PLOS CLIMATE

RESEARCH ARTICLE

Use of machine learning tools to predict health risks from climate-sensitive extreme weather events: A scoping review

Shakirah N. Ssebyala[1]*, Timothy M. Kintu[2], David J. Muganzi[2], Caleb Dresser[3], Michelle R. Demetres[4], Yuan Lai[5], Kobusingye Mercy[2], Chenyu Li[6], Fei Wang[7,8], Soko Setoguchi[9], Leo Anthony Celi[10,11,12], Arnab K. Ghosh[1,13]*

1 Weill Cornell Medical College, New York, New York, United States of America, **2** Department of Medicine, Mbarara University of Science and Technology, Mbarara, Uganda, **3** Department of Emergency Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts, United States of America, **4** Samuel J. Wood Library & C.V. Starr Biomedical Information Center, Weill Cornell Medicine, New York, New York, United States of America, **5** Department of Urban Planning, Tsinghua University, Beijing, China, **6** Department of Biomedical Informatics, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, United States of America, **7** Department of Population Health Sciences, Weill Cornell Medicine, New York City, New York, United States of America, **8** Institute of AI for Digital Health, Weill Cornell Medicine, New York, New York, United States of America, **9** Department of Medicine, Rutgers Robert Wood Johnson Medical School, Rutgers University, New Brunswick, New Jersey, United States of America, **10** Department of Pulmonary, Critical Care and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts, United States of America, **11** Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States of America, **12** Laboratory for Computational Physiology, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, **13** Department of Medicine, Weill Cornell Medicine, New York, New York, United States of America

* sns4002@med.cornell.edu (SNS); akg9010@med.cornell.edu (AKG)

Check for updates

OPEN ACCESS

Citation: Ssebyala SN, Kintu TM, Muganzi DJ, Dresser C, Demetres MR, Lai Y, et al. (2024) Use of machine learning tools to predict health risks from climate-sensitive extreme weather events: A scoping review. PLOS Clim 3(1): e0000338. https://doi.org/10.1371/journal.pclm.0000338

Editor: Olivier Damette, Universite de Lorraine and Climate Economic Chair Paris Dauphine, FRANCE

Received: August 9, 2023

Accepted: December 11, 2023

Published: January 17, 2024

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pclm.0000338

Copyright: © 2024 Ssebyala et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: This is a scoping review and all data used was publicly available. Details are noted in the supplemental information.

Abstract

Machine learning (ML) algorithms may play a role in predicting the adverse health impacts of climate-sensitive extreme weather events because accurate prediction of such effects can guide proactive clinical and policy decisions. To systematically review the literature that describe ML algorithms that predict health outcomes from climate-sensitive extreme weather events. A comprehensive literature search was performed in the following databases from inception–October 2022: Ovid MEDLINE, Ovid EMBASE, The Cochrane Library, Web of Science, bioRxiv, medRxiv, Institute of Electrical and Electronic Engineers, Google Scholar, and Engineering Village. The retrieved studies were then screened for eligibility against predefined inclusion/exclusion criteria. The studies were then qualitatively synthesized based on the type of extreme weather event. Gaps in the literature were identified based on this synthesis. Of the 6096 records screened, seven studies met the inclusion criteria. Six of the studies predicted health outcomes from heat waves, and one for flooding. Health outcomes described included 1) all-cause non-age standardized mortality rates, 2) heat-related conditions and 3) post-traumatic stress disorder. Prediction models were developed using six validated ML techniques including non-linear exponential regression, logistic regression, spatiotemporal Integrated Laplace Approximation (INLA), random forest and decision tree methods (DT), and support vector machines (SVM). Use of ML algorithms to assess adverse health impacts from climate-sensitive extreme weather events is possible. However, to fully utilize these ML techniques, better quality data suitable for use is desirable.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Development of data standards for climate change and health may help ensure model robustness and comparison across space and time. Future research should also consider health equity implications.

Introduction

Climate change represents the greatest public health challenge of our time. In the United States, seven in ten Americans experienced extreme weather events in 12022 [1]. While there has been increasing acknowledgement in government [2, 3] and health-related organizations [4, 5] of the impending threat that climate change and associated extreme weather events pose to at-risk populations, there has been a focus on mitigating its severity rather than adapting to anticipated impacts. However, adapting to future climate impacts will be increasingly important [6]. Extreme weather events are likely to increase in frequency and destructiveness [7]; the warming climate is leading to escalating risks related to heat waves, droughts, wildfires, severe weather, tropical cyclones, extreme rainfall, and flooding.

Machine learning (ML) algorithms have the potential to play an important role in predicting individual and population-level health impacts of climate-sensitive extreme weather. By leveraging vast amounts of clinical, socioeconomic, and environmental data, ML-guided tools can be harnessed to identify and quantify the risk to individuals and populations from specific threats [8, 9], similar to their use in predicting breast cancer survival [10], occurrence of coronary artery disease [11], and in classifying and detecting cancerous lesions on images [12, 13]. Additionally, ML-guided tools have the potential to build on existing systems that assess populations at risk from extreme weather events and take advantage of secondary data to do so. By leveraging healthcare system data (e.g., electronic medical records), healthcare organizations and public health authorities may have the opportunity to tailor algorithms to specific populations, making them more context- and disease-specific, and perhaps addressing limitations of existing emergency preparedness systems.

It is unclear to what extent ML techniques are being used to predict health outcomes from extreme weather events, and to what extent they may be useful for doing so. In this scoping review, we aimed to comprehensively describe peer-reviewed manuscripts in the published and grey literature reporting use of ML methods to predict health outcomes from climate-sensitive extreme weather events worldwide.

Methods

We performed a scoping review to examine the use of ML techniques to predict health outcomes during and after extreme weather events. This study was performed following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for scoping reviews (PRISMA-ScR) [14]. In accordance with these standards, a protocol was submitted and preregistered by the International Prospective Register of Systematic Reviews (PROSPERO; CRD42023391186). The PRISMA flow diagram is described in the Fig 1.

Search strategy

A medical librarian (MRD) performed comprehensive searches to identify studies that addressed ML approaches to predicting at-risk populations or health impacts from climate-sensitive extreme weather events. Searches were finalized October 20, 2022, in the following databases: Ovid MEDLINE, Ovid EMBASE, The Cochrane Library (Wiley), Web of Science

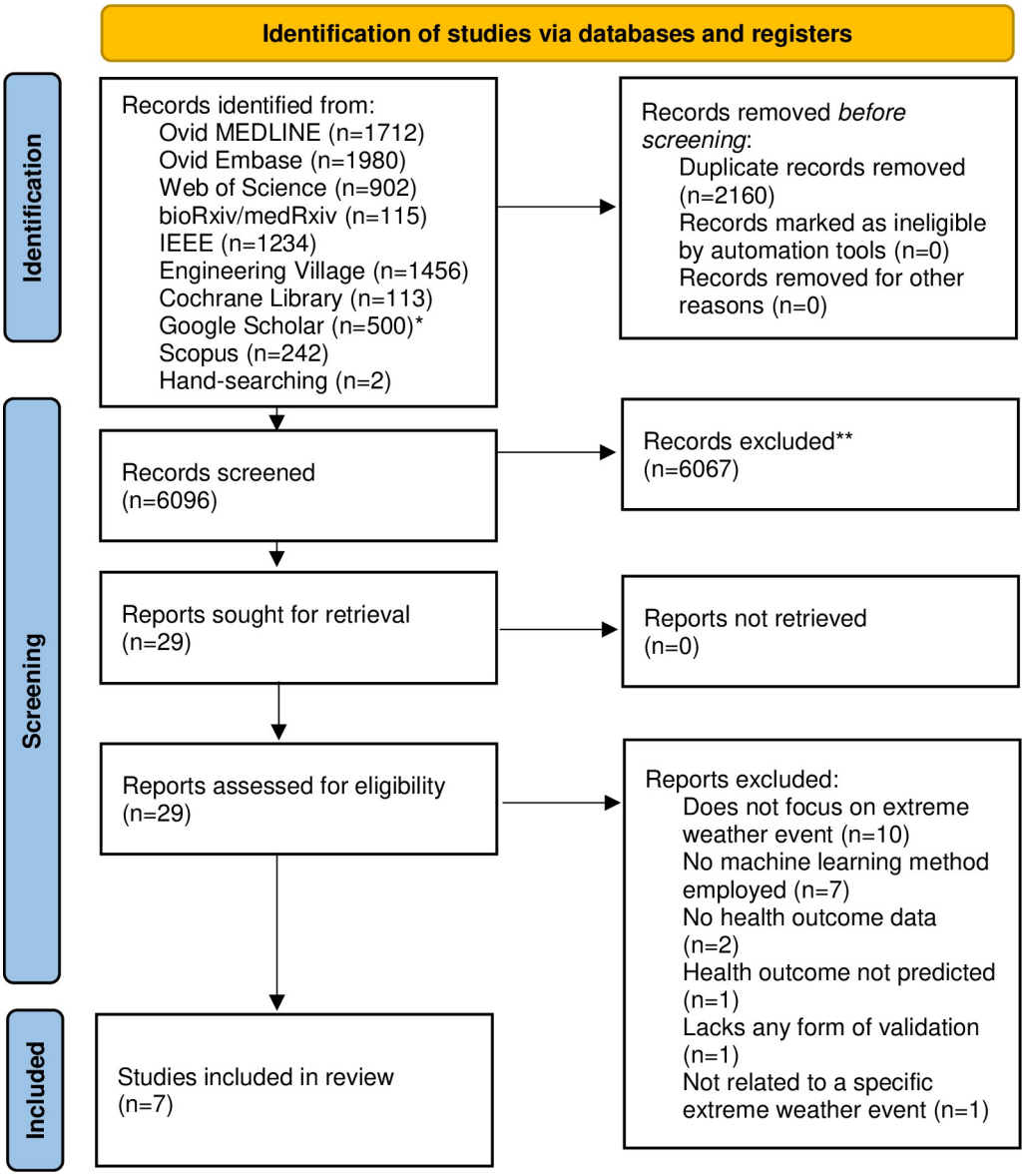

Fig 1. PRISMA 2020 flow diagram for new systematic reviews which included searches of databases and registers only. *Due to the inherently large number of results retrieved from a Google Scholar search, only the first 500 results were screened.

https://doi.org/10.1371/journal.pclm.0000338.g001

(Core Collection–Clarivate), bioRxiv, medRxiv, Institute of Electrical and Electronic Engineers (IEEE) database, Google Scholar, and Engineering Village (Elsevier). The search strategy included all appropriate controlled vocabulary and keywords for the concepts of "machine learning" and "natural disasters" or "extreme weather." The full search strategies for all databases are available in Supplement (S1 Table). To limit publication bias, there were no language, publication date, or article type restrictions on the search strategy. For articles selected for inclusion in this study, reference lists and citing articles were pulled from Scopus (Elsevier) and screened. Two additional studies were identified from hand-searching relevant journals.

Study selection

Retrieved studies were screened (AKG, MRD, SNS, SS, FW, LAC, CD, YL, DJM, KM, TMK) for inclusion using Covidence systematic review software. Titles and abstracts were reviewed against predefined inclusion/exclusion criteria by two independent reviewers. Discrepancies were resolved by consensus. For final inclusion, the full text was retrieved and then screened by two independent reviewers.

Inclusion and exclusion criteria

Our inclusion criteria were: (1) Population: participants $\geq$ 18 years old; (2) Exposure: Extreme weather and related events where the onset is acute, defined as within 2 weeks, and that are thought to be exacerbated by climate change (i.e., hurricanes /cyclones/typhoons, wet precipitation leading to flooding, wildfires, heat waves, extreme cold, mudslides); (3) Outcomes studied: health events (death, hospitalization, any presentation to clinic or emergency department) within a defined time-frame (e.g., 30 days, 2 months, 1 year, 5 years); (4) Employs ML-based modeling methods (e.g., linear regression, decision trees, k-nearest neighbors, random forests, kernel methods, deep neural networks) [15] to predict the health outcomes; (5) The ML methods undergoes evaluation by assessment and/or creation of accuracy, precision, recall, specificity, F1 score, Receiver Operator Characteristics curve, Precision Recall curve; (6) The ML models undergo internal validation using either training and testing split, resubstituting, K fold cross validation, bootstrapping, nested cross validation, or external validation using a different dataset.

Studies were excluded according to the following criteria: (1) Written in non-English language; (2) Review articles, commentaries, or editorials; and (3) Authors do not include measures for ML model validation.

Data extraction

Data extraction was performed independently by a pair of reviewers with predefined, standardized templates. Each extraction was reviewed independently by a secondary reviewer (SNS) after extraction by the primary reviewer (AKG). Data points defined for extraction were extreme weather event; study design; population; health outcome; time horizon for health outcome; ML techniques; validation methods; and key findings.

Data synthesis

Following data extraction, findings were synthesized qualitatively to describe the geographies covered and study settings, types of extreme weather events, health outcomes, ML techniques used, and validation methods employed. Gaps in the literature were then identified based on this qualitative synthesis. No quantitative assessment of the literature (i.e., meta-analysis) was performed because of the heterogeneity in ML techniques used and outcomes assessed.

Results

Summary of articles

The study selection process is outlined in the PRISMA flow diagram in Fig 1. After removal of duplicates, a total of 6096 records were screened. A total of 7 studies, summarized in Table 1, met criteria for inclusion in the analysis.

1. Geographies covered and study settings. The seven studies analyzed in here report data from the United States of America [16], Europe [17–19], China [20, 21] and South Korea [22]. Data was collected at the country [18, 19], county [16] or division [20, 22], and city level

Table 1. Summary of study results.

Author, Year	Extreme Weather Event	Study design	Population	Health outcome	Time horizon for health outcome	ML techniques	Validation Methods	Findings	Comments
Dessai et al., 2002	June 1981 heat wave, and the July 1991 heat wave in Lisbon, Portugal Heat-associated outcomes >29˚C	Secondary data analysis of weather and administrative mortality data	Population of Lisbon, Portugal (1980–1998)	All-cause mortality	2 days after heat wave	Non-linear, exponential regression with daily lags	Internal validation: split data into 2 time periods: 1980 to 1989, and 1990–1998. Trained model on one time period and tested it on the other, and qualitatively compared predicted-to-observed mortality rates using residuals from regression model	No summary statistics provided for discrimination or calibration; model trained using data from 1980s on 1991 heat wave discriminates more than model trained using data from 1990s on 1981 heat wave across no lag, 1-day lag, and 2-day lag models	1. Uses all-cause mortality rather than heat-related causes of mortality (e.g., cardiovascular disease) 2. Model specification is unclear, and no other modeling strategies employed 3. Only qualitative assessment of residuals used for validation
Huang et al., 2010	1998 flood affecting Hunan province, China	Multi-stage stratified, cluster sampling of flood-impacted individuals between January and May 2000	29,285 individuals with survey completion rate of 87% (25,478); 44.7% under 16 years old, 52.9% female, 9.7% with high school education or higher, and 91.3% were occupied as farmers; 9.2% diagnosed with post-traumatic stress disorder (PTSD)	Diagnosis of post-traumatic stress disorder (PTSD) using Diagnostic and Statistical Manual of Mental Disorders Fourth Edition (DSM-IV) definition	Approximately two years after flood events	Logistic regression (LR) (stepwise, forward); variables added based on author's opinion and prevailing literature	Internal validation: 70%/30% split into test/training datasets Receiver Operator Curves (ROC) constructed and area under ROC quoted	AUC = 0.853; calibration of model was not performed	1. Calibration of model not performed 2. Other modeling strategies not considered
Lowe et al., 2015	Heat wave (temperature not defined) August 1–15, 2003, and Cold spell January 1–15, 2003, Europe	Secondary data analysis of weather and administrative mortality data	16 countries across Europe (Austria, Belgium, Croatia, the Czech Republic, Denmark, France, Germany, Italy, Luxembourg, Netherlands, Poland, Portugal, Slovenia, Spain, Switzerland and the United Kingdom)	Daily mortality rates across 54 regions with homogenous populations from 1998–2003	Daily after the beginning of the heat wave and cold wave	Spatiotemporal Integrated Laplace Approximation (INLA)	Quantitative comparison between predicted daily mortality rates and observed rates, defined through receiver operator curve (ROC), specificity and sensitivity for pre-specific mortality thresholds (defined as 75% of mortality distribution)	Area under the ROC 97% with positive predictive value (PPV) of 89% for heat wave model; and ROC 78% with PPV 69% for cold spell model	1. Modeling for geographical regions that are large, potentially obfuscating the finer spatial and other confounders that may influence mortality rates

(*Continued*)

Table 1. (Continued)

Author, Year	Extreme Weather Event	Study design	Population	Health outcome	Time horizon for health outcome	ML techniques	Validation Methods	Findings	Comments
Lowe et al., 2016	Heat wave (temperature not defined) August 1–15, 2003 and Cold spell January 1–15, 2003, Europe	Secondary data analysis of weather and administrative mortality data using lag model with sub-seasonal-to-seasonal temperature data	16 countries across Europe (Austria, Belgium, Croatia, the Czech Republic, Denmark, France, Germany, Italy, Luxembourg, Netherlands, Poland, Portugal, Slovenia, Spain, Switzerland and the United Kingdom)	Daily mortality rates across 54 regions with homogenous populations from 1998–2003	Daily after the beginning of the heat wave and cold wave after considering different climate forecast data at different lead times (from 1 day to 3 months)	Spatiotemporal Integrated Laplace Approximation (INLA)	Quantitative comparison between predicted daily mortality rates and observed rates, defined through receiver operator curve (ROC), specificity and sensitivity for pre-specific mortality thresholds (defined as 75% of mortality distribution)	The mean area under the ROC for greater than 4 days lead time was < 0.5	Same as Lowe at al., 2015 Assess the use of forecasted temperature instead of the observed
Wang et al., 2019	Heat waves defined as days > 35 degrees C, China, 2012–2014	Secondary data analysis of weather and administrative mortality data	Residents of the cities of Shanghai, Chongqing, Wuhan, Jinan, Ningbo, Hefei and Shaoxing	Daily rate of heat stroke cases	Daily after the beginning of the heat wave	Random Forest (RF) model	Internal validation: test (10%) and train (90%); linear fitting method and linear R^2 used to assess model performance; calibration undertaken using Bland-Altman method	Maximum linear R^2 was 0.636 with 14 model variables and 500 trees; mean difference between predicted to observed pairs was not significant	1. This modeling incorporated a range of socioeconomic and other environmental factors including internet search index, urban vs rural populations, air conditioner number per hundred houses, and normalized difference in vegetation rates
Park et al., 2020	Heat waves defined as maximum daily temperature $> 33^{\circ}$C for 2 consecutive days, 2015–2018, Korea	Secondary data analysis using nationwide thermal disease monitoring from the Korea Disease Control and Prevention Agency	17 administrative divisions of Korea	Heat-related damage, defined as heatstroke, cardiovascular and respiratory diseases	1 week period after the heat wave	Random Forest (RF) compared to Decision tree (DT) methods, support vector machines (SVM), and logistic regression (LR)	Quantitative comparison of modeling strategies using mean absolute error (MAE), root mean squared error (RMSE), root mean squared logarithmic error (RMSLE), coefficient of determination ($R^{2)}$	Compared to the 3 other modeling strategies, the performance of the random forest was better across all measures: MAE– 3.81 (RF) Vs 5.5, 5.1, 5.3 (DT, SVM, LR) RMSE: 8.7 vs 13.3, 8.8, 12.5 RMSLE: 0.65 vs 0.8, 0.96, 0.86 $R^{2:}$ 0.8 vs 0.53, 0.80, 0.6	1. Outcome under study (health outcome) not well-defined. 2. Modeling incorporated several measures of vulnerability including (area-level measures of income, insurance premiums, occupational groups) as well as dynamic meteorological variables 3. Not calibrated

(*Continued*)

Table 1. (Continued)

Author, Year	Extreme Weather Event	Study design	Population	Health outcome	Time horizon for health outcome	ML techniques	Validation Methods	Findings	Comments
Jiang et al., 2021	Extreme heat events, defined, over 1993–2012 study period, as either weather measures greater than 1) 95% percentile threshold of daily maximum temperature; 2) 98th percentile threshold of daily minimum temperature; or 3) 99% percentile threshold over apparent temperature, Georgia, USA	Retrospective secondary data analysis	Patient-level emergency department visits from hospitals within the 20-county Atlanta metropolitan area (1993–2004) or from the Georgia Hospital Association (2005–2012) during summer months (May to September)	Emergency room presentations for all internal causes (ICD-9 001–799) with subgroups (diabetes, fluid and electrolyte imbalance, circulatory system diseases, hypertension, ischaemic heart disease, myocardial infarction, congestive heart failure, ischaemic stroke, renal diseases, acute renal failure, and heat-related illnesses)	Not applicable; model specification comparing across 20-year period extreme heat event days to non-extreme heat event days	Logic regression (tree-based method)	Simulation study performed for each exposure and lagged effects, sensitivity, specificity, relative bias, and relative root mean square error quoted for specific ICD-9 groupings	Across all disease states associated with emergency room presentations (i.e., circulatory system diseases, renal disease, and heat-related illnesses), logic regression models with extreme heat event-related single day exposures performed the best of all models in terms of sensitivity, specificity, smallest relative bias, and smallest relative root mean square error	1. Model specification did not include patient-level characteristics including age, sex, measure of clinical comorbidities

Abbreviations

PTSD = post-traumatic stress disorder

DSM-IV = Diagnostic and Statistical Manual of Mental Disorders Fourth Edition

ROC = Receiver Operator Curves

AUC = Area Under the Curve

ICD-9 = International Classification of Disease, 9

PPV = Positive Predictive Value

SVM = Support Vector Machine

DT = Decision Tree

RF = Random Forest

LR = Logistic Regression

MAE = mean absolute error

RMSE = root mean squared error

RMSLE = root mean squared logarithmic error

INLA = Integrated Laplace Approximation

https://doi.org/10.1371/journal.pclm.0000338.t001

[17, 21]. There were 16 European countries included, namely Austria, Belgium, Croatia, the Czech Republic, Denmark, France, Germany, Italy, Luxembourg, Netherlands, Poland, Portugal, Slovenia, Spain, Switzerland and the United Kingdom (England and Wales only) [18, 19]. The main study design used was retrospective secondary data analysis, utilized in six of the articles [16–19, 21, 22]. The remaining study from Hunan, China used a multistage, stratified cluster sampling design [20].

2. Extreme weather events studied. Six studies examined the impact of heat waves on population health either daily [16, 18, 19, 21], within 2 days [17], or 1 week after the heat wave [22]. There was variation in the definition of heat waves in different regions. Heat-associated outcomes were documented for temperature >29˚C in Lisbon, Portugal [17], >33˚C in Korea [22] and >35˚C in China [21]. The study from Georgia, USA documented extreme heat events as either a temperature measure greater than 95% percentile threshold of daily maximum temperature, or 98th percentile threshold of daily minimum temperature or 99% percentile threshold over apparent temperature [16]. Both Lowe et al. studies from Europe examined a heat wave and cold spells for which the exact temperatures were not defined [18, 19]. One study from the Hunan province in China examined the effect of a 1998 flood approximately two years later [20].

3. Health outcomes evaluated. The health outcomes described in the final studies were all-cause, non-age-standardized mortality [17–19], post-traumatic stress disorder (PTSD) diagnosed using validated instruments [20], emergency department presentations [16], and heat-related conditions including heat stroke, cardiovascular and respiratory diseases using relevant International Classification of Disease version 9 codes from a thermal disease monitoring system [21, 22].

4. Description, and comparison of machine learning techniques employed. Among the seven studies, there were six ML techniques utilized: non-linear exponential regression [17], logistic regression [16, 20], spatiotemporal Integrated Laplace Approximation (INLA) [18, 19], random forest models [21, 22] decision tree methods (DT), and support vector machines (SVM) [22]. Only one study [22] compared their proposed model to other ML strategies: random forest model to decision tree methods (DT), support vector machines (SVM) and logistic regression (LR).

The techniques examined in this review are all supervised learning whose goal is to use labelled data to predict outcomes. Of note, these techniques have varied levels of interpretability and performance. In particular, the regression methods such as non-linear exponential and logistic regression describe outputs that are more easily interpreted but may have lower performance. On the other hand, the classification models including INLA, DT, and SVM use categorical labels, have lower interpretability and higher performance compared to the regression models [15].

A common ML algorithm employed was logistic regression, an adaptive regression method that attempts to construct predictors as Boolean combinations of binary covariates. For example, Dessai et al. [17] employed non-linear exponential regression to model the aggregate, non-linear effects of climate-related mortality. Jiang et al. [16] applied logistic regression to a time-series analysis of ED visit data and extreme heat indicators. The different combinations of the binary covariates allowed the authors to generate better predictive models that capture both lagged and sustained effects of extreme heat. In their study, Huang et al. used a step-wise forward regression to select predictive risk factors for the binary dependent variable, the presence or absence of PTSD [20].

Differently, both studies authored by Lowe employed INLA, a Bayesian technique that efficiently models data structures with spatiotemporal components. They developed models to estimate mortality-apparent temperature relationships in a Bayesian model framework, which

allows the simulation of probabilistic predictions of daily mortality in space and time. These models were then fitted with the INLA to simulate the mortality predictions for the heat and cold spells [18, 19].

The third type of ML modeling was the random forest model. Random forest models are used to construct hundreds or thousands of deep decision trees by employing a bootstrapping method (a resampling technique that involves random sampling with replacement which makes the forests stronger predictors). The outputs of these trees are then combined which makes this method less prone to overfitting and multi-collinearity, allowing high predictability with robustness that improve its generalizability. Additionally, random forests usually require less parameter tuning which makes them a good method for smaller datasets [15]. Wang et al. used the Boruta algorithm, a wrapper built around the random forest algorithm that uses a z-score for feature selection, ensuring that the variables in their model was significantly correlated with the outcome variable. They added socioeconomic variables including location, gross domestic product, and population density which improved prediction and the model fitness [21]. The Park et al. group also used the Boruta method to consider a range of socioeconomic, meteorological, and demographic variables in their model. Of note, they report that regional economic indicators such as income and insurance had less impact on prediction of the diseases related to heatwaves [22].

Park et al. used other ML techniques including DT and SVM. DT are an essential building block for many ML techniques (including random forest models) as they cluster data based on classes or probabilities. However, a disadvantage of DT methods is that they are prone to overfitting. SVM methods are used for classification of continuous output, and they are robust to noise and particularly effective in high-dimensional datasets. The goal of SVM is to find the optimal decision boundary that separates the classes. Compared to DT and RF methods, the outputs of an SVM may be less in interpretable [15]. Park et al. compared their RF model to the DT and SVM to determine the accuracy in making predictions. Using the same training data set, they found that the RF model was more accurate than other models for making predictions [22].

5. Validation methods used. The ML techniques were validated using three methods including internal validation [17, 20, 21], simulation [16], and quantitative comparison between predicted and observed data [18, 19, 22]. In one study, the authors undertook internal validation of the ML method by splitting the data sets into two time periods, trained the model with one time period and tested it on another, and then compared the observed and expected values through residual analysis and regression coefficient [17]. In other studies, there was a 70/30 [20] or 90/10 [21] training/test dataset which was used to calculate area under the receiver operator curve (ROC) [20, 21].

There was slight variation in the modes of quantitative comparison between the studies. The Lowe studies also reported the area under the ROC, and positive predictor value (PPV) for the heat wave model [18, 19]. Park et al. compared all three modeling strategies using the mean absolute error (MAE), root mean squared error (RMSE), root mean squared logarithmic error (RMSLE), and coefficient of determination R^2 [22].

Qualitative synthesis

1. Heat wave- and temperature-focused papers. Six of the seven studies that were included focused on predicting health outcomes associated with heat waves [16–19, 21, 22]; two also evaluated cold spells [18, 19]. These large retrospective studies were conducted at city [17, 21], county [16], division [22], or country [18, 19] level. There was variation in the definition of heat waves, with a range of temperatures above 29–35˚C [17, 21, 22]. Most of the

studies evaluated health outcomes daily [16, 18, 19, 21], while others examined at 2 days [17] or 1 week after the heat wave [22]. These included all-cause, non-age-standardized mortality [17–19], emergency department presentations [16], or specific heat-related health conditions including heat stroke, exhaustion, cramps, fainting and edema [21, 22]. The validated ML techniques used have been described above. Two studies incorporated a range of socioeconomic data such as area-level measures of income, insurance premiums, occupational groups and other environmental factors including internet search index, urban vs rural populations, air conditioner number per hundred houses, and normalized difference in vegetation rates [21, 22].

2. Flooding paper. Huang et al. examined the effect of a 1998 flood in the Hunan province in China approximately two years after the event [20]. They collected data from a qualitative survey of 29285 individuals impacted by the flood. Seven independent predictive factors (age, gender, education, type of flood, severity of flood, flood experience, and mental status before the flood) were identified and used as key variables in a risk score model using stepwise, forward logistic regression. The area under the ROC curve for the model was 0.853 in the validation data. The sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of this risk score model were 84.0%, 72.2%, 23.4%, and 97.8%, respectively, at a cutoff value of 67.5 in the validation data.

Discussion

This scoping review identified seven articles predicting the relationship between exposure to climate-sensitive extreme weather events and health outcomes using ML tools that underwent validation. Heat waves were the most studied extreme weather event; two groups also compared cold spells [18, 19]. All but one [16] of the studies was from outside the United States.

Findings from these studies suggest that the use of ML to predict health outcomes in populations at risk of climate-amplified extreme weather events is possible, can be done in a rigorous manner, and has significant public health potential at the local level, regional, national, and international level. Nonetheless, despite ML's ever increasing role in the healthcare landscape over the last two decades [15], the relatively small number of studies identified in this review highlights the lack of realized potential in using these tools to identify and safeguard future populations from climate-sensitive extreme weather events.

ML powerfully enables complex analysis of huge datasets that can be leveraged to develop risk prediction tools tailored to specific populations [8]. Additionally, ML-focused projects have the potential to augment existing emergency preparedness systems by making them more adaptable to different contexts and specific diseases. A key question stemming from this scoping review is why, given the growing role of ML across the spectrum of healthcare in general, has ML not been more rigorously utilized in assessing risk from climate-sensitive extreme weather events, particularly in the US setting?

We posit three possible reasons. First, few countries and organizations have access to the data necessary to examine the population-level health effects of climate change. Second, there is a growing need for robust health and environmental data infrastructures including data/metadata standards in climate and health. And third, there are unclear uses for ML-formed risk prediction models in climate change.

To be useful as a public health tool, ML methods need to define populations at risk of climate-sensitive extreme weather events completely. We note that all except one of the reviewed studies was from Europe and Asia where robust national healthcare data infrastructures exist, thus allowing complex population-level analyses. In the US, national population-level healthcare is only organized for a few set populations (e.g., Medicare data for those $>$ 65 years, those on hemodialysis). Moreover, health data focused on disease surveillance and healthcare

utilization is often collected by different organizations at the federal, state, and local levels, leading to fragmentation across different geographic regions, populations, and levels of specificity. Although methods have been developed to organize healthcare data in the US, there is still work to be done. Consequently, the United States is missing a significant opportunity to harness its extensive expertise in ML and artificial intelligence to examine the impact of climate-sensitive exposures on populations, predict individuals at risk, and implement interventions to safeguard lives and livelihoods in both the short and long term.

Our review highlights the need for more robust health and environmental data infrastructures that incorporate standard definitions and measurements of climate-sensitive exposures. We note the heterogeneity in definitions of the heat waves with ranges in temperature from 29–35˚C or undefined but > 95th percentile of the daily minimum range. Other variables such as the health outcomes also varied in that some articles evaluated the daily mortality versus emergency department visits versus heat-related conditions, some of which lacked specificity. Not only does this lack of standardization pose challenges to compare and possibly fit proposed models to different places without standardized metrics for the variables, but it also limits any comparisons between geographic regions and across time. Thus, as a result, the lack of standards prevents the identification and development of unified measures and guidelines to examine the health effects of climate change broadly and build consensus towards potential interventions. This is important, as climate-amplified extreme weather events are increasing in frequency, intensity, and duration, and novel tools are critical to protecting populations from their consequences.

Last, once developed, the appropriate agency or group which might employ these ML-informed risk prediction tools is unclear. Currently, except for a few notable cases including the Department of Health and Human Services emPOWER program [23], few actors have the breadth, ability, and resources to leverage the knowledge generated from these tools to protect the well-being of at-risk populations, particularly in the US. Lack of proper infrastructure for surveillance, reporting, and integrated evidenced-based decision-making likely made the US more susceptible to impacts from such events, as compared to other high-income countries [24]. Historically, it has been the purview of emergency management services to address the devastation wrought by extreme weather events. However, as these become more frequent, there is concern that these services–typically provided only in the short-term–may not be able to provide the necessary care in a repeated fashion and may miss opportunities for prevention-focused efforts on longer timescales. As a result, countries, states, and cities may seek to reorganize their preparation for and responses to extreme weather hazards, particularly as they become more damaging in terms of health and financial cost. As these changes occur, ML prediction tools may have a role in planning public health approaches and priorities in both pre- and post-event phases.

A final notable feature of our review was the limited focus on factors related to the inequitable impacts of climate-responsive environmental hazards. Addressing health inequities in communities affected by climate-amplified extreme weather events requires nuanced socioeconomic data. ML methods are well-equipped to manage the nested, hierarchical structure of such data forms. However, only two studies incorporated socioeconomic and other environmental factors in their model to capture the complexity of their influence on health outcomes [21, 22]. It is increasingly reported that extreme weather events disproportionately affect the health of populations based on several factors [25] including advanced age [26], gender [26, 27], preexisting conditions [27–29], geographical features [28] especially cities [30], poverty [28, 31], and population density [31]. Incorporation of socioeconomic and spatial parameters in future work may be beneficial.

Limitations in this study exist. Although we performed a comprehensive bibliometric analysis of existing peer-reviewed and grey literature in English, this review included only a small number of final articles. While this limits synthesis of the data to generate strong conclusions on how ML is used as a tool to predict health outcomes, we identify an opportunity to expand use of this apparently underleveraged tool. Moreover, our methodology was robust in that our initial screening was not language specific, and re-evaluated bibliographies of the final selected manuscripts to ensure the completeness of the research.

In conclusion, this scoping review demonstrated the feasibility of using ML methods to predict the risk of adverse health outcomes from climate-responsive extreme weather events including heat waves, cold spells, and floods. Despite the comprehensive approach, this review yielded only seven articles meeting criteria for inclusion. While there is great opportunity to use ML as a tool to identify and potentially develop strategies to protect vulnerable populations from health harms resulting from extreme weather events, utilization of this tool has been limited. Future efforts may benefit from focusing on utilizing more comprehensive and higher quality data with ML tools, creating data standards for climate change and health datasets to ensure robustness of the models and comparability of results, and expanding the capacity of agencies, health professionals, and other organizations to utilize this data and translate its findings into actionable public health interventions.

Supporting information

S1 Checklist. Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist. JBI = Joanna Briggs Institute; PRISMA-ScR = Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews. * Where *sources of evidence* (see second footnote) are compiled from, such as bibliographic databases, social media platforms, and Web sites. † A more inclusive/heterogeneous term used to account for the different types of evidence or data sources (e.g., quantitative and/or qualitative research, expert opinion, and policy documents) that may be eligible in a scoping review as opposed to only studies. This is not to be confused with *information sources* (see first footnote). ‡ The frameworks by Arksey and O'Malley (6) and Levac and colleagues (7) and the JBI guidance (4, 5) refer to the process of data extraction in a scoping review as data charting. § The process of systematically examining research evidence to assess its validity, results, and relevance before using it to inform a decision. This term is used for items 12 and 19 instead of "risk of bias" (which is more applicable to systematic reviews of interventions) to include and acknowledge the various sources of evidence that may be used in a scoping review (e.g., quantitative and/or qualitative research, expert opinion, and policy document). *From*: Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMAScR): Checklist.
(DOCX)

S1 Table. Search strategies for databases.
(DOCX)

Author Contributions

Conceptualization: Shakirah N. Ssebyala, Arnab K. Ghosh.

Data curation: Shakirah N. Ssebyala, Michelle R. Demetres, Arnab K. Ghosh.

Formal analysis: Shakirah N. Ssebyala, Timothy M. Kintu, David J. Muganzi, Caleb Dresser, Michelle R. Demetres, Yuan Lai, Kobusingye Mercy, Chenyu Li, Arnab K. Ghosh.

Funding acquisition: Arnab K. Ghosh.

Investigation: Shakirah N. Ssebyala, Arnab K. Ghosh.

Methodology: Shakirah N. Ssebyala, Michelle R. Demetres, Arnab K. Ghosh.

Project administration: Shakirah N. Ssebyala, Arnab K. Ghosh.

Resources: Shakirah N. Ssebyala, Arnab K. Ghosh.

Software: Michelle R. Demetres, Arnab K. Ghosh.

Supervision: Shakirah N. Ssebyala, Arnab K. Ghosh.

Validation: Shakirah N. Ssebyala, Arnab K. Ghosh.

Visualization: Shakirah N. Ssebyala, Arnab K. Ghosh.

Writing – original draft: Shakirah N. Ssebyala, Arnab K. Ghosh.

Writing – review & editing: Shakirah N. Ssebyala, Timothy M. Kintu, David J. Muganzi, Caleb Dresser, Michelle R. Demetres, Yuan Lai, Kobusingye Mercy, Chenyu Li, Fei Wang, Soko Setoguchi, Leo Anthony Celi, Arnab K. Ghosh.

References

1. Kennedy B. Americans Divided Over Direction of Biden's Climate Change Policies https://www.pewresearch.org/topic/science/: Pew Research Center; 2022 [updated July 14; cited 2022 July 14. 42]. Available from: https://www.pewresearch.org/science/2022/07/14/americans-divided-over-direction-of-bidens-climate-change-policies/.
2. US Department of Health & Human Services. The Office of Climate Change and Health Equity. Available from: https://www.hhs.gov/ocche/index.html.
3. Executive Order on Protecting Public Health and the Environment and Restoring Science to Tackle the Climate Crisis [press release]. The White House, January 20, 2021.
4. Centers for Disease Control and Prevention. Climate Effects on Health Atlanta: CDC; 2021. Available from: https://www.cdc.gov/climateandhealth/effects/default.htm.
5. National Academy of Medicine. Climate Change, Human Health, and Equity 2021. Available from: https://nam.edu/programs/climate-change-and-human-health/.
6. Masson-Delmotte V P. Zhai A, Pirani SL, Connors C, Péan S, Berger N, et al. Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC; 2021.
7. National Centers for Environmental Information: U.S. Billion-Dollar Weather and Climate Disasters 2020. Available from: https://www.ncdc.noaa.gov/billions/, https://doi.org/10.25921/stkw-7w73
8. Bayati M, Braverman M, Gillam M, Mack KM, Ruiz G, Smith MS, et al. Data-driven decisions for reducing readmissions for heart failure: general methodology and case study. PLoS One. 2014; 9(10): e109264. https://doi.org/10.1371/journal.pone.0109264 PMID: 25295524
9. Wiens J, Shenoy ES. Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Epidemiology. Clin Infect Dis. 2018; 66(1):149–53. https://doi.org/10.1093/cid/cix731 PMID: 29020316
10. Burke HB, Goodman PH, Rosen DB, Henson DE, Weinstein JN, Harrell FE Jr, et al. Artificial neural networks improve the accuracy of cancer survival prediction. Cancer. 1997; 79(4):857–62. https://doi.org/10.1002/(sici)1097-0142(19970215)79:4<857::aid-cncr24>3.0.co;2-y PMID: 9024725
11. Lapuerta P, Azen SP, LaBree L. Use of neural networks in predicting the risk of coronary artery disease. Comput Biomed Res. 1995; 28(1):38–52. https://doi.org/10.1006/cbmr.1995.1004 PMID: 7614823
12. Vial A, Stirling D, Field M, Ros M, Ritz C, Carolan M, et al. The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: a review. Transl Cancer Res. 2018; 7(3):803–16.
13. Fakoor R, Ladhak F, Nazi A, Huber M, editors. Using deep learning to enhance cancer diagnosis and classification. Proceedings of the international conference on machine learning; 2013: ACM New York, NY, USA.
14. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med. 2018; 169(7):467–73. https://doi.org/10.7326/M18-0850 PMID: 30178033

15. Badillo S, Banfai B, Birzele F, Davydov II, Hutchinson L, Kam-Thong T, et al. An Introduction to Machine Learning. Clin Pharmacol Ther. 2020; 107(4):871–85. https://doi.org/10.1002/cpt.1796 PMID: 32128792

16. Jiang S, Warren JL, Scovronick N, Moss SE, Darrow LA, Strickland MJ, et al. Using logic regression to characterize extreme heat exposures and their health associations: a time-series study of emergency department visits in Atlanta. BMC Med Res Methodol. 2021; 21(1):87. https://doi.org/10.1186/s12874-021-01278-x PMID: 33902463

17. Dessai S. Heat stress and mortality in Lisbon part I. model construction and validation. Int J Biometeorol. 2002; 47(1):6–12.

18. Lowe R, Ballester J, Creswick J, Robine JM, Herrmann FR, Rodo X. Evaluating the performance of a climate-driven mortality model during heat waves and cold spells in Europe. Int J Environ Res Public Health. 2015; 12(2):1279–94. https://doi.org/10.3390/ijerph120201279 PMID: 25625407

19. Lowe R, Garcia-Diez M, Ballester J, Creswick J, Robine JM, Herrmann FR, et al. Evaluation of an Early-Warning System for Heat Wave-Related Mortality in Europe: Implications for Sub-seasonal to Seasonal Forecasting and Climate Services. Int J Environ Res Public Health. 2016; 13(2):206. https://doi.org/10.3390/ijerph13020206 PMID: 26861369

20. Huang P, Tan H, Liu A, Feng S, Chen M. Prediction of posttraumatic stress disorder among adults in flood district. BMC Public Health. 2010; 10:207. https://doi.org/10.1186/1471-2458-10-207 PMID: 20420677

21. Wang Y, Song Q, Du Y, Wang J, Zhou J, Du Z, et al. A random forest model to predict heatstroke occurrence for heatwave in China. Sci Total Environ. 2019; 650(Pt 2):3048–53. https://doi.org/10.1016/j.scitotenv.2018.09.369 PMID: 30373081

22. Park M, Jung D, Lee S, Park S. Heatwave Damage Prediction Using Random Forest Model in Korea. Appl Sci-Basel. 2020; 10(22).

23. HHS emPOWER Program https://empowerprogram.hhs.gov/in-action.html: US Department of Health & Human Services; 2023 [Available from: https://empowerprogram.hhs.gov/de-identified-dataset.html.

24. Maani N, Galea S. COVID-19 and Underinvestment in the Public Health Infrastructure of the United States. Milbank Q. 2020; 98(2):250–9. https://doi.org/10.1111/1468-0009.12463 PMID: 32333418

25. Kovats RS, Hajat S. Heat stress and public health: a critical review. Annu Rev Public Health. 2008; 29:41–55. https://doi.org/10.1146/annurev.publhealth.29.020907.090843 PMID: 18031221

26. Robine JM, Michel JP, Herrmann FR. Excess male mortality and age-specific mortality trajectories under different mortality conditions: a lesson from the heat wave of summer 2003. Mech Ageing Dev. 2012; 133(6):378–86. https://doi.org/10.1016/j.mad.2012.04.004 PMID: 22564660

27. Monteiro A, Carvalho V, Oliveira T, Sousa C. Excess mortality and morbidity during the July 2006 heat wave in Porto, Portugal. Int J Biometeorol. 2013; 57(1):155–67. https://doi.org/10.1007/s00484-012-0543-9 PMID: 22547142

28. Madrigano J, Mittleman MA, Baccarelli A, Goldberg R, Melly S, von Klot S, et al. Temperature, myocardial infarction, and mortality: effect modification by individual- and area-level characteristics. Epidemiology. 2013; 24(3):439–46. https://doi.org/10.1097/EDE.0b013e3182878397 PMID: 23462524

29. McGuinn L, Hajat S, Wilkinson P, Armstrong B, Anderson HR, Monk V, et al. Ambient temperature and activation of implantable cardioverter defibrillators. Int J Biometeorol. 2013; 57(5):655–62. https://doi.org/10.1007/s00484-012-0591-1 PMID: 22990411

30. Sheridan SC, Dolney TJ. Heat, mortality, and level of urbanization: measuring vulnerability across Ohio, USA. Clim Res. 2003; 24(3):255–65.

31. Zhang XH, Li YH, Cheng YB, Wang Y, Wang Y, Yao XY. Assessment of Regional Health Vulnerability to Extreme Heat—China, 2019. China Cdc Weekly. 2021; 3(23):490–+. https://doi.org/10.46234/ccdcw2021.126 PMID: 34594920