Sustainable Cities and Society 115 (2024) 105785

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs

Multi-region models built with machine and deep learning for predicting several heat-related health outcomes

Jérémie Boudreault [a,b,*], Annabel Ruf [a], Céline Campagna [a,b,c], Fateh Chebana [a]

[a] Centre Eau Terre Environnement, Institut national de la recherche scientifique, 490 de la Couronne, Québec, QC G1K 9A9, Canada
[b] Direction de la santé environnementale, au travail et de la toxicologie, Institut national de santé publique du Québec, 945 avenue Wolfe, Québec, QC G1V 5B3, Canada
[c] Département de médecine sociale et préventive, Faculté de médecine, Université Laval, 2300 rue de la Terrasse, Québec, QC G1V 0A6, Canada

ARTICLE INFO

Keywords:
Temperature
Extreme heat events
Mortality
Morbidity
Ensemble tree-based methods
Deep neural networks

ABSTRACT

As a result of climate change, populations worldwide will be exposed to more heat episodes. To ensure a sustainable future, cutting-edge tools must be developed to predict the health effects of heat and limit its consequences. However, current research has mainly focused on one health outcome in a single city/region, thus providing limited knowledge to improve society's resilience to extreme heat. In this study, a machine learning (ML) framework is introduced to predict several heat-related health outcomes in multiple regions simultaneously, using the province of Quebec (Canada) as a case study. Five ML models including penalized regression, ensemble tree-based models and deep neural networks were considered and compared. Models were trained to predict these health outcomes using various meteorological, regional and temporal predictors across all regions. Our results showed that deep learning models were the most promising, with out-of-sample R^2 of >60 % for most of the studied health outcomes. However, ensemble tree-based approaches also had the best performance for some health outcomes, and were more sensitive to weather variables and to heatwaves. By introducing novel ML-based tools for predicting heat risks in several regions, this study can guide climate change adaptation and help cities and society to become more healthy, resilient and sustainable.

1. Introduction

Climate change is recognized as one of the greatest threats to human health of the current century (Watts et al., 2021). Among its consequences are heat episodes that will occur earlier, last longer, be more frequent and intense (Meehl & Tebaldi, 2004). Extreme heat already affects population health worldwide, both in terms of mortality (e.g., Basu, 2009; Bi et al. 2023; Gosling et al., 2009; Kotharkar et al, 2024; Son et al., 2019) and morbidity (e.g., Li et al., 2015; Lin et al., 2023; Seong et al., 2024; Ye et al., 2012), but its consequences will be even more devastating in the future (Curtis et al., 2017; Huang et al., 2011; 2013). Yet state-of-art tools still need to be developed (or improved) to better allocate health resources and adequately target interventions that will, ultimately, reduce heat effects and lead to more a sustainable future (Ebi & Schmier, 2005; Kotharkar & Ghosh, 2022; McGregor et al., 2015).

With this in mind, recent developments in artificial intelligence (AI) and, specifically, machine learning (ML), have opened the door to a wide variety of novel applications towards sustainability, for example to improve population health (Fisher & Rosella, 2022; Morgenstern et al., 2020; Wiemken & Kelley, 2019), optimize green infrastructures (Shaamala et al, 2024) and mitigate urban heat (Li et al., 2023), among others. Indeed, ML can exploit larger datasets than ever (i.e., big data), model complex interactions and non-linear relationships, learn from multiple data points simultaneously and does not require strong assumptions on data distribution, leading to enhanced performance over traditional methods (Bi et al., 2019). In the context of climate change and increasing extreme heat events, various ML-based models were proposed to predict heat-related health impacts (e.g., Boudreault et al. 2023, Boudreault et al., 2024a, Ke et al. 2023, Kim and Kim 2022, Lee et al. 2022, Navares et al. 2018, Nishimura et al. 2021, Ogata et al. 2021, Park et al. 2020, Wang et al. 2019, Zhang et al. 2014). However, the aforementioned studies have shortcomings that need to be acknowledged.

First, most studies modelled a single health outcome such as mortality (Boudreault et al., 2023; Kim & Kim, 2022; Lee et al., 2022; Zhang et al., 2014) or heat-related illnesses (Ke et al., 2023; Nishimura et al., 2021; Ogata et al., 2021; Wang et al., 2019). Knowing that heat affects both mortality and morbidity, these models need to target multiple

* Corresponding author at: Centre Eau Terre Environnement, 490 de la Couronne, Québec, QC G1K 9A9 Canada.
E-mail address: jeremie.boudreault@inrs.ca (J. Boudreault).

https://doi.org/10.1016/j.scs.2024.105785
Received 14 February 2024; Received in revised form 23 July 2024; Accepted 27 August 2024
Available online 10 September 2024

health variables to provide a representative portrait of the heat-related burden for public health authorities. Second, existing studies have mainly focussed on a single city/region (Kim & Kim, 2022; Lee et al., 2022; Navares et al., 2018) or on multiple cities by (re)fitting the same model across locations (Boudreault et al., 2024a; Ogata et al., 2021; Zhang et al., 2014). This limits the potential use of big data in population health and ML's ability to take advantage of more data from multiple locations to improve its performance. Third, the few multi-region studies found in the literature often lacked appropriate validations during particular events of interest such as heatwaves (Kim & Kim, 2022; Ogata et al., 2021; Park et al., 2020; Wang et al., 2019). This can lead to selecting a model that will under-perform during periods when the most important health effects are observed (i.e., heatwaves). Fourth, most studies have only considered one ML model, mainly ensemble tree-based methods (Ke et al., 2023; Kim & Kim, 2022; Wang et al., 2019; Zhang et al., 2014). This completely ignores the neural networks family (i.e., deep learning) that could be more effective. Moreover, the lack a benchmark/comparison model (e.g., linear regression) limits the ability to conclude on the gain of ML over more traditional approaches. Finally, some ML models in the literature were literally presented as "black boxes", with no indications on models' explainability (Lee et al., 2022; Navares et al., 2018) or on how to adequately tune these models (Ke et al., 2023; Park et al., 2020; Zhang et al., 2014).

These drawbacks can lead to an incomplete picture of heat-related health impacts, suboptimal predictions and a lack of confidence in these models by non-AI experts, thus limiting the ability of cities and society to become more resilient and sustainable in a climate change context. In this study, a ML-based framework is proposed to model both mortality and morbidity (5 health variables) in two populations (general and elderly) during summer in multiple regions simultaneously. To that end, five ML models based on penalized regression, ensemble tree-based models and deep neural networks were considered and applied to the province of Quebec, Canada. These models were calibrated with various weather, socio-environmental and temporal predictors in a transparent way and were thoroughly validated. Finally, models' explainability was also provided, giving valuable insights for health authorities' preparedness and surveillance activities of extreme heat events to foster climate adaptation and act towards more resilient cities.

2. Material and methods

2.1. Study design and health outcomes

An ML-based framework was proposed to predict several heat-related health outcomes in multiple regions simultaneously (hereinafter referred to "multi-region models"), instead of creating individual models for each region, thus leveraging information and data from all available regions. The 15 southernmost health regions in the province of Quebec, Canada, were considered (top-right of Fig. 1), excluding regions #10, #17 and #18 due to their different climatology and their smaller population compared to the other regions (Boudreault et al., 2024b). The spatial scale of the health region (referred to simply as region for the rest of the paper) was chosen for this study as a trade-off between relevance for decision-makers (i.e., each region has its own public health team and director) and statistical power (i.e., having a large enough population in each region). These 15 regions represented over 99 % of the Quebec population in 2021, i.e. around 8.5M inhabitants. The summer season was defined as May to September. This project received ethics approval from the Human Research Ethics Committee of the National Institute of Scientific Research (CER-22-693). Fig. 1 presents an overview of the methodology.

Health data was made available by the *Institut national de santé publique du Québec* (INSPQ). Studied health variables were the daily numbers in each region of all-cause: (1) deaths, (2) hospitalizations, (3) emergency department visits (EDV), (4) ambulance transports and (5) *Info-Santé* 811 calls (a free telephone consultation service for non-urgent

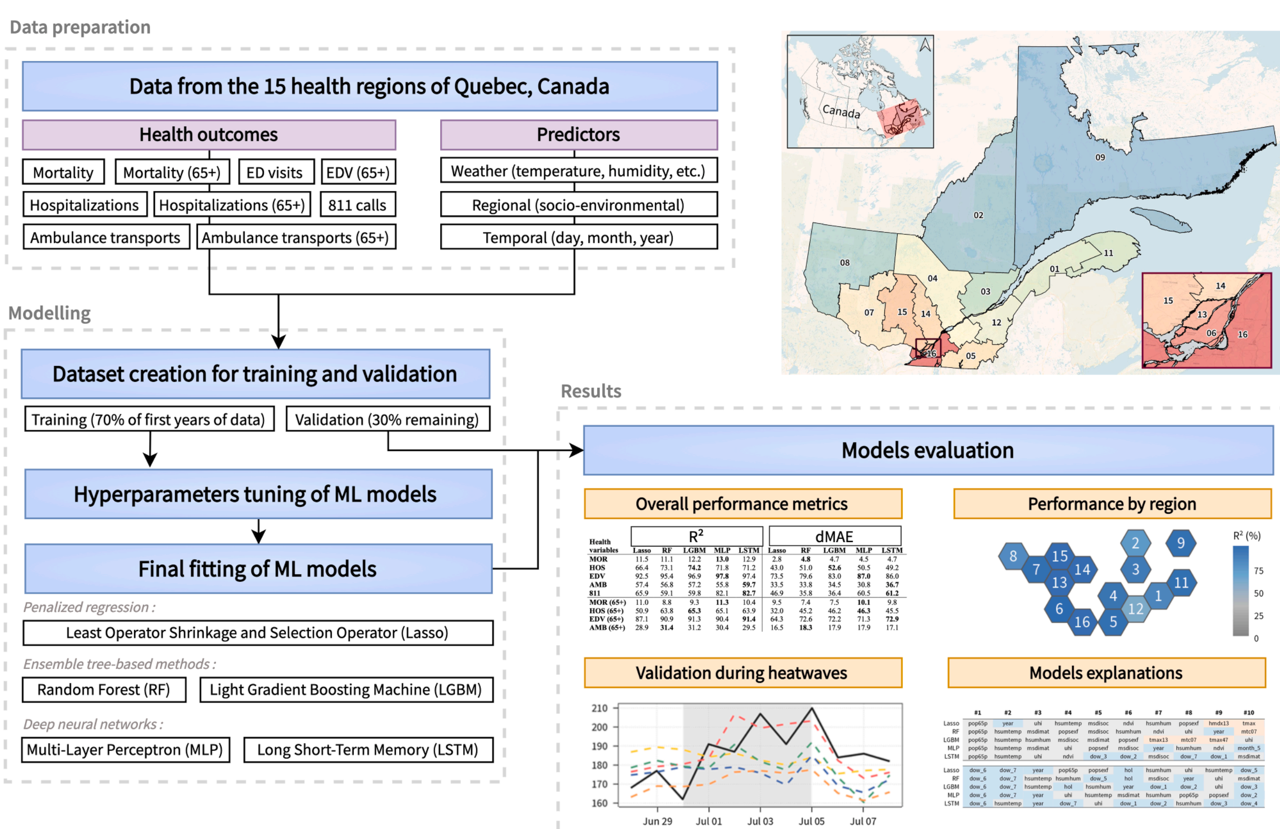

Fig. 1. Overview of the methodology.

health issues) (Table 1). All-cause health variables were chosen over cause-specific variables (e.g., cardiovascular diseases, heat-related illnesses) because they are more readily available in Quebec and already tracked by public health authorities through the health surveillance system of extreme weather events called SUPREME (Toutant et al., 2011). In addition, all these health variables were shown to be heat-related in previous epidemiological studies performed in the same study area (Boudreault et al., 2024b; Lebel et al., 2019) and in international literature reviews (Campbell et al., 2018; Cheng et al., 2019; Ye et al., 2012). Health variables were provided for two populations: (1) the general population (of all ages), and (2) the elderly sub-population (aged 65 and over), known to be more vulnerable to the effects of heat (Benmarhnia et al., 2015). This led to 9 studied health outcomes for this study (i.e., 5 health variables in two populations), as the data for 811 calls was only available for the general population (Table 1). Mortality and hospitalization data were available from 1996, EDV and ambulance transports from 2014 and 811 calls from 2008. All health outcomes were considered until 2019, prior to the COVID-19 pandemic.

2.2. Weather, regional and temporal predictors

Three types of predictors were considered to model the 9 health outcomes of interest: (1) meteorological predictors, for the short-term effect of weather on health, (2) regional predictors, to account for the socioeconomic and environmental differences between regions, and (3) temporal predictors, to consider potential temporal patterns or trends in the modelled health outcomes (Table 2).

For weather predictors, gridded daily data at 1 km × 1 km over North America from 1980 to 2021 from NASA's Daymet database were used (Thornton et al., 2022). Pixel values of minimum (*tmin*) and maximum (*tmax*) temperature, as well as vapour pressure, were extracted and aggregated at the region level by weighting each pixel with the number of units located in each pixel using the *AQgéobâti* database (Adresses Québec, 2023). This led to population-weighted weather times series for each region. Mean temperature was computed from the average of daily maximum and minimum temperatures. Various humidity metrics related to different health outcomes such as relative humidity, dew point and humidex (Davis et al., 2016) were computed from vapour pressure and temperature using *MetPy* in Python (May et al., 2022). Lagged values of meteorological predictors up to 7 days were considered using the value at lag 0, the mean of values at lags 1 to 3 days and the mean of values at lags 4 to 7 days (Boudreault et al., 2024a). Temperature variation variables were also considered including Mean Temperature Change (MTC), computed from ($tmean_0$ – $tmean_d$), where $tmean_0$ is the current mean temperature and $tmean_d$ is the mean temperature at $d = 1$, 3 and 7 previous days, and Diurnal Temperature Range (DTR), computed from (*tmax* – *tmin)* at lags 0, 1, 2 and 3 days. Air pollution could not be included in this study because daily data was only available in a few regions throughout the study period. No explicit heatwave variables (e.g., intensity, duration, etc.) were included in the models. Indeed, ML models should be able to develop their own understanding of extreme heat events on the basis of the weather variables provided, as shown in a previous study (Boudreault et al., 2023).

Regional predictors included demographic, socioeconomic, environmental and climatological characteristics (Table 2). Total population (used as an offset in the models), proportion of elderly (i.e., people aged 65 years and more) and proportion of women were provided by the *Ministère de la Santé et des Services Sociaux* (MSSS) du Québec from 1996 to present (MSSS, 2022). The Material and Social Deprivation Index (MSDI) developed by INSPQ from data of the canadian census at the dissemination area level was used (INSPQ, 2024). MSDI values for the social and material component of the index were aggregated at the region level by weighting the MSDI with the population in each dissemination area. Then, a linear interpolation was performed to obtain values between each census years (1996, 2001, 2006, 2011 and 2016) for the whole studied period. Two built environment variables related to urban heat were considered: the Normalized Difference Vegetation Index (NDVI), a commonly known index of density of vegetation ranging from -1 to 1, and a measure of the Urban Heat Island (UHI), computed as the difference between surface temperature and a surrounding milder rural reference value, both provided by the *Centre d'enseignement et de recherche en foresterie de Sainte-Foy* (CERFO, 2022). NDVI and UHI values available at 15m x 15m resolution in 2013 were weighted by the number of units in each pixel (with a buffer of 15 m) to obtain NDVI and UHI population-weighted values in all regions (same as what was done for weighting meteorological variables above). The 2013 values of NDVI and UHI were applied to the whole studied period (i.e., 1996–2019) as done previously (Pascal et al., 2021). Finally, 20-year averages of mean temperature and relative humidity by region during summer (May to September) were considered as climatological predictors. They were calculated from daily Daymet values, centred on the years 1980, 1990, 2000, 2010 and 2020 (only 10 and 11 years were used respectively for 1980 and 2020, as Daymet data was only available from 1980 to 2021). Then, a linear interpolation was performed to get smooth yearly values between these 10-year periods. Mean values of regional predictors for each region are reproduced in Table S1.

Regarding temporal predictors, they included days of the week (one binary variable for each day), public holiday (0 or 1), month (one variable for each month from May to September) and year (as a numeric variable) (Table 2). In total, there were 47 predictors considered for the different models.

2.3. Machine learning models

Predicting each of the 9 health outcomes (Table 1) based on weather, regional and temporal predictors (Table 2) can be seen as supervised learning problem. Hence, five machine learning (ML) models adapted to that end were considered: Least Absolute Shrinkage and Selection Operators (Lasso), Random Forest (RF), Light Gradient Boosting Machine (LGBM), Multi-Layer Perceptron (MLP) and Long Short-Term Memory (LSTM). These models were selected to represent a diversity of architectures, amounts of data required, potential performance and novelty of

Table 1
Overview of the studied health outcomes.

	Health outcome	Age	Abbreviation	Mean ± SD	Range	Years
1	**Mortality**	All	MOR	0.2 ± 0.1	0.0–1.2	1996–2019
2	**Hospitalizations**	All	HOS	2.4 ± 0.7	0.6–6.4	1996–2019
3	**EDV**	All	EDV	16.0 ± 7.8	6.0–45.1	2014–2019
4	**Ambulance transports**	All	AMB	2.1 ± 0.5	0.4–5.1	2014–2019
5	**811 calls**	All	811	5.4 ± 1.3	1.3–10.9	2008–2019
6	**Mortality (65+)**	65+	MOR (65+)	1.0 ± 0.5	0.0–9.9	1996–2019
7	**Hospitalizations (65+)**	65+	HOS (65+)	6.0 ± 2.2	0.0–27.0	1996–2019
8	**EDV (65+)**	65+	EDV (65+)	20.9 ± 8.7	8.2–63.9	2014–2019
9	**Ambulance transports (65+)**	65+	AMB (65+)	6.1 ± 1.3	0.8–14.4	2014–2019

Values are expressed in daily counts per 10^4 inhabitants across all regions for the May to September period. SD = Standard deviation. Range = Minimum and maximum daily values. EDV = Emergency department visits. 811 = Telephone consultation line for non-urgent health issue.

Table 2
Weather, regional and temporal predictors considered in this study.

	Units	Abbreviation	Mean ± SD	Range	Source
Weather predictors					
Mean temperature	°C	*tmean*	16.0 ± 4.8	-2.6–29.6	Daymet
Minimum temperature	°C	*tmin*	10.5 ± 4.9	-6.5–24.3	Daymet
Maximum temperature	°C	*tmax*	21.5 ± 5.4	-0.9–35.9	Daymet
Mean humidex	-	*hmdx*	21.5 ± 8.1	-10.7–44.5	Daymet
Mean dew point	°C	*tdew*	10.3 ± 5.1	-10.6–24.2	Daymet
Relative humidity	%	*relh*	69.6 ± 9.8	19.8–95.7	Daymet
MTC over 1 day	°C	*mtc01*	0.0 ± 2.6	-15.3–11.8	Daymet
DTR	°C	*dtr*	11.1 ± 3.8	1.3–29.4	Daymet
Regional predictors					
Total population (*offset*)	10^4 hab.	*poptot*	51.6 ± 46.4	9.0–206.6	MSSS
Elderly population (65+)	%	*pop65p*	15.2 ± 3.6	7.4–27.6	MSSS
Women population	%	*popsexf*	50.2 ± 0.6	48.7–52.0	MSSS
MSDI – Social component	-	*msdisoc*	-0.6 ± 1.0	-2.6–1.8	INSPQ
MSDI – Material component	-	*msdimat*	0.9 ± 1.8	-2.1–6.3	INSPQ
NDVI	-	*ndvi*	0.48 ± 0.06	0.36–0.60	CERFO
UHI measurement	°C	*uhi*	7.3 ± 1.9	3.3–10.5	CERFO
Historical summer temperature	°C	*hsumtemp*	16.0 ± 1.8	12.2–18.5	Daymet
Historical summer rel. humidity	%	*hsumhum*	69.5 ± 2.3	64.4–73.8	Daymet
Temporal predictors					
Day of week	-	*dow_1* to *dow_7*	-	0 or 1	-
Holiday	-	*hol*	-	0 or 1	-
Month	-	*month_5* to *month_9*	-	0 or 1	-
Year	-	*year*	Depends on health outcomes		-

All data span from 1996 to 2019 across all regions for May to September months. Lagged variables were also used in the models, but not shown in the table (i.e., tmean13 [at lags 1 to 3 days], tmean47 [lags 4 to 7 days], tmin13, tmin47, tmax13, tmax47, hmdx13, hmdx47, tdew13, tdew47, relh13, relh47, mtc03, mtc07, dtr1 [dtr at lag 1 day], dtr2, dtr3). Total population by region was used as an offset for the response variable (i.e., daily count / total population) rather than a predictor. SD = Standard deviation. Range = Minimum and maximum values. MTC = Mean temperature change. DTR = Diurnal temperature range. MSDI = Material and social deprivation index. NDVI = Normalized difference vegetation index. UHI = Urban heat island. MSSS = Ministère de la santé et des services sociaux. INSPQ = Institut national de santé publique du Québec. CERFO = Centre d'enseignement et de recherche en foresterie de Sainte-Foy.

approach. In addition, most of these models were used in previous related studies, thus allowing for comparisons of our results with the latter (e.g., Boudreault et al. 2023, 2024a, Lee et al. 2019, Nishimura et al. 2021, Ogata et al. 2021, Park et al. 2020, Wang et al. 2019).

Lasso is a multiple regression model that automatically performs variable selection and regularization by penalizing the absolute size of the coefficients during model training (Tibshirani, 1996). RF is an ensemble tree-based method in which a forest of fully developed decision trees is built using bootstrapped datasets (Breiman, 2001). LGBM is also an ensemble tree-based method, but it uses trees with fewer leaves (called weak learners) that are sequentially added to the ensemble based on the residual of the last fitted tree(s) (Friedman, 2001; Ke et al., 2017). MLP and LSTM both belong to the family of deep learning models. MLP is the vanilla feedforward neural network that is characterized by multiple layers of fully connected nodes (neurons) inspired by the architecture of the human brain (Chapter 6 in Goodfellow et al. 2016). LSTM is a type of recurrent neural networks, which are naturally adapted for time series data, with the addition of memory cells than can retain information over long sequences (Hochreiter & Schmidhuber, 1997). More information about these models can be found in the above references or in Boudreault et al. (2023).

Following the holdout method for time series, ML models were calibrated using the first 70 % of available years of data (*training*) and validated using the remaining 30 % of years of data (*test*). Training years (and test years) were 1996–2012 (2013-2019) for mortality and hospitalizations, 2014–2018 (2019) for EDV and ambulance transports, and 2008–2016 (2017–2019) for 811 calls in both populations. Given this split of data, the models were thus validated in a close to reality context, i.e. with weather, regional and temporal predictors over years of data that the models have never seen in their training. It should be noted here that all regions were used in the training of the models. In other words, there is no external validation of the models on new regions. This was intended because the models developed here are meant to be used in Quebec only, and not in other provinces/countries where the administrative structure of healthcare may differ.

In Lasso, MLP and LSTM, predictors were first scaled to the 0-1 range as these models require scaled predictors. Then, given the low evidence found about optimal tuning of ML models in our specific context, an extensive grid search was performed. Tested hyperparameters (including models' architectures) are reported in Table S2. The training dataset was used to find the hyperparameters that minimize the mean square error (MSE) in a 5-fold cross-validation grouped by years for Lasso, RF, LGBM and MLP and in a holdout validation of 30 % of years of the training dataset for LSTM (because of the time series nature of this model). Models' fitting was performed in Python using *scikit-learn* for Lasso, RF, LGBM and MLP (Pedregosa et al., 2011) and *keras* for LSTM (Gulli & Pal, 2017).

2.4. Models' performance and explainability

Models' performance was assessed in three different ways using the test dataset and based on two performance metrics: coefficient of determination (R^2) (1) and decrease in Mean Absolute Error (*dMAE*) compared to a non-informative model (i.e., a model containing only an intercept) (2) :

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \tag{1}$$

$$dMAE = \frac{MAE_0 - MAE}{MAE_0} = 1 - \frac{\sum |y_i - \hat{y}_i|}{\sum |y_i - \bar{y}|} \tag{2}$$

where y_i is the observed response variable, $\hat{y}_i$ is the predicted response variable, MAE is the mean absolute error of the model, MAE_0 is the mean absolute error of a non-informative model and $\bar{y}$ is the mean response variable in the training dataset (equivalent to the intercept in a null model).

First, the overall model's performance across all summer days of test years of data and regions was computed. Second, performance metrics were computed separately across the 15 regions to see how the performance was distributed among regions (again, on the test years of data only). These two first performance metrics correspond to the performance for the "heat-related" effect, as they are computed over all summer days. Third, the behaviour of the models during two well-known heatwaves in Quebec was visually inspected: the 2018 heatwave from June 30th to July 5th (Lebel et al., 2019) and the 2021 heatwave from June 6th to June 9th (Lamothe et al., 2023) (Fig. S1).

Note that the 2018 heatwave was part of the training set for EDV and ambulance transports, while the 2021 heatwave occurred during the COVID-19 pandemic that had an important effect on reported numbers (Lamothe et al., 2023). These two heatwaves had to be selected due to the absence of heatwaves in 2019. Health and predictors data in 2021 were only used for this specific heatwave validation (i.e., they were not used in models' calibration nor in other types of evaluation). Validation during heatwaves was presented for daily mortality/morbidity values aggregated across the 15 regions (i.e., for the whole province of Quebec). This last validation of the models during specific extreme heat episodes was added to validate models' performance for the "heatwave" effect.

In addition to performance, models' explainability was also provided. Although many methods exist to that end and are still emerging (see Rasheed et al. (2022) for a review), feature importance (FI) was chosen for the current study as it can give a quick overview of the most important predictors for any type of model, thus allowing to compare results of various modelled health outcomes (n=9), machine learning models (n=5) and tested predictors (n=47) at a glance (e.g., Boudreault et al. 2023, Zhang et al. 2014). FI was computed based on permutation, which consists of randomly shuffling each predictor and quantifying the increase in mean square error (MSE) on the training dataset after shuffling. The more the MSE increases, the more the predictor was important to the model's predictive power. This process was repeated 100 times for each predictor and the average increase of MSE was taken. FI metrics were converted to ranks, rank #1 being the most important predictor and #p (i.e., the number of predictors) the least important one. Only the 15 most important variables were reported (out of a total of 47).

3. Results

The obtained results are described below, while their discussion and interpretation are presented in the Discussion (Section 4). Sections 3.1 and 3.2 present respectively the performance of the models across Quebec and by region through the summer. Overall, deep learning models (MLP and LSTM) seemed to outperform other approaches for most health outcomes in terms of out-of-sample R^2 and dMAE values. In section 3.3, performance during two major heatwaves (2018 and 2021) are reported. LGBM was the most sensible model to extreme heat events and better predicted the health effects during these two periods. Finally, Section 3.4 highlights the most important predictors for all models and health outcomes. Overall, regional and temporal predictors were the most important variables, followed by weather factors such as temperature metrics. Optimal hyperparameters found for each ML model and health outcomes were reported in Supplementary Material (Table S3).

3.1. Performance across all regions

Models' performance in terms of R^2 and dMAE was computed on the test years of data across all regions in Table 3. All ML approaches performed well to predict daily morbidity variables (i.e., HOS, EDV, AMB and 811) with out-of-sample R^2 values of 60–95 % for both general and elderly populations (Table 3, left side). However, mortality in both populations and ambulance transports in the elderly population were less well predicted with R^2 values of ~10 % and ~30 %, respectively, indicating weaker performance of all models to predict these health outcomes. Overall, ensemble tree-based methods (RF, LGBM) and neural networks (MLP and LSTM) outperformed Lasso. However, the difference in performance depended on the modelled health outcome. For example, there was a 10–15 % increase in R^2 for hospitalizations and a 5 % increase for EDV in both populations for RF, LGBM, MLP and LSTM compared to Lasso. The highest increase in R^2 was 15–20 % for 811 calls in neural networks (MLP and LSTM) compared to Lasso. For mortality and ambulance transports, the difference between Lasso and other models was <5 %. A few combinations of models and health outcomes had slightly worse performance compared to Lasso (e.g., RF, LGBM and LSTM for mortality in the elderly population; RF, LGBM and MLP for ambulance transports in the general population). These findings were similar when looking at the decrease in Mean Absolute Error (dMAE) performance criteria (Table 3, right side), although some slight differences were noted. For example, RF was the best model for mortality based on dMAE, while it was MLP based on R^2. Across all health outcomes, LSTM and MLP were overall the most performing models for both performance criteria, but RF and LGBM were also sometimes the best approaches.

3.2. Performance by region

The performance metrics were also reported separately for each of the 15 regions in terms of R^2 for the general population (Fig. 2). dMAE values by region were also analyzed but were not shown because the same regional patterns were observed. Results for mortality (a) showed that R^2 reached up to 30–35 % for regions #1 and #4 in all models. In other regions, R^2 was closer to the overall value of 10 % found in Section 3.1. For hospitalizations (b), only region #8 was less well predicted by all models. Hospitalizations in regions #6 and #13, respectively Montreal and Laval (which are highly populated regions, recall Table S1), were not well modelled by Lasso (R^2 of ~30 %), but had a much better performance with the other ML models (R^2 > 80 %). The same behaviour was also observed for EDV (c) in regions #1 and #12 where performance was weaker for Lasso than for the other ML models. There was a high variability of performance in the modelled ambulance transports (d), with R^2 values ranging 0–80 %. Finally, 811 calls (e) were poorly modelled in regions #2, #3, #4 and #5, but regions #3 and #5 were improved with Lasso, MLP and LSTM models compared to RF and LGBM.

Results in the elderly population were comparable to the ones observed in the general population, with notable improvements in some regions by ensemble tree-based models and neural networks compared to Lasso (Fig. S2). The greatest regional variability of performance was observed for mortality (a) and ambulance transports (d) in the elderly population, as also observed in the general population. In terms of hospitalizations (b) and EDV (c), these two health variables were slightly

Table 3
Performance metrics during out-of-sample test years in terms of R^2 (left side) and decrease in Mean Absolute Error (dMAE) (right side) across all regions.

Health outcomes	R^2 (%)					dMAE (%)				
	Lasso	RF	LGBM	MLP	LSTM	Lasso	RF	LGBM	MLP	LSTM
MOR	11.5	11.1	12.2	**13.0**	12.9	2.8	**4.8**	4.7	4.5	4.7
HOS	66.4	73.1	**74.2**	71.8	71.2	43.0	51.0	**52.6**	50.5	49.2
EDV	92.5	95.4	96.9	**97.8**	97.4	73.5	79.6	83.0	**87.0**	86.0
AMB	57.4	56.8	57.2	55.8	**59.7**	33.5	33.8	34.5	30.8	**36.7**
811	65.9	59.1	59.8	82.1	**82.7**	46.9	35.8	36.4	60.5	**61.2**
MOR (65+)	11.0	8.8	9.3	**11.3**	10.4	9.5	7.4	7.5	**10.1**	9.8
HOS (65+)	50.9	63.8	**65.3**	65.1	63.9	32.0	45.2	46.2	**46.3**	45.5
EDV (65+)	87.1	90.9	91.3	90.4	**91.4**	64.3	72.6	72.2	71.3	**72.9**
AMB (65+)	28.9	**31.4**	31.2	30.4	29.5	16.5	**18.3**	17.9	17.9	17.1

Best performance metrics are in bold for each health outcome.

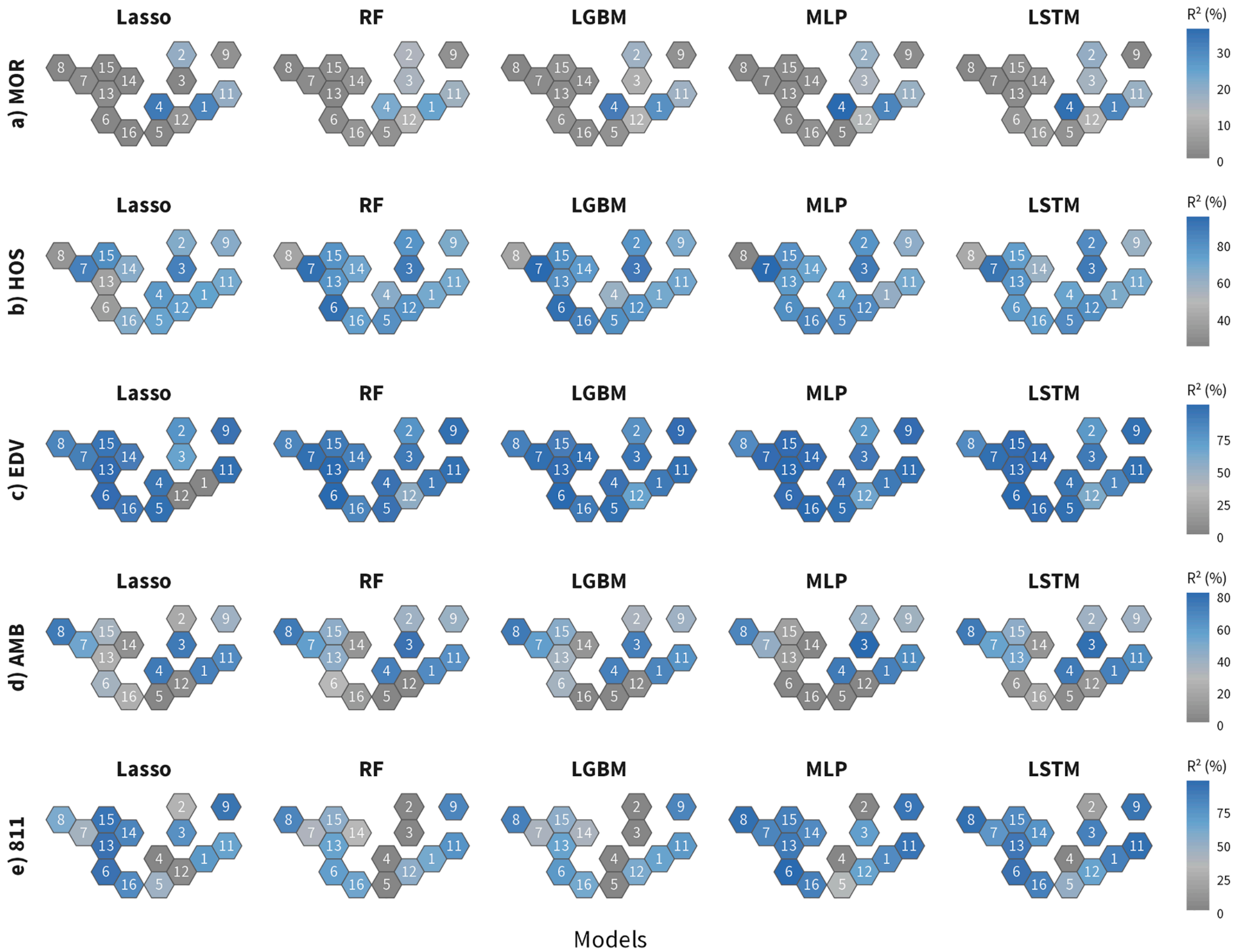

Fig. 2. Regional performance in terms of R^2 in the 15 regions for all models in the general population for (a) mortality, (b) hospitalizations (c) emergency department visits, (d) ambulance transports and (e) 811 calls. Each region is represented by a hexagon of equal size, irrespective of the actual area of the region. The geographical position of the regions has been preserved. The number in the hexagon indicates the region code as in Fig. 1. Note that the scale of R^2 values is different across health outcomes as the goal is to compare regional performance for each health outcome, rather than to compare health outcomes with each other (refer to Table 3 to that end).

less well predicted in some regions for the elderly population compared to the general population.

3.3. Validation during heatwaves

Observed and predicted values during 2018 and 2021 heatwaves across the province of Quebec were reproduced in Fig. 3. The 2018 heatwave showed a high increase in observed daily mortality and ambulance transports, but no clear increasing trend in hospitalizations, EDV and 811 calls (solid black lines in Fig. 3i). For mortality (i.a), LGBM predicted with the greatest accuracy the increased mortality, followed by MLP. For hospitalizations (i.b) and 811 calls (i.e), all models predicted well the trends in daily counts, but there was not an increased heatwave effect. For EDV (i.c), only RF predicted a peak in visits, while the observed data did not show any. For ambulance transports (i.d), the two peaks also seemed better modelled with these two models (LGBM and MLP) compared to others (Lasso, RF and LSTM), even though the second peak was slightly underestimated by all models at different levels. During the 2018 heatwave, same results were generally found in the elderly population (Fig. S3i) with some exceptions. For example, RF also modelled the peak of mortality well in the elderly population in 2018 along with LGBM (i.a). Also, only LGBM and LSTM reproduced the first increase in ambulance transports during 2018 heatwave in the elderly population, but all models underestimated the second peak later during that heatwave (i.d). There was no notable difference in hospitalizations (i.b) and EDV (i.c) in 2018 in the elderly population compared to the general population.

The 2021 heatwave was highly influenced by changes in mortality and morbidity due to the COVID-19 pandemic. Indeed, most models were no longer aligned with the observed values (Fig. 3ii). Still, the patterns in the models during that heatwave could be analyzed. For example, the mortality peaked less in 2021 than in 2018 and none of the models predicted an increase in mortality during that period (ii.a). It is worth recalling that this heatwave was not as hot and long as the 2018 one (Fig. S1). The hospitalizations were still insignificantly impacted by the heatwave and all models reproduced that behaviour well (ii.b). The peak of EDV (ii.c) and ambulance transports (ii.d) observed in 2021 was well detected by most models, except RF for ambulance transports. Finally, 811 calls were predicted to increase by all models, but not such an increase was observed (ii.e). There was no noteworthy difference in the results for the elderly population during the 2021 heatwave compared to the general population (Fig. S3ii).

3.4. Models' explainability

The 15 first most important predictors (out of 47) for all studied health outcomes and models were extracted for the general population (Table 4). Overall, regional predictors, such as *pop65p* (percentage of population 65 and over), *hsumtemp* (historical summer temperature), *hsumhum* (historical summer humidity) and *msdisoc* (social deprivation

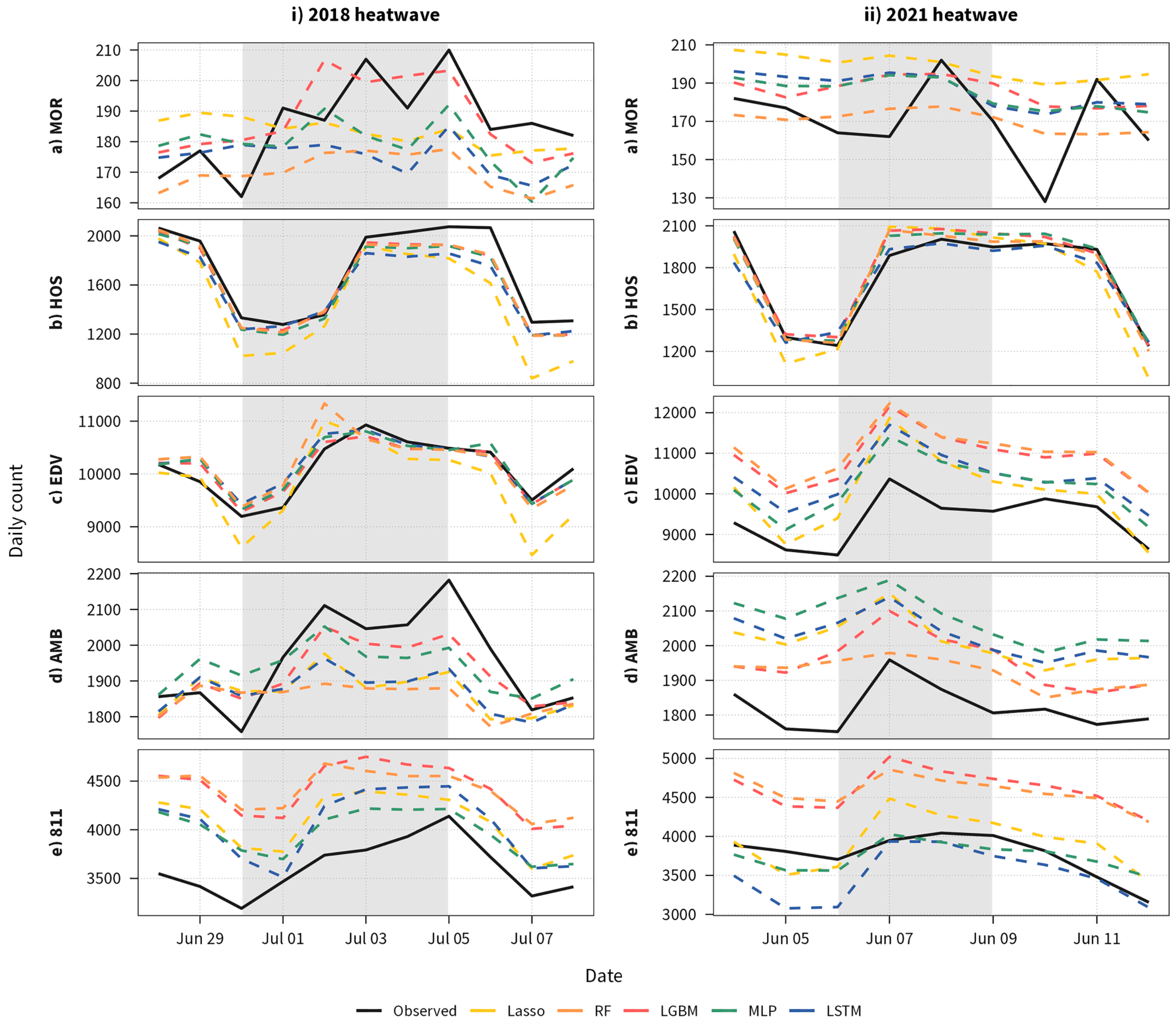

Fig. 3. Observed (solid black line) and predicted (dashed coloured lines) daily counts of (a) mortality, (b) hospitalizations (c) emergency department visits, (d) ambulance transports and (e) 811 calls in all studied regions for the general population during (i) 2018 and (ii) 2021 heatwaves. Heatwaves' periods are highlighted in grey.

index), appeared in the first 5 ranks for all health outcomes and models. These variables can account for regional differences in the modelled health outcomes. Temporal predictors were also among the most important variables in ranks 1 to 15, depending on the health outcome and/or model. The most important temporal variables were associated with daily (*dow_*), yearly (*year*) and, less frequently, monthly (*month_*) patterns in the modelled health outcomes. It is worth nothing that Saturday (*dow_6*) and Sunday (*dow_7*) were the two most important predictors in hospitalizations for all models (Table 4b).

Weather variables seemed less important (at ranks 7 and above) than regional and temporal predictors across all models (Table 4). However, major differences between models and health outcomes were noted. In terms of models, weather variables were generally more important in Lasso, RF and LGBM than in MLP and LSTM. In terms of health outcomes, weather predictors were more important for mortality (a) and ambulance transports (d), at ranks 7 to 10 approximately. They were respectively *tmax* (maximum temperature), *mtc07* (mean temperature change over 7 days), *tmean* (mean temperature) and *hmdx* (humidex) for mortality and *mtc07, tmax47* (maximum temperature at lags 4 to 7 days), *tmax13* and *hmdx13* for ambulance transports, according to Lasso, RF and LGBM. For hospitalizations (b), temperature exposures at 4 to 7 previous days appeared in ranks 13 to 15 in RF, but these variables did not seem important in the other models. For EDV (c), *tmax13, mtc07* and *tmax47* were the most important weather predictors (ranks 14–15) in RF and LGBM. For 811 calls (e), 4 to 7 days lagged humidex (*hmdx47*) and dew point (*tdew47*) exposure were among the 15 most important predictors in Lasso and LGBM, while *tdew* and *tmin* (minimum temperature) were found in MLP.

In the elderly population, similar results as above were found, i.e. regional and temporal variables were the most important predictors followed by weather variables (Table S4). Weather variables were again more important for mortality (a) and ambulance transports (d) in the elderly population, but with some notable differences. Change of temperature over 3 (*mtc03*) and 7 (*mtc07*) days seemed more important than direct temperature exposure in the elderly population compared to the general population for these two health outcomes. Even though weather

Table 4
Top 15 most important predictors for all models in the general population for (a) mortality, (b) hospitalizations, (c) emergency department visits, (d) ambulance transports and (e) 811 calls.

Feature importance (FI) rank

Models		#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15
a) MOR	Lasso	pop65p	year	uhi	hsumtemp	msdisoc	ndvi	hsumhum	popsexf	hmdx13	tmax	msdimat	month_5	tdew	mtc07	hmdx47
	RF	pop65p	hsumtemp	msdimat	popsexf	msdisoc	hsumhum	ndvi	uhi	year	mtc07	tmax47	mtc03	tmean47	hmdx47	relh13
	LGBM	pop65p	hsumtemp	hsumhum	msdisoc	msdimat	popsexf	tmax13	mtc07	tmax47	uhi	relh13	mtc01	ndvi	tmax	relh47
	MLP	pop65p	hsumtemp	msdimat	uhi	popsexf	msdisoc	year	hsumhum	ndvi	month_5	month_6	month_9	dow_6	dow_3	dow_1
	LSTM	pop65p	hsumtemp	uhi	ndvi	dow_3	dow_2	msdisoc	dow_7	dow_1	msdimat	month_8	month_9	month_7	dow_4	dow_5
b) HOS	Lasso	dow_6	dow_7	year	pop65p	popsexf	hol	hsumhum	uhi	hsumtemp	dow_5	msdimat	month_7	month_8	month_9	ndvi
	RF	dow_6	dow_7	hsumtemp	hsumhum	dow_5	hol	msdisoc	year	uhi	msdimat	popsexf	pop65p	hmdx47	tdew47	tmean47
	LGBM	dow_6	dow_7	hsumtemp	hol	hsumhum	year	dow_1	dow_2	uhi	dow_3	msdisoc	dow_4	msdimat	month_9	month_7
	MLP	dow_6	dow_7	year	uhi	hsumtemp	msdimat	hsumhum	pop65p	popsexf	dow_2	dow_1	dow_3	hol	dow_4	msdisoc
	LSTM	dow_6	hsumtemp	year	dow_7	uhi	dow_1	dow_2	hsumhum	dow_3	dow_4	pop65p	popsexf	month_9	month_5	msdimat
c) EDV	Lasso	hsumtemp	msdimat	hsumhum	uhi	ndvi	pop65p	msdisoc	dow_6	dow_1	dow_7	year	dow_2	hol	month_5	popsexf
	RF	msdisoc	hsumtemp	msdimat	pop65p	dow_6	hsumhum	dow_1	dow_7	uhi	year	dow_2	popsexf	dow_5	mtc07	tmax13
	LGBM	msdisoc	hsumtemp	msdimat	hsumhum	dow_6	pop65p	dow_1	dow_7	dow_2	popsexf	uhi	hol	month_5	tmax47	tmax13
	MLP	hsumtemp	popsexf	msdimat	msdisoc	uhi	ndvi	pop65p	dow_1	hsumhum	year	dow_6	month_5	month_9	dow_2	dow_7
	LSTM	hsumtemp	msdisoc	msdimat	uhi	popsexf	pop65p	hsumhum	dow_1	dow_6	year	ndvi	dow_7	dow_2	month_5	dow_4
d) AMB	Lasso	pop65p	hsumhum	tmean	hmdx	hmdx13	msdisoc	uhi	tmax	tmax13	tdew13	hmdx47	msdimat	tdew47	hsumtemp	tmin47
	RF	pop65p	hsumhum	hsumtemp	ndvi	uhi	msdimat	popsexf	msdisoc	tmax	mtc07	year	dtr	mtc03	relh	tmean
	LGBM	pop65p	hsumhum	hsumtemp	popsexf	ndvi	msdisoc	tmax	msdimat	dow_1	mtc07	year	uhi	dtr	tmax13	mtc03
	MLP	pop65p	hsumhum	popsexf	hsumtemp	uhi	msdimat	msdisoc	ndvi	dow_2	dow_3	dow_6	month_8	month_5	month_9	dow_7
	LSTM	pop65p	hsumhum	msdisoc	ndvi	popsexf	uhi	hsumtemp	msdimat	month_8	month_9	month_7	month_5	dow_2	dow_3	month_6
e) 811	Lasso	year	hsumtemp	uhi	popsexf	pop65p	hmdx47	tdew47	hmdx13	tdew13	msdimat	ndvi	dow_6	hmdx	dow_7	hsumhum
	RF	hsumhum	msdisoc	msdimat	year	dow_6	popsexf	hsumtemp	dow_7	uhi	ndvi	pop65p	dow_1	tdew47	hmdx47	relh47
	LGBM	hsumhum	msdisoc	year	dow_6	msdimat	dow_7	popsexf	uhi	dow_1	pop65p	hsumtemp	ndvi	month_5	month_6	month_8
	MLP	year	msdimat	uhi	pop65p	hsumhum	msdisoc	hsumtemp	popsexf	ndvi	dow_6	dow_7	dow_1	tdew	month_5	tmin
	LSTM	dow_1	year	dow_2	dow_3	msdimat	month_5	dow_4	dow_5	month_6	hsumhum	month_9	msdisoc	month_7	pop65p	month_8

Variable type: Weather Temporal Regional

Weather variables are in light orange, temporal predictors in blue and regional predictors in grey. Refer to Table 2 for a description of each abbreviation.

variables were less important for hospitalizations (b), the 4 to 7 days lagged temperature exposure was consistent in the two populations. For EDV (c), temperatures with longer lags were found to be more important in the elderly compared to the general population.

4. Discussion

This study considered 5 machine learning (ML) models to predict several heat-related health outcomes in 15 regions in the province of Quebec (Canada). ML models trained with weather, regional and temporal predictors from multiple locations allowed to predict a health outcome in each of the regions simultaneously, without the need to (re) fit the same model to different locations (e.g., Ogata et al. 2021, Zhang et al. 2014, Boudreault et al., 2024a) or being limited to a single-location study (e.g., Boudreault et al. 2023, Lee et al. 2022, Navares et al. 2018). The models' calibration (initial and final hyperparameters, packages and software used), validation (overall, by regions and during heatwaves) and explainability (with feature importance) was transparent and comprehensive, which contrasts with the existing literature in population health (Morgenstern et al., 2020) and previous studies (e.g., Lee et al. 2022, Navares et al. 2018, Park et al. 2020, Wang et al. 2019, Zhang et al. 2014). To our knowledge, this is the first study to consider such a wide range of ML approaches, including deep learning, to model 9 health outcomes in several locations simultaneously.

In all the considered models, morbidity variables in both populations were well modelled with out-of-sample R^2 values of >60 %, while mortality and ambulance transports in 65+ population were less accurately predicted with R^2 values of <30 % (recall Table 3). This may be explained by different reasons, such as (1) the use of all-cause indicators instead of heat-related illnesses, (2) the presence of more noise/ randomness in these health outcomes compared to the other ones, (3) a change of relationship in the test vs. the training period, and (4) a lack of predictors to correctly capture the relationship. Unfortunately, the focus on modelling a single health outcome in the literature limited the possibility to compare our results on different health outcomes with other studies. In terms of modelling approaches, deep neural networks (MLP and LSTM) were overall the best-performing approaches across the 9 health outcomes modelled, followed by ensemble tree-based approaches (RF and LGBM) when looking at the overall summer period. For some health outcomes, increases in out-of-sample R^2 of up to 20 % were noted for neural networks compared to Lasso. Lasso was never the best modelling approach in our study.

In the literature, a few other studies compared the results of various ML models to predict heat-related health outcomes. For example, Lee et al. (2022) found that penalized regression and tree-based methods outperformed LSTM for all-cause and cause-specific mortality predictions in Seoul, South Korea. Boudreault et al. (2023) reported that ensemble tree-based models had better performance than neural networks and statistical models for mortality prediction during summer in Montreal, Canada. LSTM was found to be more efficient than RF and non-linear equations to model heat-related illnesses in the study of Nishimura et al. (2021) in Nagoya, Japan. Navares et al. (2018) found the best performance with the Auto-Regressive Integrated Moving Average (ARIMA) model and neural networks compared to tree-based methods for hospital admissions prediction in Madrid, Spain. These comparisons with international studies should be made with caution given the differences in study population, healthcare management, models used and studied health outcome. In addition, all the above studies were conducted with data of only one region/city and for a single health outcome. Indeed, other past multi-region studies mainly employed a single tree-based approach (e.g., Ke et al. 2023, Park et al. 2020, Wang et al. 2019) instead of various machine and deep learning models as in our study, an important novelty. Thus, this can explain the differences in performance observed in other studies and why deep learning was found to be superior in ours.

In addition to overall performance, further validations by region and during heatwaves presented a more complete picture of models' performance. First, the performance by regions showed that some regions were considerably better modelled by neural networks models compared to Lasso, while some regions remained poorly modelled by all approaches (recall Figs. 2 and S2). The 2018 heatwave showed sharp increases in mortality and ambulance transports that were well predicted by LGBM (for mortality), LGBM/RF (elderly mortality) and LGBM/MLP (ambulance transports) (recall Figs. 3 and S3). These results were consistent during the 2021 heatwave, although COVID-19 had a strong influence on the reported numbers (e.g., mortality displacement, lower use of health services) and that heatwave was not as hot and long as 2018's (recall Fig. S1). These additional validations, that have not been computed nor presented in previous studies, have highlighted some key information that overall performance metrics cannot reveal. They thus opened the door to future improvements to these multi-region models. For example, weighting observations when training models could ensure more accurate predictions in the most densely populated regions, or in those with the weakest performance. Furthermore, as the best model may vary depending on how performance is evaluated as seen above and in a previous study (Boudreault et al., 2023), stacked models could be considered in the future to "pool" the strengths of different approaches and result in a model likely to work better in various situations (e.g., Lu and Qiu 2023, Navares et al. 2018).

To supplement these validations, some models' explainability was also provided, limiting the "black-box" feeling of the ML methodology when this information is not provided (e.g., Lee et al. 2022, Navares et al. 2018). Feature importance (FI) showed that regional and temporal predictors were more important than weather variables for all health variables, underlining the presence of strong regional and temporal patterns in the modelled health variables (recall Tables 4 and S4). This result was expected given that the raw daily counts of the different health outcomes (adjusted by population) were modelled, which is commonly done in the literature (e.g., Lee et al. 2022, Navares et al. 2018, Ogata et al. 2021, Park et al. 2020, Wang et al. 2019). Modelling raw daily data is an advantage as it limits the data preparation steps such as trends/patterns removal required in other studies (e.g., Boudreault et al. 2023, 2024a, Masselot et al. 2018, Zhang et al. 2014). The lower importance found for weather variables can also be explained by the inclusion of highly correlated temperature variables in the models. This may have prevented the detection of these variables with permutation-based FI metrics. For example, if *tmean* is no longer available to the model, but *tmax* and *tmin* are, the model can still perform relatively well, leading to a lower FI value for *tmean* (the same also occurs for the other temperature variables, following the same logic). However, considering all these temperature variables (even if highly correlated) was important as they can contribute to a better understanding of weather patterns and, consequently, improve the predictive power of heat-related health outcomes by these models.

Regarding weather, some key meteorological information for public health planning and surveillance was revealed. First, mortality and ambulance transports in the general population were the most sensitive to short-term heat exposure (recall Tables 4 and S4) and heatwaves (recall Figs. 3 and S3), supporting their use as early indicators of heat effects. Second, hospitalizations, EDV and 811 calls were found to be associated with more delayed temperature effects. Finally, temperature changes could potentially be more important than direct temperature exposure in the elderly population. These divergent results highlight the need for differentiated analyses by population strata, which can lead to better-targeted heat-related prevention actions. These potentially new insights were made possible because of the introduced ML models combining weather information from multiple regions and by the large number of health outcomes studied. They could help health authorities to better plan the health burden during heatwaves and deliver appropriate messages to different vulnerable populations. However, explaining the pathways by which these weather variables may be linked to health effects is beyond the scope of the current study and will need to be confirmed by further research.

As the proposed models were trained with readily available predictors, they can be used to monitor health effects of heat over different time horizons in future applications, thus acting towards more resilient, healthier and sustainable cities. On the one hand, used in combination with weather forecasts, they can predict the short-term effects of heat on various health indicators and limit the heat effect on population through prevention. On the other hand, by leveraging climate and socioeconomic projections, these models can assess the long-term heat-related health burden in the coming decades, helping authorities to better plan climate change adaptation measures to reduce the heat burden. With these new tools in hand, public health authorities and decision-makers will have a better understanding of complex interactions between climate and health and a clearer portrait of the heat impacts in multiple locations and for several health outcomes. Ultimately, adopting these approaches will increase the resiliency of the society considering climate change ahead.

The strengths of the study must be highlighted. First, five ML models were considered and compared, including penalized regression (Lasso), ensemble tree-based methods (RF and LGBM) and deep learning (MLP and LSTM). Second, these models were trained using a comprehensive database of meteorological (e.g., temperature exposure and changes, humidity), regional (e.g., demographic, socioeconomic, built environment, climatology) and temporal (e.g., day, month and year) predictors. Third, these models allowed for health outcomes predictions simultaneously in multiple locations without the need to create one model per location, as it would be the case for traditional statistical/ML models. Fourth, 9 health outcomes, including both mortality and morbidity (hospitalizations, EDV, ambulance transports and 811 calls) in two populations (general and the elderly), were modelled, providing a more complete picture of the heat effects than in other studies. Finally, our study was thoroughly transparent in terms of ML models' calibration, validation and explainability, which contrasted with existing literature.

Some drawbacks must also be noted. First, only all-cause health variables were modelled as they were more readily available information. This could explain the low performance obtained for some health outcomes. Studying cause-specific mortality or morbidity, such as heat-related illnesses that may be more closely related to temperature variables, could lead to better performance and is left for future research. Second, the models were developed using data from all southernmost regions of Quebec, Canada. Developing models for each region individually could also be of interest to compare with the results obtained here, but was deemed out-of-scope for the current study. Third, the validation of the models was performed on out-of-sample years of data, but not on out-of-sample regions. This latter validation was less relevant in our study, since the developed models were intended to be used only in Quebec's regions. Fourth, not all predictors could be included in our models (such as air pollution due to data unavailability), which can explain the weaker performance obtained for some health outcomes. Finally, other models' explainability techniques, in addition to predictions during heatwaves and feature importance already presented, could have been explored to further explain ML models, such as the predicted responses for different temperature values (e.g., 10°C, 20°C or 30°C) or SHAP values.

5. Conclusion

Deep learning models (MLP and LSTM) were the most promising tools to predict several heat-related health impacts in multiple regions of Quebec (Canada) during summer, with out-of-sample R^2 of >60 % for most of the health outcomes studied. However, tree-based methods (RF and LGBM) also performed best for some health outcomes, and LGBM was found to be more accurate during heatwaves. These new tools based on machine and deep learning can be used by decision-makers to better predict the heat burden during summer, plan resources and actions appropriately, and limit the consequences of increasing heat episodes in a climate change context. They therefore contribute directly to making cities more resilient and, more broadly, to create a more sustainable and healthier society. Finally, this general framework can also be used in other regions of the world experiencing the negative effects of heat, as well as extended to study the joint effect of multiple exposures such as extreme heat and air pollution.

Data availability

Authors do not have permission to share health data. Socio-environmental data are freely available from the following organization: NASA for weather data, Ministère de la Santé et des Services Sociaux du Québec (MSSS) for demographic data and Institut national de la santé publique du Québec (INSPQ) for socio-economic and built environment data.

CRediT authorship contribution statement

Jérémie Boudreault: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Annabel Ruf:** Writing – review & editing, Visualization, Software, Funding acquisition, Formal analysis. **Céline Campagna:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Fateh Chebana:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

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

The authors would like to thank Denis Hamel and Louis Rochette from the *Bureau d'information et d'étude de la santé de la population* (BIESP) of INSPQ for the help with the health data extraction, Antoine Saint-Amand, Mathieu Tandonnet and Nathalie Gravel from the Geomatics team at INSPQ for their help with the geospatial data extraction, as well as Yohann Chiu, Magalie Canuel, Ray Bustinza, Felix Lamothe and Ernest Lo, also from INSPQ, for their comments on early versions of this work. The authors would also like to thank the editor, the associate editor, the guest editor and two anonymous reviewers for their constructive comments that helped improve this manuscript.

Funding

The first author would like to acknowledge funding from the Natural Sciences and Engineering Research Council of Canada (Vanier Scholarship #CGV-180821), the Canadian Institute of Health Research (Health System Impact Fellowship #IF1-184093), Ouranos (Real-Décoste Excellence Scholarship #RDX-317725) and the National Institute of Public Health of Quebec (no grant number). Annabel Ruf received funding from the German Academic Exchange Service (DAAD) through the Mitacs Globalink internship program (#57679026).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.scs.2024.105785.

References

Adresses Québec. (2023). AQgéobâti. https://adressesquebec.gouv.qc.ca/aqgeobati.asp.

Basu, R. (2009). High ambient temperature and mortality: A review of epidemiologic studies from 2001 to 2008. *Environmental Health, 8*(1), 1–13.

Benmarhnia, T., Deguen, S., Kaufman, J. S., & Smargiassi, A. (2015). Vulnerability to heat-related mortality. *Epidemiology, 26*(6), 781–793.

Bi, Q., Goodman, K. E., Kaminsky, J., & Lessler, J. (2019). What is machine learning? A primer for the epidemiologist. *American Journal of Epidemiology, 188*(12), 2222–2239.

Bi, X., Wu, C., Wang, Y., Li, J., Wang, C., Hahs, A., Mavoa, S., Song, C., Konrad, C., & Emch, M. (2023). Changes in the associations between heatwaves and human mortality during two extreme hot summers in Shanghai, China. *Sustainable Cities and Society, 95*, Article 104581.

Boudreault, J., Campagna, C., & Chebana, F. (2023). Machine and deep learning for modelling heat-health relationships. *Science of the Total Environment, 892*, Article 164660.

Boudreault, J., Campagna, C., & Chebana, F. (2024a). Revisiting the importance of temperature, weather, and air pollution variables in heat-mortality relationships with machine learning. *Environmental Science and Pollution Research, 31*(9), 14059–14070.

Boudreault, J., Lavigne, É., Campagna, C., & Chebana, F. (2024b). Estimating the heat-related mortality and morbidity burden in the province of Quebec, Canada. *Environmental Research, 257*, Article 119347.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

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.

CERFO. (2022). *Cartographie des îlots de chaleur et de fraîcheur dans le Québec urbain à l'aide d'imagerie satellitaire Landsat-8 (2013-2014)*. Centre d'enseignement et de recherche en foresterie de Sainte-Foy. https://www.donneesquebec.ca/recherche/dataset/ilots-de-chaleur-fraicheur-urbains-et-ecarts-de-temperature-relatifs-2013-2014/resource/a33969ba-143a-4524-88c3-8ec7485676b1.

Cheng, J., Xu, Z., Bambrick, H., Su, H., Tong, S., & Hu, W. (2019). Impacts of exposure to ambient temperature on burden of disease: A systematic review of epidemiological evidence. *International Journal of Biometeorology, 63*, 1099–1115.

Curtis, S., Fair, A., Wistow, J., Val, D. V., & Oven, K. (2017). Impact of extreme weather events and climate change for health and social care systems. *Environmental Health, 16*(1), 23–32.

Davis, R. E., McGregor, G. R., & Enfield, K. B. (2016). Humidity: A review and primer on atmospheric moisture and human health. *Environmental Research, 144*, 106–116.

Ebi, K. L., & Schmier, J. K. (2005). A stitch in time: Improving public health early warning systems for extreme weather events. *Epidemiologic Reviews, 27*(1), 115–121.

Fisher, S., & Rosella, L. C. (2022). Priorities for successful use of artificial intelligence by public health organizations: A literature review. *BMC Public Health, 22*(1), 2146.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 1189–1232.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Gosling, S. N., Lowe, J. A., McGregor, G. R., Pelling, M., & Malamud, B. D. (2009). Associations between elevated atmospheric temperature and human mortality: A critical review of the literature. *Climatic Change, 92*(3), 299–341.

Gulli, A., & Pal, S. (2017). *Deep learning with Keras*. Packt Publishing Ltd.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.

Huang, C., Barnett, A. G., Wang, X., Vaneckova, P., FitzGerald, G., & Tong, S. (2011). Projecting future heat-related mortality under climate change scenarios: A systematic review. *Environmental Health Perspectives, 119*(12), 1681–1690.

Huang, C., Barnett, A. G., Xu, Z., Chu, C., Wang, X., Turner, L. R., & Tong, S. (2013). Managing the health effects of temperature in response to climate change: Challenges ahead. *Environmental Health Perspectives, 121*(4), 415–419.

INSPQ. (2024). *Index of material and social deprivation compiled by the Bureau d'information et d'études en santé des populations (BIESP) from 1991, 1996, 2001, 2006, 2011, 2016 and 2021 Canadian Census data*. Institut national de santé publique du Québec. https://www.inspq.qc.ca/en/deprivation/material-and-social-deprivation-index.

Ke, D., Takahashi, K., Takakura, J., Takara, K., & Kamranzad, B. (2023). Effects of heatwave features on machine-learning-based heat-related ambulance calls prediction models in Japan. *Science of the Total Environment, 873*, Article 162283.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30.

Kim, Y., & Kim, Y. (2022). Explainable heat-related mortality with random forest and SHapley Additive exPlanations (SHAP) models. *Sustainable Cities and Society, 79*, Article 103677.

Kotharkar, R., Dongarsane, P., Ghosh, A., & Kotharkar, V. (2024). Numerical analysis of extreme heat in Nagpur city using heat stress indices, all-cause mortality and local climate zone classification. *Sustainable Cities and Society, 101*, Article 105099.

Kotharkar, R., & Ghosh, A. (2022). Progress in extreme heat management and warning systems: A systematic review of heat-health action plans (1995-2020). *Sustainable Cities and Society, 76*, Article 103487.

Lamothe, F., Dubé, M., & Bustinza, R. (2023). *Surveillance des impacts sanitaires des vagues de chaleur extrême au Québec – Bilan de la saison estivale 2021*. Institut national de santé publique du Québec.

Lebel, G., Dubé, M., & Bustinza, R. (2019). Surveillance des impacts des vagues de chaleur extrême sur la santé au Québec à l'été 2018. *Bulletin d'information en santé environnementale, 1*, 1–10.

Lee, J. Y., Kim, H., Gasparrini, A., Armstrong, B., Bell, M. L., Sera, F., Lavigne, E., Abrutzky, R., Tong, S., & Coelho, M. D. S. Z. S. (2019). Predicted temperature-increase-induced global health burden and its regional variability. *Environment International, 131*, Article 105027.

Lee, W., Lim, Y. H., Ha, E., Kim, Y., & Lee, W. K. (2022). Forecasting of non-accidental, cardiovascular, and respiratory mortality with environmental exposures adopting machine learning approaches. *Environmental Science and Pollution Research, 29*(58), 88318–88329.

Li, F., Yigitcanlar, T., Nepal, M., Nguyen, K., & Dur, F. (2023). Machine learning and remote sensing integration for leveraging urban sustainability: A review and framework. *Sustainable Cities and Society, 96*, Article 104653.

Li, M., Gu, S., Bi, P., Yang, J., & Liu, Q. (2015). Heat waves and morbidity: Current knowledge and further direction-a comprehensive literature review. *International Journal of Environmental Research and Public Health, 12*(5), 5256–5283.

Lin, X., Tian, T., Shi, C., Wang, P., Chen, S., Guo, T., Li, Z., Liang, B., Zhang, W., & Qin, P. (2023). What are the individual and joint impacts of key meteorological factors on the risk of unintentional injuries? A case-crossover study of over 147,800 cases from a sentinel-based surveillance system. *Sustainable Cities and Society, 91*, Article 104413.

Lu, X., & Qiu, H. (2023). Explainable prediction of daily hospitalizations for cerebrovascular disease using stacked ensemble learning. *BMC Medical Informatics and Decision Making, 23*(1), 1–13.

Masselot, P., Chebana, F., Bélanger, D., St-Hilaire, A., Abdous, B., Gosselin, P., & Ouarda, T. B. (2018). Aggregating the response in time series regression models, applied to weather-related cardiovascular mortality. *Science of the Total Environment, 628*, 217–225.

May, R. M., Goebbert, K. H., Thielen, J. E., Leeman, J. R., Camron, M. D., Bruick, Z., Bruning, E. C., Manser, R. P., Arms, S. C., & Marsh, P. T. (2022). MetPy: A meteorological Python library for data analysis and visualization. In *Bulletin of the American Meteorological Society, 103* pp. E2273–E2284).

McGregor, G. R., Bessmoulin, P., Ebi, K., & Menne, B. (2015). *Heatwaves and health: Guidance on warning-system development*. World Meteorological Organization and World Health Organization.

Meehl, G. A., & Tebaldi, C. (2004). More intense, more frequent, and longer lasting heat waves in the 21st century. *Science, 305*(5686), 994–997.

Morgenstern, J. D., Buajitti, E., O'Neill, M., Piggott, T., Goel, V., Fridman, D., Kornas, K., & Rosella, L. C. (2020). Predicting population health with machine learning: A scoping review. *BMJ Open, 10*(10), Article e037860.

MSSS. (2022). *Estimations et projections de population par territoire sociosanitaire*. Ministère de la santé et des services sociaux du Québec. https://publications.msss.gouv.qc.ca/msss/document-001617/.

Navares, R., Díaz, J., Linares, C., & Aznarte, J. L. (2018). Comparing ARIMA and computational intelligence methods to forecast daily hospital admissions due to circulatory and respiratory causes in Madrid. *Stochastic Environmental Research and Risk Assessment, 32*, 2849–2859.

Nishimura, T., Rashed, E. A., Kodera, S., Shirakami, H., Kawaguchi, R., Watanabe, K., Nemoto, M., & Hirata, A. (2021). Social implementation and intervention with estimated morbidity of heat-related illnesses from weather data: A case study from Nagoya City, Japan. *Sustainable Cities and Society, 74*, Article 103203.

Ogata, S., Takegami, M., Ozaki, T., Nakashima, T., Onozuka, D., Murata, S., Nakaoku, Y., Suzuki, K., Hagihara, A., & Noguchi, T. (2021). Heatstroke predictions by machine learning, weather information, and an all-population registry for 12-hour heatstroke alerts. *Nature Communications, 12*(1), 1–11.

Park, M., Jung, D., Lee, S., & Park, S. (2020). Heatwave damage prediction using random forest model in Korea. *Applied Sciences, 10*(22), 8237.

Pascal, M., Goria, S., Wagner, V., Sabastia, M., Guillet, A., Cordeau, E., Mauclair, C., & Host, S. (2021). Greening is a promising but likely insufficient adaptation strategy to limit the health impacts of extreme heat. *Environment International, 151*, Article 106441.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research, 12*, 2825–2830.

Rasheed, K., Qayyum, A., Ghaly, M., Al-Fuqaha, A., Razi, A., & Qadir, J. (2022). Explainable, trustworthy, and ethical machine learning for healthcare: A survey. *Computers in Biology and Medicine*, Article 106043.

Seong, K., Jiao, J., Mandalapu, A., & Niyogi, D. (2024). Spatio-temporal patterns of heat index and heat-related Emergency Medical Services (EMS). *Sustainable Cities and Society, 111*, Article 105562.

Shaamala, A., Yigitcanlar, T., Nili, A., & Nyandega, D. (2024). Algorithmic green infrastructure optimisation: Review of artificial intelligence driven approaches for tackling climate change. *Sustainable Cities and Society*, Article 105182.

Son, J. Y., Liu, J. C., & Bell, M. L. (2019). Temperature-related mortality: A systematic review and investigation of effect modifiers. *Environmental Research Letters, 14*(7), Article 073004.

Thornton, M.M., Shrestha, R., Wei, Y., Thornton, P.E., Kao, S., & Wilson, B.E. (2022). Daymet: Daily surface weather data on a 1-km grid for North America, Version 4. ORNL DAAC, Oak Ridge, Tennessee, USA.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology, 58*(1), 267–288.

Toutant, S., Gosselin, P., Bélanger, D., Bustinza, R., & Rivest, S. (2011). An open source web application for the surveillance and prevention of the impacts on public health of extreme meteorological events: The SUPREME system. *International Journal of Health Geographics, 10*(1), 1–11.

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.

Watts, N., Amann, M., Arnell, N., Ayeb-Karlsson, S., Beagley, J., Belesova, K., Boykoff, M., Byass, P., Cai, W., & Campbell-Lendrum, D. (2021). The 2020 report of the Lancet Countdown on health and climate change: Responding to converging crises. *The Lancet, 397*(10269), 129–170.

Wiemken, T. L., & Kelley, R. R. (2019). Machine learning in epidemiology and health outcomes research. *Annual Review of Public Health, 41*, 21–36.
Ye, X., Wolff, R., Yu, W., Vaneckova, P., Pan, X., & Tong, S. (2012). Ambient temperature and morbidity: A review of epidemiological evidence. *Environmental Health Perspectives, 120*(1), 19–28.
Zhang, K., Li, Y., & Schwartz, J. D. (2014). What weather variables are important in predicting heat-related mortality? A new application of statistical learning methods. *Environmental Research, 132*, 350–359.