section 5.

2. Methods

Our aim was to develop a model to predict heat-related morbidity that can be applied to all the regions of Japan. For this purpose, four meteorological variables provided by the Japan Meteorological Agency—air temperature, relative humidity, solar radiation, and wind speed—were collected from 1991 to 2020. The agency has collected meteorological data since 1872 and provides weather data at the city level. We selected 47 main weather stations in 47 prefectures to express the heat stress experienced by people in different regions. In most prefectures of Japan, the population is concentrated in one or more main cities. Therefore, we selected weather stations that are located near major urban areas to represent the weather conditions in each prefecture. A more detailed description of the weather stations and their locations can be found in the supplementary materials, along with descriptions of special conditions that had to be considered. For example, even though Mito is the most populous city in Ibaraki Prefecture, solar radiation data were unavailable at that weather station, so we used data from Tsukuba, which is the second-most populous city in Ibaraki. Some weather stations also had missing values for some days. In these cases, we used data from the nearest weather station in place of the missing values.

The framework of this model was inspired by the concepts proposed by the Intergovernmental Panel on Climate Change (IPCC) in its Fifth Assessment Report (IPCC, 2014a). The risk of heat-related health impacts results from the interaction of heat-related hazards with vulnerability and exposure. In this research, the term “hazards” refers to the properties of heat-related hazards. It includes severity, duration, location, frequency, and other concepts related to heat-related hazards. The term “exposure” applies to the residents affected by heat-related hazards. Demographic factors for each prefecture were used to evaluate exposure. The term “vulnerability” refers to the sensitivity of residents’ health to extreme heat events. Factors including age, gender, and socio-economic status will influence the ability to tolerate extreme environments. Both “exposure” and “vulnerability” are influenced by a wide range of demographic and socio-economic features and processes that have been incompletely considered to date and that make quantitative assessments of future trends complex (IPCC, 2014b).

Hazards, exposure, and vulnerability were all used as inputs in the model to evaluate heat-related risk. Heat-related risk is generally quantified as health outcomes caused by heat-related hazards. Mortality has been assessed as the primary health outcome in previous research (Campbell et al., 2018). Although morbidity data are sometimes more sensitive to heat-related health impacts, the limitation of access to morbidity data makes it difficult to perform large-scale population-level studies (Campbell et al., 2018). We selected ambulance call data as the outcome in our model because it includes different levels of health impacts, ranging from mild (i.e., does not require hospitalization) to severe (requires hospitalization of 3 weeks and more). Death is also included in the data, but only in cases where death was confirmed at the initial examination. We used data from Japan's Fire and Disaster Management Agency (FDMA), which has provided heat-related ambulance call data since 2008. According to Japan's FDMA, Heat-related ambulance call refers to the situation that people are transported to medical institutions by emergency ambulance and are diagnosed with heat stroke after initial examination. Heat stroke here is defined as "a general term for disorders that occur when the body's water and salt (sodium, etc.) imbalance occurs due to a breakdown in the body's ability to regulate the heat in a hot environment. It includes sunstroke, heat cramps, heat exhaustion, and heatstroke. In this research, we used heat-related ambulance call data from June to September of each year in Japan's 47 prefectures from 2008 to 2020.

2.1. Heatwave features

The wet-bulb globe temperature (WBGT) is a worldwide heat index to estimate the co-effects of temperature, relative humidity, wind speed, and solar radiation. WBGT was first developed to estimate heat exposure in the marine military, where humidity was of concern for soldiers in military training clothes and subject to significant risk of heat stress (Yaglou, 1957). After WBGT was accepted and recommended by ISO 7243 (ISO, 1989), it became a popular tool used by researchers to measure the general heat stress for people working in hot and humid environments. Unlike in some countries where the summer temperatures are high but humidity is low, Japan experiences both high temperatures and high humidity in summer. Once the rainy season starts, the humidity remains high for months. Therefore, we chose WBGT rather than air temperature to evaluate heat stress for this research. Observed WBGT data have only been provided at limited weather stations since 2006 in Japan. To obtain WBGT data from 1991 to 2020, we used the estimation method proposed by Ono and Tonouchi (2014), which can estimate WBGT with a bias of less than 1.0°C with 98.3–99.8% confidence. The estimation formula is as follows:

$$WBGT = 0.735 \times Ta + 0.0374 \times RH + 0.00292 \times Ta \times RH + 7.619 \times SR - 4.557 \times SR^2 - 0.0572 \times WS - 4.064, \quad (1)$$

where *Ta* is standard air temperature (°C), *RH* is relative humidity (%), *SR* is daily solar radiation (kW/m^2), and *WS* is wind speed (m/s).

It should be noted that Ono’s formula was designed for hourly WBGT estimation while our research used daily average WBGT. Previous studies have noted that the estimation procedures reflected in Eq. (1) do not strictly follow the ISO standards for measuring WBGT (Takakura et al., 2019). When estimating daily average WBGT, it is necessary to calculate daytime WBGT and nighttime WBGT separately because of changes in solar radiation. Therefore, we tested four different methods based on Ono's formula (Eq. 1) to estimate daily average WBGT (see the supplemental material). The results indicate that Ono's formula has sufficient prediction accuracy even though it lacks physical rationality.

Heat stress accumulates when hot days continue for a long period, and it becomes more difficult to discharge the heat stress at night. Epidemiological studies in different countries have proved that heat-related health impacts are usually delayed for 2 or 3 days (Saez et al., 1995; Hajat et al., 2002; Honda et al., 2014), and the lag effect plays a significant role in properly evaluating accumulated heat stress. A method was proposed by Nairn that included a 3-day period to calculate overall heat stress during heatwave events (Nairn et al., 2015). However, the lag effect pattern varied in different countries and regions, making it difficult to set a reasonable lag effect for all regions. Honda et al. (2014) found that a lag of 0 days had the highest risk, followed by a lag of 1 day and so on. Therefore, specified weights need to be allocated for different lag days. To accomplish this, we combined concepts from previous research and defined the overall heat stress of day t as follows:

$$Heatstress_t = \sum_{n=0}^{k}(w_{t-n} \times WBGT_{t-n}) \quad , \tag{2}$$

where k is the number of days that the lag effect exists in a specific region; w_{t-n} is the daily weight of the lag effect, which should decrease as time passes. We used a 2-day lag effect in this research because evidence has indicated that significant lag effects continue for 2 days in Japan (Honda et al., 2014).

Heat acclimatization is the beneficial physiological adaptive response to a hot environment that helps people reduce the risks of serious heat illness (Kinney et al., 2008). Heat acclimatization occurs when repeated heat exposures are sufficiently stressful to invoke profuse sweating and elevate whole-body temperatures (Saunders et al., 2019). Evidence shows that progress of heat acclimatization might take 2 weeks with sufficient daily heat exposure (McGregor et al., 2015). Considering that the time required for heat acclimatization for different groups of people varies, we used two variables, WBGT_Previous15 and WBGT_Previous30, to represent the difference of WBGT between the current day and the average of the previous 15 (30) days. A relatively high value of WBGT_Previous15 (30) means that the WBGT has suddenly increased, and people might experience difficulties adapting to hot days. Heat acclimatization is used to capture the situation that a sudden increase of temperature cause people to be unable to adapt to the heat, even though absolute temperature not very high.

Previous studies have confirmed a V-shaped relationship between temperature and the mortality rate (Honda et al., 2014; Curriero et al., 2002). The mortality rate increases rapidly when the temperature becomes higher than an optimal temperature. This optimal temperature is location-specific and plays a significant role in heat-related health risk evaluation research (Vicedo-Cabrera et al., 2021). Similar result was found in Japan that different WBGT thresholds are required for hazard classification across different regions (Nakamura et al., 2022). The relationship between temperature and morbidity is more complicated because the level of heat-related illness needs to be considered. Therefore, we used three optimal WBGTs to characterize the WBGT thresholds in each prefecture: the 85th, 90th, and 95th percentile of previous 30 years WBGTs.

Accumulated heat stress, heat acclimatization, and optimal temperature are almost always considered in heatwave-related research. Therefore, to distinguish these features from weather information, we denoted these features as “heatwave features” in this research.

2.2. Demographic and socio-economic features

We also used the following demographic and socio-economic features of the 47 prefectures: population size and the ratio of the population over 65 years old in each prefecture. In addition, three integrated socio-economic factors, the Income Index, Health Index, and Education Index, were calculated on a sub-national scale by Global Data Lab (https://globaldatalab.org/) based on the Human Development Index (HDI). These indices can be used to assess the development of a country by assessing life expectancy at birth, mean years of schooling of adults, and gross national income per capita in different areas. In addition, 10 other socio-economic features were also selected to represent levels of heat vulnerability in each prefecture. All of these features have been shown to be related to heat-related health impacts (McGregor et al., 2015). Table 1 summarizes the details of all of the variables used in our research and the data sources.

2.3. Model training

We used a supervised ML method, Extreme Gradient Boosting (XGBoost), to develop our heat-related morbidity prediction model. XGBoost is a scalable, distributed gradient-boosted decision tree ML library (Chen and Guestrin, 2016). It provides a parallel tree boosting method and is a highly effective ML library for regression and classification problems. First, we pre-processed the training dataset, which included all variables listed in Table 1. The scales of demographic variables and socio-economic variables varied from

1 to 10,000,000. Because variables with large differences in scale might make the learning progress unstable. we used a normalization method to pre-process input data by scaling each variable to a range between 0 and 1. Then, we used randomized search cross validation with repeated 10-fold cross-validation to find the optimal hyperparameters of the XGBoost model. Finally, the adjusted coefficient of determination (adjusted R^2) and root mean squared error (RMSE) were utilized to examine model performance. Generally, a smaller RMSE value indicates better performance, whereas a higher adjusted R^2 value indicates better performance. Moreover, we applied SHapley Additive exPlanations (SHAP) on the national-level prediction model to explain the impacts of variables. SHAP is an open-source library that applies the Shapley value of cooperative game theory to ML models which handles well multicollinearity (Lundberg and Lee, 2017). It allows researchers to understand the impact of each explanatory variable used in an ML model. Features with higher SHAP values have higher impacts on the model.

Table 1 List of variables and data sources

Variables	Abbreviation	Units	Dataset sources
Weather information			
Wet Bulb Globe Temperature	WBGT	℃	Calculated based on data provid
The WBGT of the previous day	WBGT_1	℃	Calculated based on WBGT
The WBGT of the previous two day	WBGT_2	℃	Calculated based on WBGT
85th percentile WBGT in this region for 1991 – 2020 period	WBGT_85	℃	Calculated based on WBGT
90th percentile WBGT in this region for 1991 – 2020 period	WBGT_90	℃	Calculated based on WBGT
95th percentile WBGT in this region for 1991 – 2020 period	WBGT_95	℃	Calculated based on WBGT
The WBGT compared against the average WBGT over the past 15 days	WBGT_Previous15	℃	Calculated based on WBGT
The WBGT compared against the average WBGT over the past 30 days	WBGT_Previous30	℃	Calculated based on WBGT
Socio-economic and Demographic information			
Total population (prefecture level)	Total_Population		E-Stat (Ministry of Internal Affa
The rate of the population over 65 years old	Eld_Population_Rate		E-Stat (Ministry of Internal Affa
Health index	Health_index		Global Data Lab
Income index	Income_index		Global Data Lab
Education index	Education_index		Global Data Lab
The number of air conditioners per 1000 people	Air_Conditioner_per1000		E-Stat (Ministry of Internal Affa
The rate of single elder people	Isolated_Elder_Rate	%	E-Stat (Ministry of Internal Affa
The rate of people enrolled in high school	High_School_Enrollment	%	E-Stat (Ministry of Internal Affa
The rate of male compared against female	Male_Female_Rate	%	E-Stat (Ministry of Internal Affa
The rate of people who need nursing care	Need_Care_Rate	%	E-Stat (Ministry of Internal Affa
The number of convenience stores per 1,000 people	Konbini_per1000		E-Stat (Ministry of Internal Affa
The number of ambulance vehicles per 1,000 people	Ambulanca_per1000		E-Stat (Ministry of Internal Affa
The coverage of urban green space	Coverage_Urban_Green	%	Ministry of Land, Infrastructure
The coverage of urban greenway space	Coverage_Green_Way	%	Ministry of Land, Infrastructure
The coverage of urban park	Coverage_Urban_Park	%	Ministry of Land, Infrastructure
Heat-related emergency ambulance call	Total Ambulance Call		Fire and Disaster Management A

Four different prediction models were trained to find the best approach to predict heat-related morbidity using a dataset with selected variables: (1) a national-level model with heatwave features, (2) a national-level model without heatwave features, (3) regional-level models with heatwave features, and (4) regional-level models without heatwave features. The distribution of climate regions in Japan is summarized in the supplementary Material (Fig. S2). In general, prefectures located in the same climate region share a similar climate pattern and experience extreme weather simultaneously. We then compared the national and regional models to investigate the best approach to predict the number of heat-related ambulance calls. We assessed the differences between observed values in summer 2018 and predicted values of heat-related ambulance calls in different prefectures to evaluate model performance.

2.4. Global climate models and future projections

Daily temperature, relative humidity, wind speed, and solar radiation obtained from general circulation model (GCM) were used to calculate future WBGT which is the main input of weather information. Uncertainties in the projection of future heat-related risks arise from the low resolution of GCM data and the bias of weather information between GCM data and observation data. To incorporate the uncertainties, we used bias-corrected climate scenarios GCM based on the Coupled Model Intercomparison Project phase 6 (CMIP6) for Japan, which were provided by the National Institute of Environmental Studies, Japan (Ishizaki, 2021). Five GCMs (ACCESS-CM2, IPSL-CM6A-LR, MIROC6, MIP-ESM1-2-HR, and MRI-ESM2-0) were selected and processed by Ishizaki (2021) for impact and adaptation studies in Japan. These GCMs have already been bias-corrected and downscaled by Ishizaki (2021) to a spatial resolution of 1 × 1 km. In our research, the nearest grid points to the target weather station were selected to estimate future WBGT because the weather station data had already been used in training the prediction models.

Three projection scenarios, i.e, SSP-1.26, SSP-2.45, and SSP-5.85 were used to estimate future heat-related risks. The scenarios were based on the Shared Socioeconomic Pathways (SSPs 1, 2, and 5) and Representative Concentration Pathways (RCPs 2.6, 4.5, and 8.5). SSP-1.26 is known as the "sustainability" scenario, in which carbon emissions will be cut to zero by 2075. SSP-2.45 is a "middle of the road" scenario and assumes that carbon emissions will peak by 2050 and start to decline. SSP-5.85 is known as "fossil-fueled development"; in this scenario, carbon emissions will continue to rise throughout the 21st century. Historical GCM data were also used for comparison with observed values. Historical simulation data (2010–2014) were combined with GCM data under the SSP-2.45 scenario (2015–2019) for historical comparison. We calculated future WBGT under three different scenarios for each prefecture in Japan from 2020 to 2100. Finally, we obtained the future prediction of heat-related ambulance calls by providing the transformed input data to the national-level prediction model we developed. We only considered the impacts caused by climate change in the future. Demographic and socio-economic features for each prefecture were kept consistent with their status in 2020.

3. Results

We developed four different XGboost models to predict heat-related ambulance calls in Japan: two national models (one with and one without heatwave features) and two regional models (one with and one without heatwave features). The performance of each prediction model is shown in Table 2. Regional-level models were developed for 11 climate regions, and the average RMSE and adjusted R^2 of the 11 regions are used for comparison. It should be noted that the RMSE can only be compared between the models on the same level (i.e., regional or national) because the normalization process was different in the different training datasets (see the Methods for a discussion of normalization). The train RMSE and test RMSE are smaller for both the national and regional models with heatwave features. An adjusted R^2 will only increase if a newly added feature improves the model's predicting power, and adding irrelevant features to a regression model will result in a decrease in the adjusted R^2. Among our models, the value increases from 0.9061 to 0.9659 for the national models and from 0.9102 to 0.9860 for the regional models. Models with heatwave features have better prediction performance at both levels.

Table 2 Prediction of model performance

	National-level XGboost models		Regional-level XGboost models	
	With heatwave features	Without heatwave features	With heatwave features	Without heatwave features
Train RMSE	0.0087	0.0145	0.0106	0.0268
Test RMSE	0.0086	0.0143	0.0106	0.0280
Test adjusted R^2	0.9659	0.9061	0.9860	0.9102

We examined the trend of heat-related ambulance calls in 2018 to test prediction models. Figures 1 and 2 show the prediction performance of the four different models. In 2018, 95,137 people were sent to the hospital due to heat-related health problems. Three main prefectures, Tokyo, Aichi, and Fukuoka were selected to examine the models' performance in diverse regions. Although these three prefectures are geographically dispersed, their summer heat-related ambulance call trends have a similar pattern. In each case, the prediction model with heatwave features has a higher accuracy in estimating daily heat-related ambulance calls.

The national model with heatwave features performs well under most conditions but has some small errors on days that are not very hot. The regional models with heatwave features have the highest accuracy in predicting daily ambulance calls. In all cases, models without heatwave features have relatively lower prediction accuracy, but they can still match the trends of ambulance calls caused by extreme heat.

National-level model with heatwave features National-level model without heatwave features

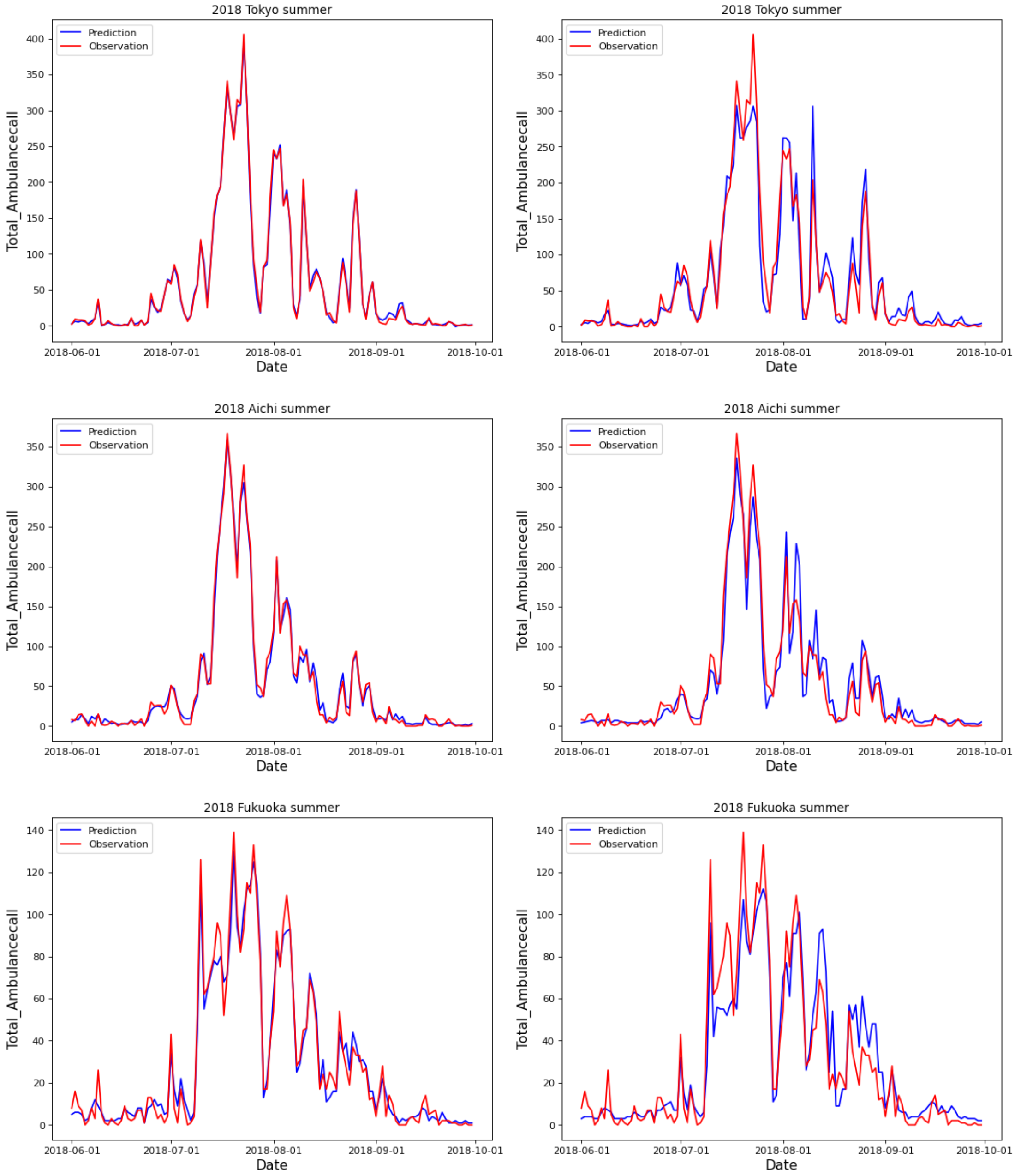


Figure 1, Prediction performance of national models for 2018 summer (from June to October) in Tokyo, Aichi and Fukuoka. Figures on the left side are results from model with heatwave features, while figures on the right side are results from model without heatwave features. Red lines indicate the observation values of total heat-related ambulance call, and the blue lines indicate the prediction values of total heat-related ambulance call. The vertical axis in these figures indicates the number of heat-related ambulance calls per 100,000 population. Prediction performance for other regions can be found in supplementary materials.

Regional-level model with heatwave features | Regional-level model without heatwave features

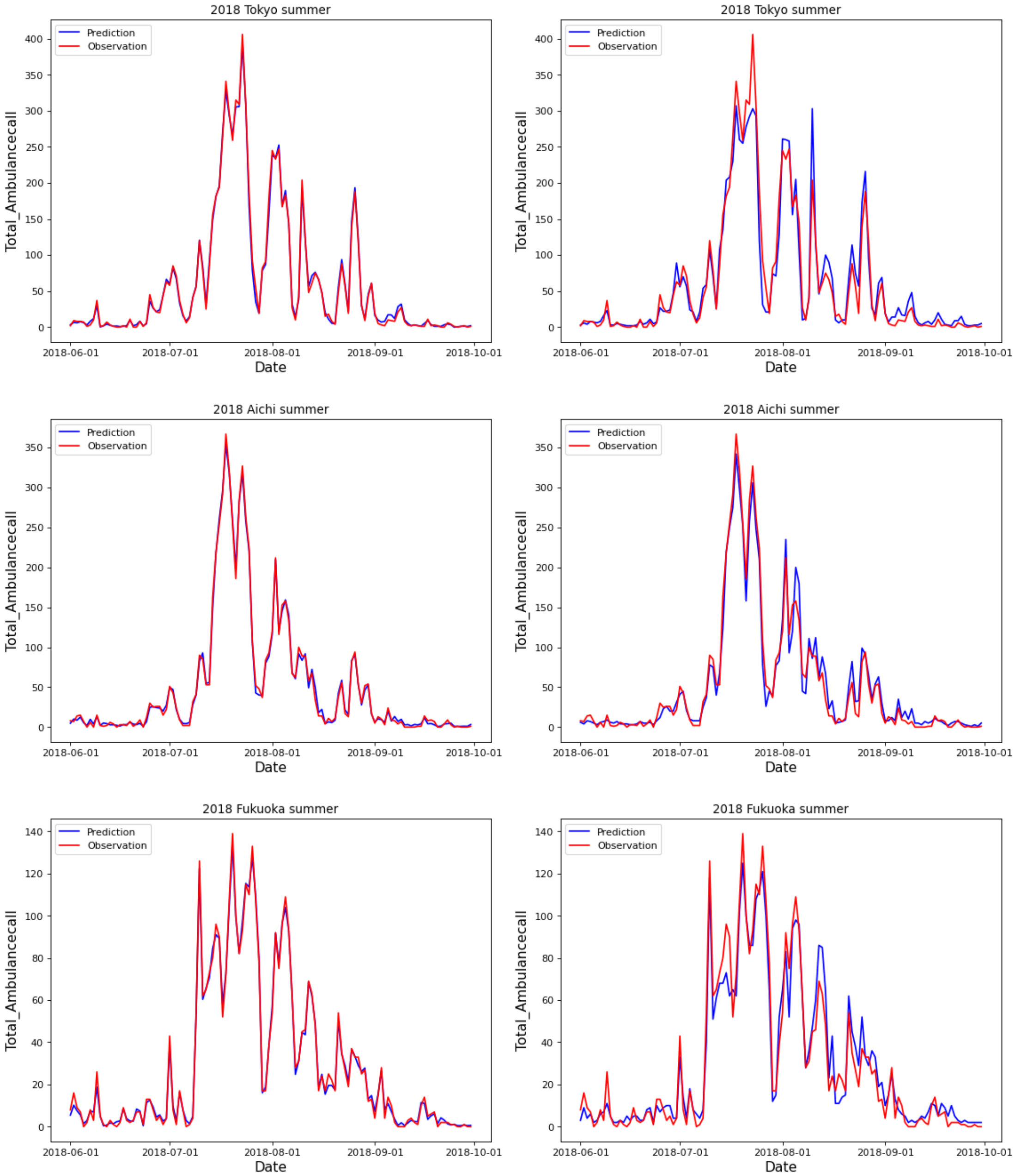


Figure 2, Prediction performance of regional-level models for 2018 summer (from June to October) in Tokyo, Aichi and Fukuoka. Figures on the left side are results from model with heatwave features, while figures on the right side are results from model without heatwave features. Red lines indicate the observation values of total heat-related ambulance call, and the blue lines indicate the prediction values of total heat-related ambulance call. The vertical axis in these figures indicates the number of heat-related ambulance calls per 100,000 population. Prediction performance for other regions can be found in supplementary materials.

SHAP values of the national model with heatwave features are shown in Fig. 3. WBGT and total population have the highest positive impacts on model output. Heatwave features (including WBGT) compared against the average WBGT over the past 30 days, the WBGT of the previous day, the WBGT compared against the average WBGT over the past 15 days, and the WBGT of the previous 2 days have significant positive impacts on the result. Another heatwave feature, the 85th percentile WBGT in a region for the 1991–

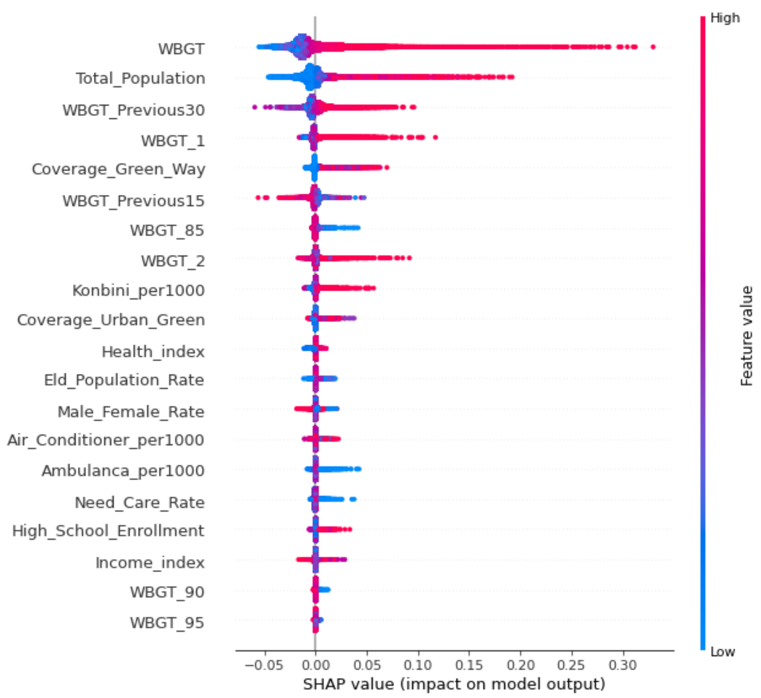


Figure 3, Importance of all features for the total heat-related ambulance call in national-level Xgboost model. Each dot represents 1 sample in the dataset. The blue to red color indicate low to high feature value. The x-axis shows the SHAP value. Red (blue) cluster with positive SHAP value means higher (lower) feature value results in higher number of total ambulance call, while. Red (blue) cluster with negative SHAP value means higher (lower) feature value results in lower number of total ambulance call. The feature WBGT in the top means the highest impact on model's output.

2020 period, has a negative impact on the result, which means a lower temperature threshold in a region might lead to a higher health risk under the same weather conditions. Socio-economic features have a relatively low influence on the model's output. The most influential socio-economic feature is urban green coverage, which positively impacts the model. It should be noted that a small SHAP value does not indicate that a feature (e.g., a socio-economic feature in this case) is not crucial in heat-related health risk research. It only explains how the difference in features influences the training progress of the prediction model in Japan.

We calculated the future trend of annual heat-related summer ambulance calls in Japan

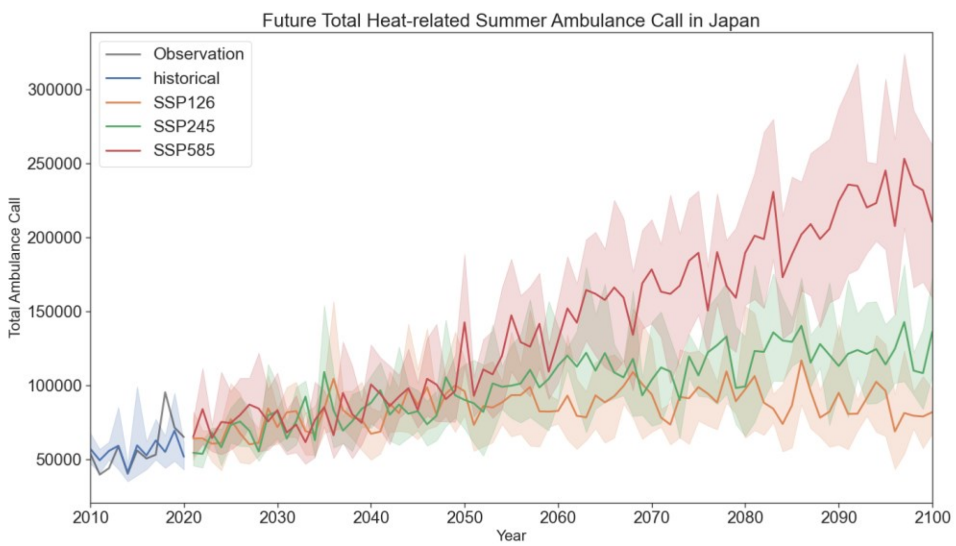


Figure 4, Temporal trends of future total heat-related summer ambulance call under different SSPs in Japan until 2100. The gray line indicates the observation of ambulance call from 2010 to 2020. The solid lines (exclude gray) indicate the mean annual ambulance call, and the shaded areas indicate the range of annual ambulance call estimated across the 5 GCMs.

by using the national model with heatwave features (Fig. 4). We first examined the difference between the observations provided by the Fire and Disaster Management Agency, Japan, and the historical values calculated by the five GCMs. The historical values are consistent with the observed values except in 2018. The annual average number of future heat-related ambulance calls under SSP-1.26, SSP-2.45, and SSP-5.85 were then estimated. There is no significant difference between the three scenarios until 2050. After 2050, the number of summer ambulance calls under SSP-5.85 increases rapidly, reaching nearly four times the current size by 2100, whereas the number of summer ambulance calls per year under SSP-1.26 is much smaller throughout the century. The trend under SSP-2.45 is distributed between those of SSP-1.26 and SSP-5.85.

The change in total number of ambulance calls for each region is shown in Fig. 5. We compared the average number of annual calls in 2011–2020 and with that of 2091–2100. The results show that Shikoku will experience the highest emergency ambulance call burden under SSP-5.85, with an almost five-fold increase by 2100, followed by the Tokai and Kinki regions (450% increase). Only three regions, Tohoku, Hokuriku, and North Kyushu, are predicted to have an increase of less than 400%.

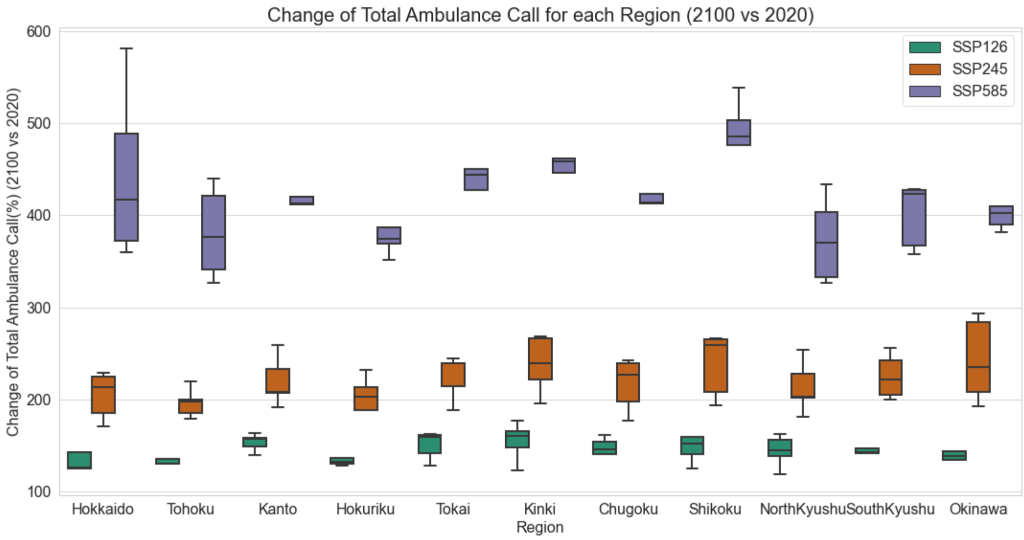


Figure 5, The changes of total ambulance call for each region in Japan under different SSPs. X axis represents the climate regions in Japan. Y axis represent the change of total ambulance call (%) in 2091 to 2100 compared to 2011 to 2020. The green, orange and purple boxes represent the the change of total ambulance call in each region under SSP126, SSP245, and SSP585. The lines inside the box are median values of change across 5 GCMs.

4. Discussion

The present study developed an ML-based method to predict daily heat-related ambulance calls with high accuracy. Both national- and regional-level models were developed to investigate effective solutions for predicting the impacts of future climate change on public health. Our results indicate that the national model has relatively high accuracy in most regions. However, there was one case where the national model showed lower accuracy (Fig. 6). Okinawa is the southern- and westernmost prefecture in Japan, and it has a humid subtropical climate that is different from the climate of mainland Japan.

These unique climate conditions limit the prediction accuracy of the national model, and only the regional model with heatwave features provides a desirable level of prediction accuracy in Okinawa. Hence, we recommend using the regional model with heatwave features for special cases such as this when the research target or application requires a high degree of accuracy. However, the national model is still the best choice and has the lowest computational cost when the research object is on the country level.

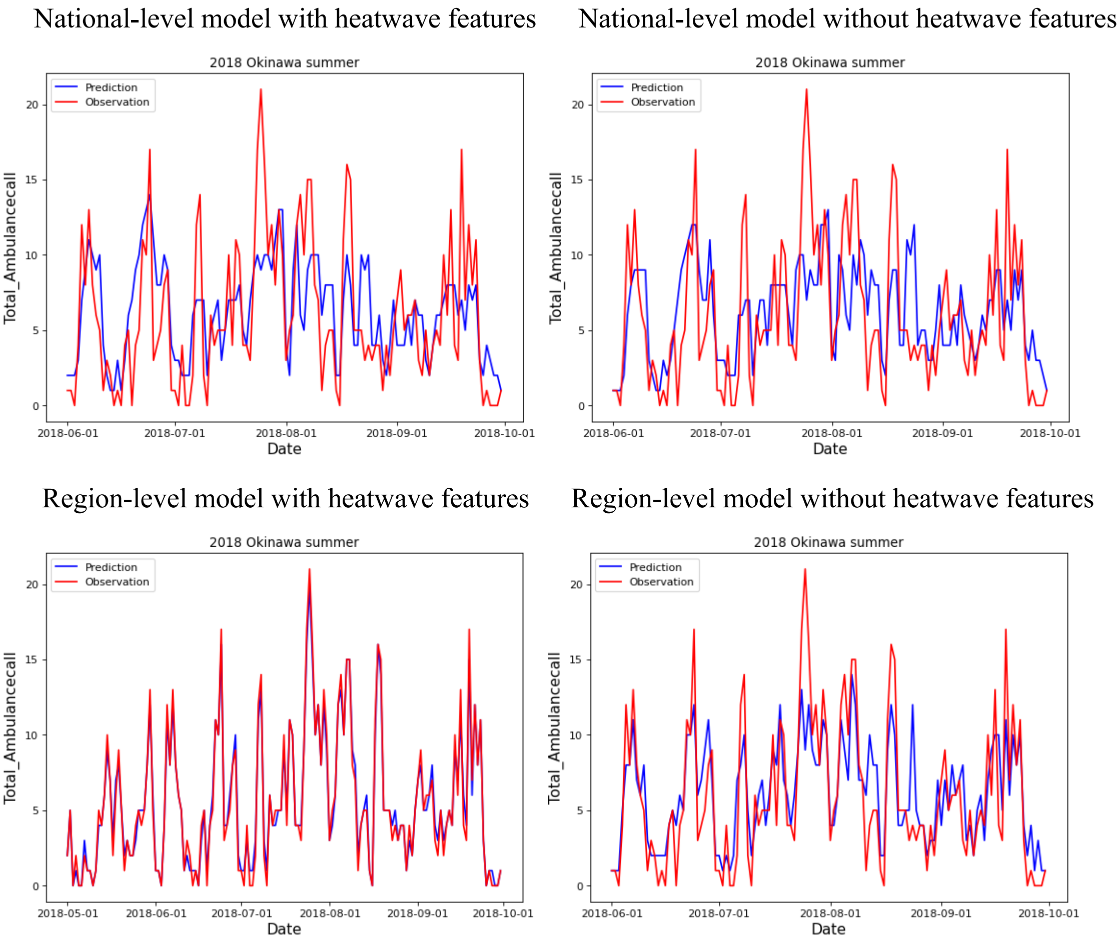


Figure 6, Prediction performance of 4 different models in Okinawa prefecture. Blues lines indicate predicted total ambulanca calls, while red lines indicate observed total ambulance calls.

This research investigated the significance of including heatwave features when developing a model that can be applied over a large area. We made comparisons of prediction performances between two well-used machine learning methods (XGBoost and Random Forest) under 3 different conditions (WBGT & heatwave features & socio-economic features, WBGT & socio-economic features, WBGT) in 4 Japanese cities. The results indicate that the improvements in prediction performance by introducing heatwave features and socio-economic features are also observed in other machine learning methods (please see supplementary material). We considered the vanilla XGboost and Randon Forest models, while the previous study proposed that a hybrid model and under-sampling are effective methods for improving a model's performance (Ogata et al., 2021). Using state-of-the-art ML methods while also including heatwave features is a promising direction for future research on heat-related health impacts. Another way to improve the prediction accuracy of the model is to consider additional socio-economic features related to heat impacts on public health. Socio-economic features have a complex interaction with individual morbidity and mortality because they interact with other determinants of health (McGregor et al., 2015). Including sufficient socio-economic features could help the model better deal with the relationships between extreme heat and health impacts, resulting in better prediction performance. However, the effects of some socio-economic features on public health are difficult to estimate quantitatively. It includes but not be limited to the countermeasures taken by local governments, implementation of heat warning systems, and social awareness of heat-related diseases (Campbell et al., 2018). Our research focused on heat-related health impacts in Japan. In the current models, the influence of these countermeasures is ignored during calculations because there are no significant differences in these features among the prefectures. However, this topic needs to be considered when building health impact prediction models for multiple countries.

We analyzed the SHAP values of the national model to explain the output. The SHAP results, however, are not in total agreement with those of previous studies. For example, previous studies have suggested that the level of dependency might be a direct risk factor for heatwave-associated mortality (Belmin et al., 2007). However, our result indicates that a lower rate of people who need nursing care increases the number of heat-related ambulance calls. In addition, single elderly people (not married or widowed) have been considered to be more vulnerable to extreme heatwave events because they generally receive less social support (Vandentorren et al., 2006), but the rate of single elderly people had no effect on the model’s output. It is difficult to explain the reasons behind this result because several socio-economic features are interrelated, and these conditions vary across countries. Older people who require caregiving can live in nursing homes in Japan, and nursing home staff members are trained to help keep their residents safe under a variety of conditions. A higher level of social welfare services might reduce the health risk of heatwaves for special populations, but the effect of socio-economic features in coping with heat-related disasters is not straightforward and requires further research. Although SHAP values can represent a feature's contribution to a change in model output (Lundberg and Lee, 2017), they do not directly explain the relationship between socio-economic features and heat-related ambulance calls.

We discussed the performance of our methods to predict heat-related health impacts. However, it is difficult to compare the prediction performance of the present study with previous studies because of the different algorithms, target variables (ambulance calls, death number, excess death, etc.), and datasets used. It is worth mentioning that ML-based methods may have poor prediction performance when meeting some unseen extreme climate conditions. When weather conditions change significantly from current conditions, the existing forecast model will become unreliable. However, we think the result in future prediction part is still reliable enough to estimate the change of future heat-related ambulance calls in country level. Considering the wide range of latitude of

Japan, it already has training data (temperature) with a wide input range. Therefore, even with the effects of global warming, only parts of Japan's cities will experience extreme hot days that not included in training data. In addition, what we are interested in prediction parts is the total number of heat-related ambulance calls in whole summer. Existing observation values are exceeded only on the hottest days of summer in some regions. Most of the summer days are still within this range and our model can have reliable prediction. Governments or research institutes can try to improve prediction accuracy by collecting more high-quality weather records and increasing the volume of the training dataset. We used heat-related ambulance calls as an indicator to investigate the exposure-response of heatwave events. However, some limitations usually exist when assessing environmental health impacts using ambulance call data. For example, not all people are willing to use emergency ambulance call services when they feel uncomfortable, especially those with mild symptoms. Young people and adults prefer to rest at home or go to nearby clinics alone, while elders who have a higher probability of suffering severe impacts prefer to use emergency ambulance call services. Due to this reason, the proportion of elders and severe patients in ambulance call data is often higher than the actual situation. In addition to this, ambulance data is still a desirable indicator for evaluating heat-related health impacts.

The ML-based heat-health impacts evaluation model proposed in the present study can also be applied to other countries or regions. However, modifications are required to adapt to the local characteristics. Although this study used daily health-related ambulance calls as an indicator, other indicators such as excess mortality rate or heat-related death can also be used to build ML-based models. Our model includes multiple socio-economic factors to improve the prediction performance. However, sufficient socio-economic data for model training are not always available, especially in some developing countries. It is recommended that researchers should carefully investigate the characteristics of local climate conditions and socio-economic conditions and then select the available potential factors associated with heat-related health outcomes. For example, wildfires occurred simultaneously with extreme heatwaves in some regions. Evidence showed that heatwave-caused air pollution also significantly affected the excess mortality rate (Katsouyanni, et al., 1993). Therefore, air pollution should be considered during model training in these regions to ensure a desirable prediction accuracy. Selecting factors based on the climate characteristics of the region or past studies can significantly improve models' performance.

There are several potential limitations to our study. First, the number of heat-related summer ambulance calls in the future was estimated using the national-level XGboost model. However, while we did consider the effects of climate change under different scenarios in the future, we did not include changes in demographic and socio-economic features. Considering the change of these features could improve our understanding of human vulnerability to extreme heat and may result in a lower number of predicted total ambulance calls. Some of the socio-economic features that were included in this research are challenging to estimate in the future, for example, the number of convenience stores and urban green spaces. Future models may be able to discard some of these features appropriately without a significant loss of prediction accuracy. Using advanced ML or deep-learning methods could also be used to compensate for any loss of prediction accuracy.

5. Conclusion

We developed an effective method to predict daily heat-related ambulance calls. The national-level model showed a high prediction accuracy that can be applied over a wide range of regions, and regional-level models were also developed when extremely high prediction accuracy is required. We found that incorporating heatwave features, including accumulated heat stress, heat acclimatization, and optimal temperature significantly improved prediction accuracy. The adjusted R^2 of the national model improved from 0.9061 to 0.9659 when heatwave features were included, showing that the model can predict the number of heat-related ambulance calls in most regions with low computing cost and high accuracy. At the same time, the adjusted R^2 of the regional models also improved from 0.9102 to 0.9860. Our analysis demonstrated that, at the end of the 21st century, the total number of heat-related ambulance calls will reach 250,000 per year (nearly four times the current number) under the SSP-5.85 scenario. Shikoku, Tokai, and Kinki regions will have the highest growth in heat-related ambulance calls, whereas Tohoku, Hokuriku, and North Kyushu are projected to have to lowest growth.

We also investigated the limitations of the developed models. For example, during model training, areas under the special climate conditions (e.g., Okinawa in this research) acted as outliers in the training dataset. Advanced strategies for regression on imbalanced data are required to improve model performance in these specific regions. Moreover, a multi-region, multi-country dataset with sufficient socio-economic features to investigate the prediction performance of this model in different countries is not currently available. Therefore, multi-region, multi-country socio-economic features collected under the same standard are required to design prediction models on a larger scale. Future research should utilize advanced ML or deep-learning methods to build more complex and accurate models. Deep learning methods with physical constraints could also be a promising solution for future climate problems.

The response speed to emergency ambulance calls is significant in saving patients' lives. The increasing frequency of extreme temperature events in the future will place more pressure on public health and emergency medical resources. Policymakers and disaster management agencies can use these highly accurate models to forecast incoming heatwave events, seek to raise and improve public awareness, and take appropriate countermeasures in advance. This research links the major areas of data science, medical science, and environmental science and can be a vital source for decision makers for future development. It provides an opportunity for additional future multi- or inter-disciplinary research in the above-mentioned fields.

Data Availability

All data supporting this study's findings are included in the article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was partly supported by the Environment Research and Technology Development Fund (JPMEERF20202002) of the Environmental Restoration and Conservation Agency of Japan.

References

Belmin, J., Auffray, J. C., Berbezier, C., Boirin, P., Mercier, S., de Reviers, B., & Golmard, J. L. (2007). Level of dependency: a simple marker associated with mortality during the 2003 heatwave among French dependent elderly people living in the community or in institutions. Age and ageing, 36(3), 298-303.

Campbell, S., Remenyi, T. A., White, C. J., & Johnston, F. H. (2018). Heatwave and health impact research: A global review. Health & place, 53, 210-218.

Chen, R., Yin, P., Wang, L., Liu, C., Niu, Y., Wang, W., ... & Zhou, M. (2018). Association between ambient temperature and mortality risk and burden: time series study in 272 main Chinese cities. *Bmj*, *363*.

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

Curriero, F. C., Heiner, K. S., Samet, J. M., Zeger, S. L., Strug, L., & Patz, J. A. (2002). Temperature and mortality in 11 cities of the eastern United States. American journal of epidemiology, 155(1), 80-87.

Fischer, E.M. and C. Schär, (2010): Consistent geographical patterns of changes in high-impact European heatwaves. Nature Geoscience, 3(6): 398–403. meteorology, 23(12), 1601-1613.

Guo, Y. (2017). Hourly associations between heat and ambulance calls. *Environmental Pollution*, *220*, 1424-1428.

Guo, Y., Gasparrini, A., Armstrong, B. G., Tawatsupa, B., Tobias, A., Lavigne, E., ... & Tong, S. (2017). Heat wave and mortality: a multi-country, multicommunity study. Environmental health perspectives, 125(8), 087006.

Hajat, S.; Kovats, R.S.; Atkinson, R.W.; Haines, A. Impact of hot temperatures on death in london: A time series approach. J. Epidemiol. Community Health 2002, 56, 367–372.

Honda, Y., Kondo, M., McGregor, G., Kim, H., Guo, Y. L., Hijioka, Y., ... & Kovats, R. S. (2014). Heat-related mortality risk model for climate change impact projection. Environmental health and preventive medicine, 19(1), 56-63.

IPCC. (2014a). Climate Change 2014 – Impacts, Adaptation and Vulnerability: Part A: Global and Sectoral Aspects: Working Group II Contribution to the IPCC Fifth Assessment Report. Cambridge: Cambridge University Press. doi:10.1017/CBO9781107415379

IPCC. (2014a). Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on

Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

Ishizaki, N. N., 2021: Bias corrected climate scenarios over Japan based on CDFDM method using CMIP6, Ver.1, NIES, doi:10.17595/20210501.001, (Reference date: 2022/01/07)

ISO. (1989) Hot environments - estimation of the heat stress on working man, based on the WBGT-index (wet bulb globe temperature) – ISO 7243 Standard. Geneva, Switzerland: International Organization for Standardization.

Kasai, M., Okaze, T., Mochida, A., & Hanaoka, K. (2017). Heatstroke risk predictions for current and near-future summers in Sendai, Japan, based on mesoscale WRF simulations. Sustainability, 9(8), 1467.

Katsouyanni, K., Pantazopoulou, A., Touloumi, G., Tselepidaki, I., Moustris, K., Asimakopoulos, D., ... & Trichopoulos, D. (1993). Evidence for interaction between air pollution and high temperature in the causation of excess mortality. Archives of Environmental Health: An International Journal, 48(4), 235-242.

Kinney, P. L., O'Neill, M. S., Bell, M. L., & Schwartz, J. (2008). Approaches for estimating effects of climate change on heat-related deaths: challenges and opportunities. Environmental science & policy, 11(1), 87-96.

Kodera, S., Nishimura, T., Rashed, E. A., Hasegawa, K., Takeuchi, I., Egawa, R., & Hirata, A. (2019). Estimation of heat-related morbidity from weather data: A computational study in three prefectures of Japan over 2013–2018. Environment international, 130, 104907.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

Martínez-Austria Polioptro, F., & Bandala Erick, R. (2018). Heat waves: health effects, observed trends and climate change. Extreme Weather, 107.

McGregor, G. R., Bessmoulin, P., Ebi, K., & Menne, B. (2015). Heatwaves and health: guidance on warning-system development. WMOP.

Mearns, L. O., Katz, R. W., & Schneider, S. H. (1984). Extreme high-temperature events: changes in their probabilities with changes in mean temperature. Journal of Applied Meteorology and Climatology, 23(12), 1601-1613.

Nairn, J. R., & Fawcett, R. J. (2015). The excess heat factor: a metric for heatwave intensity and its use in classifying heatwave severity. International journal of environmental research and public health, 12(1), 227-253.

Nairn, J., Ostendorf, B., & Bi, P. (2018). Performance of excess heat factor severity as a global heatwave health impact index. International journal of environmental research and public health, 15(11), 2494.

Nakamura, D., Kinoshita, H., Asada, K., Arimitsu, T., Yasumatsu, M., & Ishiwata, T. (2022). Trends in ambulance dispatches related to heat illness from 2010 to 2019: An ecological study. PLoS one, 17(11), e0275641.

Ogata, S., Takegami, M., Ozaki, T., Nakashima, T., Onozuka, D., Murata, S., ... & Nishimura, K. (2021). Heatstroke predictions by machine learning, weather information, and an all-population registry for 12-hour heatstroke alerts. Nature communications, 12(1), 1-11.

Oka, K., & Hijioka, Y. (2021). Prediction of the number of heatstroke patients transported by ambulance in Japan's 47 prefectures: proposal of heat acclimatization consideration. Environmental Research Communications, 3(12), 125002.

Ono, M., & Tonouchi, M. (2014). Estimation of wet-bulb globe temperature using generally measured meteorological indices. Jpn J Biometeor, 50(4), 147-157.

Robine, J. M., Cheung, S. L. K., Le Roy, S., Van Oyen, H., Griffiths, C., Michel, J. P., & Herrmann, F. R. (2008). Death toll exceeded 70,000 in Europe during the summer of 2003. Comptes rendus biologies, 331(2), 171-178.

Russo, S., Dosio, A., Graversen, R. G., Sillmann, J., Carrao, H., Dunbar, M. B., ... & Vogt, J. V. (2014). Magnitude of extreme heat waves in present climate and their projection in a warming world. Journal of Geophysical Research: Atmospheres, 119(22), 12-500.

Saez, M.; Sunyer, J.; Castellsague, J.; Murillo, C.; Anto, J.M. Relationship between weather temperature and mortality—A time-series analysis approach in barcelona. Int. J. Epidemiol. 1995, 24, 576–582.

Saunders, P. U., Garvican-Lewis, L. A., Chapman, R. F., & Périard, J. D. (2019). Special environments: altitude and heat. International Journal of Sport Nutrition and Exercise Metabolism, 29(2), 210-219.

Takakura, J. Y., Fujimori, S., Takahashi, K., Hijioka, Y., & Honda, Y. (2019). Site-specific hourly resolution wet bulb globe temperature reconstruction from gridded daily resolution climate variables for planning climate change adaptation measures. International journal of biometeorology, 63(6), 787-800.

Vandentorren, S., Bretin, P., Zeghnoun, A., Mandereau-Bruno, L., Croisier, A., Cochet, C., ... & Ledrans, M. (2006). August 2003 heat wave in France: risk factors for death of elderly people living at home. The European Journal of Public Health, 16(6), 583-591

Vicedo-Cabrera, A. M., Scovronick, N., Sera, F., Royé, D., Schneider, R., Tobias, A., ... & Gasparrini, A. (2021). The burden of heat-related mortality attributable to recent human-induced climate change. Nature climate change, 11(6), 492-500.

Vos, F., Rodriguez, J., Below, R., & Guha-Sapir, D. (2010). Annual disaster statistical review 2009: The numbers and trends.

Wang, Y., Song, Q., Du, Y., Wang, J., Zhou, J., Du, Z., & Li, T. (2019). A random forest model to predict heatstroke occurrence for heatwave in China. Science of the Total Environment, 650, 3048-3053.

Yaglou CP Minard D. (1957) Control of heat casualties at military training centers. AMA Arch Ind Health; 16: 302–16.