Environmental Advances 24 (2026) 100713

Contents lists available at ScienceDirect

Environmental Advances

journal homepage: www.sciencedirect.com/journal/environmental-advances

Mapping of local heat-related mortality risk using a hybrid statistical and machine learning framework

Maria Miguel Oliveira [a,b,c], Ana Margarida Alho [d,e,f], Ana Patrícia Oliveira [g,h], Paulo Nogueira [a,b,c,i,*]

[a] Faculty of Medicine, University of Lisbon, Avenida Professor Egas Moniz, 1649-026 Lisbon, Portugal
[b] Institute for Environmental Health, Associate Laboratory TERRA, Faculty of Medicine, University of Lisbon, Avenida Professor Egas Moniz, 1649-026 Lisbon, Portugal
[c] Biomathematics' Laboratory, Faculty of Medicine, University of Lisbon, Avenida Professor Egas Moniz, 1649-026 Lisbon, Portugal
[d] NOVA National School of Public Health, Public Health Research Center, Comprehensive Health Research Center, CHRC, NOVA University Lisbon, Lisbon, Portugal
[e] Public Health Unit USP Francisco George, ULS Santa Maria, Lisbon, Portugal
[f] Global Health and Tropical Medicine, GHTM, Instituto de Higiene e Medicina Tropical, IHMT, Universidade NOVA de Lisboa, Lisbon, Portugal
[g] CoLAB + ATLANTIC, IPL-ESTM, Rua do Conhecimento n°4, 2520-614 Peniche, Portugal
[h] Centro de Estudos Geográficos, IGOT—Instituto de Geografia e Ordenamento do Território, University of Lisbon, Lisbon, Portugal
[i] CIDNUR - Centro de Investigação, Inovação e Desenvolvimento em Enfermagem de Lisboa, Escola Superior de Enfermagem de Lisboa, Avenida Professor Egas Moniz, 1600-190 Lisbon, Portugal

ARTICLE INFO

Keywords:
Excess mortality
Extreme heat
Machine learning
Mixed-effects models
GPBoost
Public health surveillance

ABSTRACT

Background: Climate change is increasing the frequency and severity of extreme heat, raising risks to population health. These trends highlight the need for early-warning systems that can anticipate heat-related impacts and support timely public health action. In Portugal, the national heat warning system effectively triggers national alerts but does not account for differences in risk between municipalities.

Methods: We conducted a nationwide ecological retrospective study using a hybrid statistical–machine learning framework to estimate heat-related mortality risk across all municipalities in Portugal. Associations between extreme heat and mortality were estimated using a Generalised Linear Mixed Model with random intercepts per municipality and random slopes for extreme heat, while robustness and predictive performance were evaluated using GPBoost, a machine learning method combining gradient-boosted decision trees with Gaussian process–based random effects.

Findings: Extreme heat was associated with a 14.7% increase in mortality on extreme heat days (incidence rate ratio 1.147, 95% CI 1.133–1.161; $p < 0.001$). Nearly half of the total variance in mortality (47%) was attributable to between-municipality differences, stressing the need for spatially resolved modelling. These results were validated using machine learning, which estimated an 11.29% (9.98%–12.87%) relative increase in mortality associated with extreme heat.

Interpretation: By integrating formal statistical inference with machine-learning validation, this study demonstrates that reliable, policy-relevant downscaling of heat-related mortality risk is both feasible and essential at national scale. This framework is replicable to other countries, supporting the development of high-resolution heat–health warning systems and more targeted public health interventions to reduce the escalating health burden of extreme heat.

1. Introduction

The rise of extreme temperature events has been identified as one of the most important climate hazards for human health, with Southern European nations facing disproportionate exposure (Intergovernmental Panel on Climate Change (IPCC), 2022). A recent comprehensive study of 854 European cities estimated over 20,000 annual deaths attributable to heat, with substantial geographic heterogeneity and higher vulnerability in Mediterranean and Eastern European regions (Masselot et al., 2023). Portugal has experienced substantial heat-related mortality

* Corresponding author at: Faculty of Medicine, University of Lisbon, Avenida Professor Egas Moniz, 1649-026 Lisbon, Portugal.
E-mail address: pnogueira@medicina.ulisboa.pt (P. Nogueira).

https://doi.org/10.1016/j.envadv.2026.100713
Received 6 February 2026; Received in revised form 27 April 2026; Accepted 28 April 2026
Available online 29 April 2026

spikes, evidenced by approximately 2,000 excess deaths during a 2003 heatwave event (Nogueira, 2005) and over 2,400 excess deaths attributed to a heatwave event in July 2022 (United Nations Office for Disaster Risk Reduction (UNDRR), 2024).

Portugal's national heat-health warning system, ÍCARO (*Importância do Calor: Repercussões sobre os Óbitos*), has operated since 1999 as Europe's longest-running heat surveillance system (Kovats and Hajat, 2008; Nogueira et al., 1999). Developed by the National Institute of Health Dr. Ricardo Jorge, following the 1981 and 1991 heatwaves, ÍCARO produces daily heat-related mortality predictions shared with health action plan practitioners (Leite et al., 2020; Nogueira and Paixão, 2008). While effective for triggering national-level alerts, ÍCARO is inherently limited by its spatial resolution. The system relies on nationally defined thresholds and aggregated indicators, which are insufficient to capture the substantial spatial heterogeneity in heat-related health impacts observed at the sub-national level. As a result, municipalities with markedly different demographic structures, baseline health status, urban form, and climate exposures are treated as having comparable risk, masking local vulnerabilities and potentially diluting the effectiveness of interventions (Silva et al., 2024).

Environmental epidemiology increasingly recognises that traditional time-series approaches, while statistically rigorous, may fail to capture complex, non-linear relationships between climate exposures and health outcomes (Chen et al., 2024). On the other hand, machine learning (ML) methods provide a complementary approach by modelling intricate interactions and high-dimensional data (Hamilton et al., 2021; Rodríguez-González et al., 2019), despite facing criticisms for their 'black box' nature and limited capacity for causal inference (Nishi and Inoue, 2025).

We propose that the optimal approach lies in hybrid statistical–machine learning frameworks that leverage both statistical models for causal inference and ML algorithms for validation and discovery. Exemplifying this paradigm, GPBoost integrates gradient-boosted decision trees with Gaussian process and mixed-effects components, offering an alternative to conventional boosting methods like XGBoost or LightGBM (Sigrist, 2023, 2022). This architecture directly addresses a key limitation in applying ML methods to spatially nested data, a common structure in environmental epidemiology, where observed associations are strongly influenced by local-specific effects and outcomes cluster within defined geographic units of analysis. Recent applications have demonstrated GPBoost's superiority over fixed-effects ML methods for predicting outcomes in hierarchical epidemiological data, including COVID-19 severity prediction and spatial disease mapping (Sokhansanj and Rosen, 2022; Wong et al., 2023).

Building upon the need for higher spatial resolution in Portugal's ÍCARO system, this study introduces a hybrid modelling architecture to quantify the association between extreme heat events and all-cause mortality at the municipality level. While existing surveillance systems and traditional epidemiological studies have predominantly relied on broad regional aggregations, we establish a replicable pipeline that couples the interpretability of statistical methods with the predictive precision of machine learning. By applying this framework to a dynamically adaptive, threshold-based extreme heat index across an entire country, this approach aims to overcome the dichotomy between parametric statistical explanation and machine learning prediction. We aim to establish a replicable hybrid statistical–machine learning framework that can be extended to morbidity surveillance and applied in metropolitan-scale pilot implementations to support spatially resolved public health risk assessment.

2. Methods

2.1. Study design and data sources

A retrospective ecological observational study of daily all-cause mortality across all 278 municipalities of mainland Portugal was conducted. The study period encompassed 24 extended summer seasons (May 1 to September 30) spanning the years from 2000 to 2023, with a total of approximately 1.02 million municipality-day observations.

Daily mortality data were sourced from Statistics Portugal, and meteorological data were derived from the 'E-OBS daily gridded meteorological data for Europe from 1950 to present derived from in-situ observations' available through the Copernicus Climate Change Service (version v30.0e) (Cornes et al., 2018). E-OBS provides high-resolution daily temperature estimates on a $0.1^{\circ} \times 0.1^{\circ}$ regular grid (approximately 11×11 km) and for each municipality the daily maximum temperature series was retrieved by extracting and summarizing the grid cell values in the portion of a raster dataset that is covered by the polygon of each municipality (i.e., through zonal statistic), sampled according to the latest release of the Official Administrative Charter of Portugal, which depicts, as a polygon vector file, the delimitation and demarcation of the administrative districts of the country (Direção-Geral do Território, 2025). All spatial processing was implemented in Python using the exactextract library (ISciences, 2023; Python Software Foundation, 2024).

2.2. Heat exposure assessment: the GATO-IV index

Heat exposure was characterised using the Generalised Accumulated Thermal Overload index (GATO IV), developed specifically for Portuguese heat-health applications and currently employed in the national ÍCARO heat-health surveillance system (Nogueira and Paixão, 2008). Comparative analyses have demonstrated that GATO IV provides superior predictive power for heat-related mortality in Lisbon compared with alternative indices, including the Excess Heat Factor, particularly among older populations with cardiorespiratory conditions (Morais et al., 2020).

Unlike fixed temperature thresholds, GATO IV employs a dynamic, locally calibrated approach based on the 97.5$^{\text{th}}$ percentile of the historical temperature distribution, accounting for short-term acclimatisation. This percentile-based definition ensures that extreme heat is defined relative to local climatology, enhancing specificity across Portugal's climatically heterogeneous regions. Extreme heat days were then defined as any day where the maximum temperature exceeded the 97.5$^{\text{th}}$ percentile of the historical temperature distribution, using 1950-1999 as the reference period for percentile calculation, following the methodology of previous studies for this index's calculation (e.g., (Morais et al., 2020; Nogueira and Paixão, 2008)).

2.3. Generalised linear mixed model

Given that daily mortality exhibits significant overdispersion, standard Poisson regression was deemed unsuitable. Instead, we employed a GLMM with a negative binomial distribution and a log link function. The full model is defined as:

$$\log(\mu_{it}) = \beta_0 + \beta_1 \mathrm{GATO}_{it} + \beta_2 \mathrm{Month}_t + \beta_3 \mathrm{DOW}_t + \beta_4 \mathrm{COVID}_t + b_{0i} + b_{1i} \mathrm{GATO}_{it} \tag{1}$$

Where:

- μ_{it} represents the expected daily death count in municipality i on day t.
- Fixed Effects (β): β_1 estimates the global marginal effect of extreme heat, identified by GATO IV Q975. Temporal confounders are controlled via β_2 (Month) and β_3 (Day of the Week). β_4 captures the impact of the COVID-19 pandemic, operationalised as a binary indicator for the period of the official World Health Organization Public Health Emergency of International Concern (March 11, 2020, to May 5, 2023) (Sarker et al., 2023).
- Random Effects (b): The terms b_{0i} and b_{1i} denote the municipality-specific random intercepts and random slopes, respectively.

The random intercept (b_{0i}) accounts for differences in baseline population size and underlying mortality risks between municipalities. The random slope (b_{1i}) allows the sensitivity to extreme heat to vary stochastically across spatial units. By not assuming a uniform heat effect, this specification explicitly captures spatial heterogeneity in vulnerability, addressing a critical gap identified in recent European heat-health research (Lloyd et al., 2023; Zhang et al., 2023).

To quantify the variance structure, we calculated the Intraclass Correlation Coefficient (ICC), which measures the proportion of total mortality variance attributable to between-municipality differences. Furthermore, Best Linear Unbiased Predictions were extracted for the random slope terms (b_{1i}), representing the municipality-specific deviation from the national average heat effect.

2.4. GPBoost machine learning algorithm

To validate the robustness of the GLMM findings and capture potential non-linear associations between heat exposure and mortality, we implemented the GPBoost algorithm, a hybrid machine learning framework that combines gradient boosted decision trees with mixed-effects modelling (Sigrist, 2023, 2022). Standard regression models often assume linear or simple parametric relationships on meteorological variables, despite evidence that climate–health relationships are often complex and non-monotonic. GPBoost overcomes this limitation by enabling flexible, non-linear estimation of fixed effects while explicitly accounting for the hierarchical structure of spatially nested data. The GPBoost model was specified as a semi-parametric GLMM defined by:

$$\log(\mu_{\mathrm{it}}) = \mathrm{F}(\mathrm{X}_{\mathrm{it}}) + \mathrm{b}_{0\mathrm{i}} + \mathrm{b}_{1\mathrm{i}} GATO_{\mathrm{it}} \tag{2}$$

Where:

- μ_{it} represents the expected daily death count for municipality i at time t.
- $F(X_{it})$ is the gradient boosted ensemble of decision trees. This component captures non-linear global relationships and high-order interactions between the fixed effects and the outcome. To ensure comparability with the GLMM, this feature vector included the same covariates.
- b_{0i} denotes the random intercept, capturing municipality-specific baseline mortality risks.
- b_{1i} denotes the random slope for the extreme heat indicator (GATO_{it}).

The random effects were assumed to follow a multivariate normal distribution with an unstructured covariance matrix, estimated jointly with the tree ensemble parameters. The GPBoost model was trained using a 4-fold cross-validation strategy to ensure generalisability and mitigate overfitting. We performed a grid search to optimise the hyperparameters of the boosting algorithm:

- **Tree Complexity:** Maximum tree depth $\{3, 5, 7\}$ and number of leaves $\{31, 63\}$.
- **Learning Dynamics:** Learning rate $\{0.01, 0.05, 0.1\}$.
- **Regularisation:** Minimum observations per leaf $\{10, 20\}$.

We aimed to minimise the Root Mean Squared Error (RMSE) of the predictions on the validation folds. To further prevent overfitting, we employed early stopping with a patience of 10 rounds; training was halted if the validation metric did not improve for ten consecutive iterations.

To assess the global contribution of each predictor, we quantified feature importance using the Gain metric, which measures the total reduction in the training loss function accumulated from all splits involving a specific feature across the ensemble of trees. A higher total gain indicates that the feature was more effective at reducing prediction error.

Given the non-linear nature of the tree ensemble and the presence of random slopes, coefficients are not directly interpretable in the same manner as in standard regression models. Instead, we estimated the marginal effect of GATO IV Q975 using a counterfactual prediction framework. For every municipality i in the test set, we generated two potential outcomes:

1. **Baseline Scenario ($\widehat{y_i^{\mathbf{norm}}}$):** Predicted mortality with GATO IV Q975 forced to 0 for both the tree features and the random slope component.
2. **Extreme Heat Scenario ($\widehat{y_i^{\mathbf{gato}}}$):** Predicted mortality with GATO IV Q975 forced to 1 for both components.

The municipality-specific relative effect (RE_i) was calculated as the percentage increase in expected mortality attributable to heat, defined as:

$$RE_i = \left(\frac{\overline{\widehat{y_i}^{\mathrm{gato}}} - \overline{\widehat{y_i}^{\mathrm{norm}}}}{\overline{\widehat{y_i}^{\mathrm{norm}}}} \right) \times 100 \tag{3}$$

2.5. Train-test validation and sensitivity analysis

The dataset was partitioned based on a stratified random split into an 80% training set for model fitting and hyperparameter tuning, and a 20% test set reserved exclusively for final evaluation. Predictive performance was quantified using R^2, RMSE, Mean Absolute Error (MAE), and the Pearson correlation coefficient between observed and predicted mortality counts.

Sensitivity analysis were conducted to assess the robustness of the findings to different threshold definitions and different model specifications. Extreme heat was defined based on thresholds of the 90$^{\mathrm{th}}$ and 95$^{\mathrm{th}}$ historical percentiles, as well as employing a Poisson distribution, and including a fixed effect for year to control for long-term trends.

3. Results

During the extended period from 2000 to 2023, there was a total of 931,633 recorded all-cause deaths in Portugal, of which 71,777 occurred on extreme heat days (GATO IV Q975). Although the highest number of extreme heat days was observed in the northeastern region of mainland Portugal (Fig. 1, left), heat-related mortality was disproportionately concentrated in the major metropolitan areas of Lisbon and Porto. Despite experiencing fewer extreme heat days, these metropolitan regions accounted for the largest absolute number of deaths during heat events (Fig. 1, right).

When adjusting for population size (Fig. 2), the pattern shifts again. The highest daily average mortality rates are found in the interior North and Centre regions, a trend consistent during the full study period and on extreme heat days.

3.1. Generalised linear mixed model

The negative binomial GLMM identified a statistically significant association between extreme heat and daily mortality (Table 1). Incidence Risk Ratio (IRR) of GATO IV Q975 was 1.147 (95% CI: 1.133 – 1.161, $p < 0.001$), indicating that days exceeding the 97.5$^{\mathrm{th}}$ percentile temperature threshold are associated with a 14.7% increase in all-cause mortality risk.

A distinct weekly pattern was observed, with Mondays showing slightly elevated mortality (IRR = 1.017, $p < 0.001$) compared to Sundays. Seasonally, June to September demonstrated significantly lower baseline mortality compared to May. The COVID-19 pandemic period was associated with a 15.4% increase in daily death counts (IRR = 1.154, $p < 0.001$).

The random effects structure revealed substantial spatial

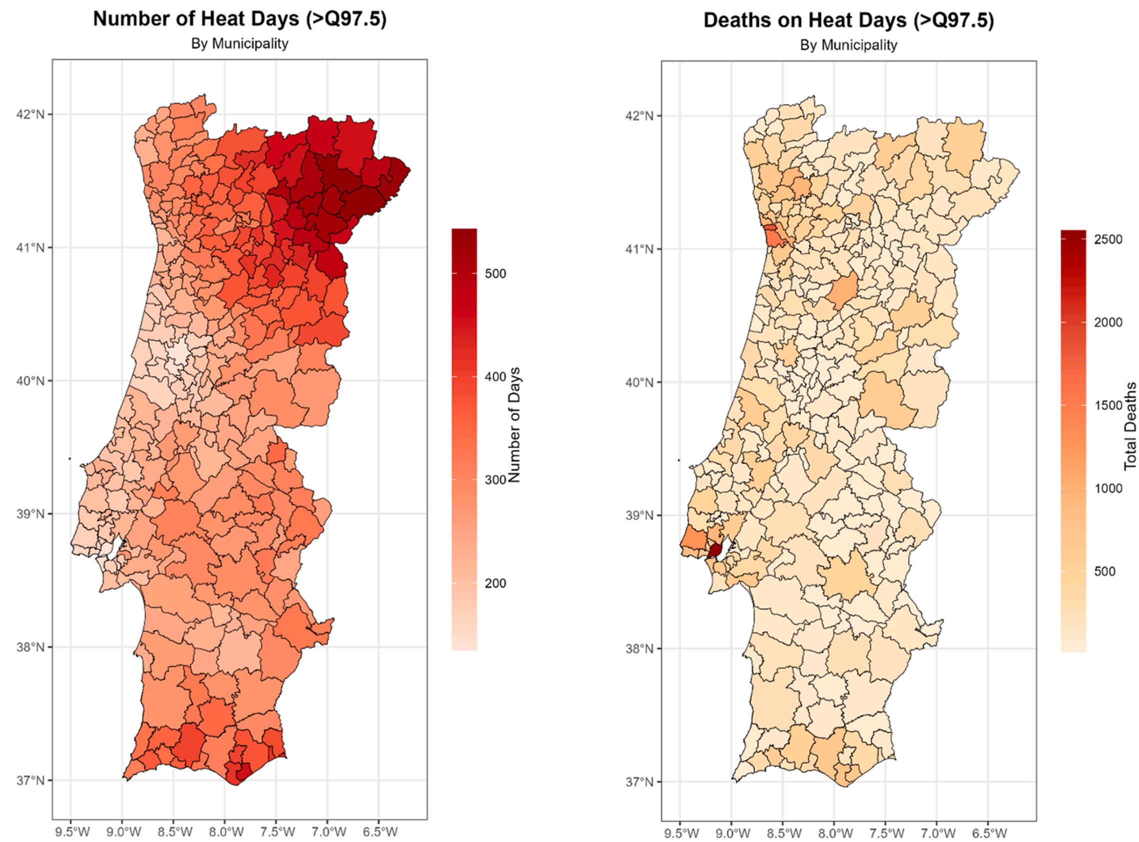

Fig. 1. Spatiotemporal patterns of extreme heat exposure and mortality burden in mainland Portugal (2000–2023). (Left) Cumulative frequency of extreme heat days (GATO IV Q975). (Right) Absolute mortality burden on extreme heat days.

heterogeneity as the residual variance (σ^2) was 0.91, with municipality-level intercept variance (τ_{00}) of 0.80 and a conditional ICC of 0.47, indicating that 47% of the variance in daily mortality is attributable to municipality-level differences, supporting the adoption of local-adjusted models. In contrast, random slope variance (τ_{11}) was 0.002, suggesting that the effect of extreme heat on mortality is relatively consistent across municipalities. Nonetheless, higher heat-associated risks were observed in Viseu, Porto, and Beira Interior regions (Fig. 3). The marginal R^2 was 0.002, while the conditional R^2 reached 0.471, demonstrating the contribution of municipality-level variation.

Training set performance produced an $R^2 = 0.682$, while test set performance was $R^2 = 0.680$, a difference of only 0.002, and the Pearson correlation coefficient of 0.826 and 0.824 demonstrated a strong linear relationship between observed and predicted values for both training and testing sets. Error metrics remained stable across data partitions, with an RMSE of 0.984 in the test set, and zero-count day prediction accuracy was 70.5%, having predicted 51.4% zeros against an observed frequency of 55.1%. Furthermore, the model exhibited minimal systematic error, with a prediction bias of −0.016 (Table 2).

3.2. GPBoost algorithm

The COVID-19 pandemic indicator was the dominant primary predictor (Gain = 0.726), indicating that viral transmission waves accounted for nearly 73% of the explained reduction in prediction error. Month ranked second (Gain = 0.167), capturing intra-seasonal baseline variations, while day of the week variations represented a minor gain (Gain = 0.057). After controlling these temporal confounders, the GATO IV Q975 indicator emerged as the top environmental predictor (Gain = 0.050).

The model demonstrated high stability, with testing metrics closely matching training performance (Test $R^2 = 0.679$ vs. Train $R^2 = 0.682$) (Table 3), indicating no significant overfitting despite the complexity of the gradient boosting architecture.

The model identified substantial spatial heterogeneity in baseline mortality ($\sigma^2_{intercept} = 0.799$), indicating that local demographic or structural factors created large disparities in daily death counts that fixed features alone could not explain. Furthermore, the non-zero variance in the heat response coefficients ($\sigma^2_{slope} = 0.013$) validates the hypothesis that vulnerability to extreme heat is not uniform across the country. Counterfactual analysis estimated a 11.29% (95% CI 9.98%–12.87%) relative increase in mortality under extreme heat scenarios, and municipality-specific effects ranged from −2.40% to 39.6%, concordant with the GLMM estimate (Fig. 4).

3.3. Sensitivity analysis

Sensitivity analysis with different threshold definitions estimated lower risk metrics for heat-related mortality (Table 4) for both the GLMM and GPBoost algorithms, with similar error metrics across all models. Including year as a fixed effect to control for long-term trends led to a marginal reduction in IRR and an average risk increase in the GPBoost algorithm. Moreover, changing from a Negative Binomial distribution to Poisson did not alter the results for the GLMM and GPBoost.

4. Discussion

This study demonstrates a hybrid statistical-machine learning approach to heat-mortality risk estimation that combines the inferential strengths of statistical mixed models with the predictive power of machine learning. Extreme heat events are associated with a 14.7% increase in the daily risk of mortality, and nearly half (47%) of the variance in mortality is attributable to municipality-level heterogeneity,

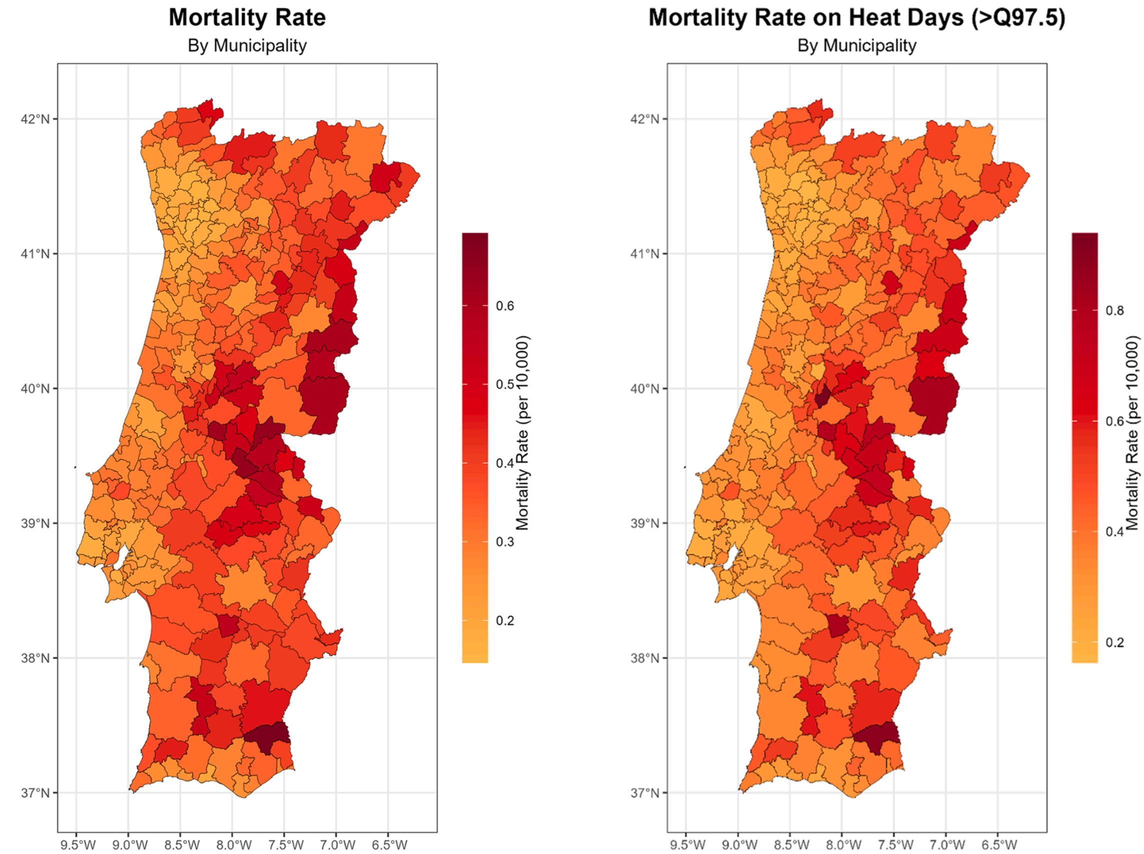

Fig. 2. Comparison of baseline and heat-related mortality rates. (Left) Average daily mortality rate per 10,000 inhabitants under non-heat conditions. (Right) Average daily mortality rate per 10,000 inhabitants during identified extreme heat days.

with implications for surveillance architecture. The convergence of results between the GLMM and GPBoost architecture provides a cross-methodological validation, confirming that the identified mortality risks and spatial patterns remain robust across both frequentist and machine learning paradigms. Nonetheless, the GLMM produced a slightly higher risk estimate compared to the GPBoost framework (11.29%), a discrepancy related to the architectural differences between the GLMM, which imposes a global log-linear assumption, and GPBoost, which captures localized threshold effects and multi-way interactions (Breiman, 2001; Hastie et al., 2009; Sigrist, 2022). As such, by deriving empirical marginal effects, the ML framework estimates a localized measure of the true mortality penalty (Athey and Imbens, 2019; Molnar, 2025).

The identified spatial heterogeneity is consistent with small-area analyses across Europe. For example, Zhang et al. (2023) used multi-country data from Norway, England/Wales, and Germany to assess the effects of heat on cardiovascular mortality, demonstrating both individual and area-level vulnerability factors, with higher susceptibility in regions with high urbanization and population density. Similarly, Barceló and colleagues (Barceló and Saez, 2025) employed Bayesian spatiotemporal models in Catalonia's 379 basic health areas, revealing over 40% excess mortality attributable to extreme heat, while Ballester and colleagues showed substantial spatial variability in heat-mortality relationships throughout Europe, with an estimated 211 deaths per million attributed to heat in Portugal during the summer of 2022 (Ballester et al., 2023). Morais and colleagues estimated that 60% of variance in elderly heat-related cardiorespiratory mortality in Lisbon is attributed to spatial heterogeneity (Morais et al., 2021).

The concurrent use of GLMMs with GPBoost represents a methodological advance in climate epidemiology, as traditional approaches have largely treated statistical models and ML as competing paradigms. We argue for their complementary deployment: GLMMs provide interpretable coefficients with uncertainty quantification, enabling hypothesis testing and causal inference under well-understood assumptions, while GPBoost offers flexible functional forms, model-agnostic feature importance, and counterfactual prediction capabilities that validate and extend statistical findings. The convergence of both approaches strengthens causal interpretation: if fundamentally different modeling strategies yield concordant estimates, confounding or model misspecification is less likely. A critical methodological divergence was observed in low-population municipalities, where estimates extended into negative values. While the GLMM returned consistently positive risk estimates due to Bayesian shrinkage, which "borrows strength" from the global dataset to stabilize local estimates, the GPBoost model exhibited higher variance in data-sparse regions. This behavior reflects the fundamental trade-off between statistical inference and machine learning: the gradient boosting algorithm captures raw local patterns, including the stochastic zeros inherent in small-area mortality data. Negative counterfactuals in these minority strata should be interpreted as stochastic noise rather than protective associations.

Furthermore, the combination of multiple data-driven modelling approaches allows to benefit from the strengths of each type of architecture, limiting the risk of recognising spurious results stemming from method-specific biases, and reinforcing confidence in the results. For example, some authors have been recommending that epidemiologists should fit multiple plausible models/algorithms and combine them as a multi-model ensemble (e.g., via cross-validated weighting) to deliver robust predictions (Clark et al., 2025).

The GPBoost algorithm addresses a critical challenge in environmental epidemiology: applying ML to hierarchically nested data. Standard boosting algorithms such as XGBoost ignore the correlation structure induced by geographic clustering, introducing potential biases

Table 1
Associations between extreme heat, temporal factors, and daily all-cause mortality estimated by the generalised linear mixed model (GLMM). Note: The model was fitted to 278 municipalities using a negative binomial distribution. GATO IV Q97.5: Indicator for days exceeding the 97.5th percentile of the generalised accumulated thermal overload index. IRR: Incidence rate ratio. CI: Confidence interval.

	All-cause mortality	
Predictors	Incidence Rate Ratios	p
Intercept	0.556 (0.501 – 0.618)	**<0.001**
GATO IV <Q97.5	1	**-**
GATO IV Q97.5	1.147 (1.133 – 1.161)	**<0.001**
May	1	**-**
June	0.973 (0.965 – 0.980)	**<0.001**
July	0.975 (0.967 – 0.982)	**<0.001**
August	0.982 (0.975 – 0.989)	**<0.001**
September	0.949 (0.942 – 0.956)	**0.001**
Sunday	1	**-**
Monday	1.017 (1.008 – 1.026)	**<0.001**
Tuesday	1.012 (1.003 – 1.021)	**0.008**
Wednesday	1.006 (0.998 – 1.015)	0.150
Thursday	1.004 (0.996 – 1.013)	0.322
Friday	1.001 (0.993 – 1.010)	0.747
Saturday	1.001 (0.992 – 1.009)	0.891
Outside the COVID-19 Pandemic Dates	1	-
Inside the COVID-19 Pandemic Dates	1.154 (1.146 – 1.162)	<0.001
Random Effects		
σ^2	0.91	
$\tau_{00\ \mathrm{municipality}}$	0.802	
$\tau_{11\ \mathrm{municipality.\ GATO\ IV\ Q97.5}}$	0.002	
$\rho_{01\ \mathrm{municipality}}$	−0.09	
ICC	0.47	
Marginal R^2 / Conditional R^2	0.002 / 0.471	

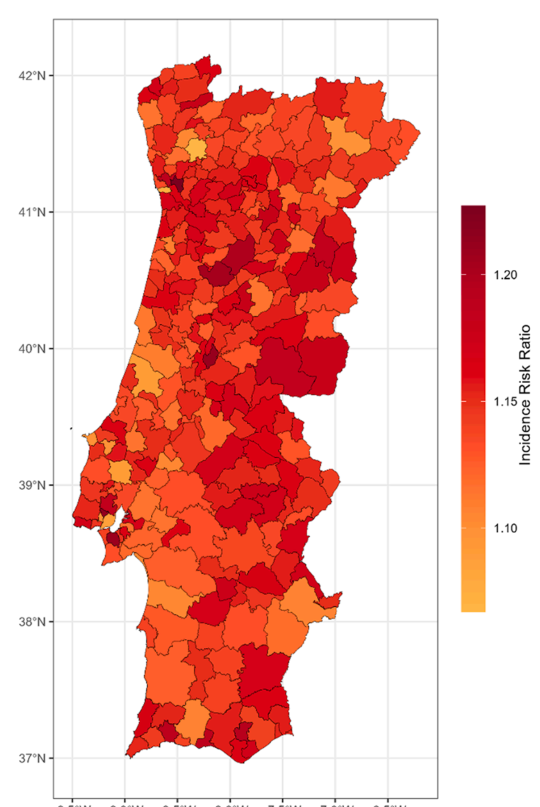

Fig. 3. Spatial distribution of heat-mortality risk (incidence risk ratios) estimated by the GLMM.

Table 2
Comparative predictive performance of the GLMM on training (80%) and testing (20%) datasets.

	Training (n = 800,640)	Testing (n = 220,176)
R^2	0.682	0.680
Root Mean Squared Error	0.984	0.969
Mean Absolute Error	0.681	0.676
Bias	-	−0.016
Zero accuracy	70.7%	70.6%
Correlation	0.826	0.824

Table 3
Performance metrics of the mixed-effects gradient boosting (GPBoost) model.

	Training (n = 800,640)	Testing (n = 220,176)
R^2	0.682	0.679
Root Mean Squared Error	0.983	0.970
Mean Absolute Error	0.681	0.676

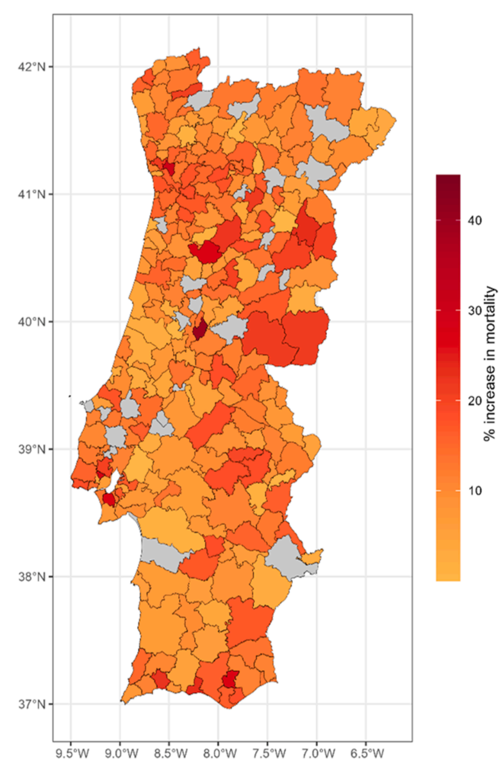

Fig. 4. Spatial distribution of heat-attributable mortality risk. Values represent the counterfactual percentage increase in daily mortality during extreme heat events compared to non-heat days. Note: Negative values were truncated to zero for visualization purposes.

into variance estimates and overconfident predictions. By embedding random effects within the boosting framework, GPBoost preserves ML's predictive advantages while respecting epidemiological data structure. This hybrid approach has gained traction in health research, with applications spanning COVID-19 epidemiology, cancer outcomes prediction, and spatial disease mapping (Levy et al., 2022; Sokhansanj and Rosen, 2022; Wong et al., 2023). Wang et al. (2024) recently developed a multi-scale ML model for heat-mortality in Germany, estimating 48, 000 heat-related deaths between 2014 and 2023 and projecting 2.5–9 × increases by 2100 under climate change scenarios. Our framework extends this tradition by explicitly combining statistical and ML inference for enhanced robustness.

Our findings provide robust empirical evidence that national-level

Table 4
Sensitivity analysis results.

	Model	IRR	Train RMSE	Train MAE	Train R2	Test RMSE	Test MAE	Test R2
Generalised Linear Mixed Model	Main Model (NegBin2, Q97.5)	1.147	0.984	0.681	0.682	0.969	0.676	0.679
	Poisson Distribution	1.147	0.984	0.681	0.682	0.969	0.676	0.679
	Included Year	1.137	0.985	0.680	0.681	0.970	0.675	0.679
	Alternative Threshold (Q95)	1.124	0.983	0.680	0.682	0.969	0.676	0.680
	Alternative Threshold (Q90)	1.111	0.983	0.680	0.682	0.969	0.675	0.679
GPBoost		% increase in mortality	Train RMSE	Train MAE	Train R2	Test RMSE	Test MAE	Test R2
	Main Model (Q97.5)	11.29	0.983	0.681	0.682	0.970	0.676	0.679
	Poisson Distribution	11.40	0.983	0.681	0.682	0.970	0.676	0.679
	Including Year	9.09	0.978	0.678	0.686	0.972	0.676	0.677
	Alternative Threshold (Q95)	10.34	0.982	0.680	0.682	0.969	0.676	0.679
	Alternative Threshold (Q90)	9.29	0.982	0.680	0.683	0.969	0.676	0.679

aggregate models are insufficient for capturing the granular realities of climate change impacts. The substantial spatial heterogeneity identified by the mixed-effects framework confirms that vulnerability is locally determined, challenging the "one-size-fits-all" paradigm often employed in early heat-health warning systems. This shift towards precision public health is critical: effective adaptation requires moving beyond broad national alerts to municipality-level surveillance that can guide targeted municipal interventions.

The feasibility of real-time, machine-learning-enhanced forecasting has been demonstrated in recent international contexts (Janoš et al., 2025), and our results suggest such systems are viable for Portugal. Integrating these municipality-specific risk predictions into existing surveillance infrastructures, such as the ÍCARO system, would enable a move from reactive crisis management to proactive resource allocation, allowing local authorities to deploy cooling centers and emergency teams precisely where and when they are needed most.

Several limitations in our study merit consideration. First, the ecological study design precludes individual-level inference: the identified associations represent population-average effects that may not apply uniformly within municipalities. Second, mortality data were limited to all-cause deaths; age group and cause-specific analysis would refine mechanistic understanding of extreme heat risk profiles. Third, while GPBoost handles spatial correlation through random effects in the model's likelihood function, temporal autocorrelation within municipalities was addressed only through temporal controls (month, day of week) rather than explicit time-series structures such as distributed lag non-linear models. Fourth, the E-OBS gridded dataset, while providing consistent coverage based on in-situ observations, may underestimate temperatures in complex coastal and mountainous terrain compared to the actual local-specific observed values. Finally, our analysis did not incorporate socioeconomic status or air pollution as potential confounders or effect modifiers, though recent European studies suggest important interactions between heat, ozone, and $PM_{2.5}$ exposure (Solano et al., 2025). Nonetheless, the role of socioeconomic status as a spatial confounder may be neutralized by the mixed-effect modelling structure (Zuur et al., 2009).

The hybrid framework validated here offers a scalable methodological template for next-generation environmental epidemiology. Future research priorities must now extend this architecture to encompass morbidity indicators, which often serve as earlier warning signals than mortality. Furthermore, integrating socioeconomic and demographic determinants as effect modifiers is essential to identify not just where risk is high, but who is most vulnerable, aligning with European frameworks for inequality-aware adaptation (Lloyd et al., 2023). Future research should also explore the integration of downscaled climate inputs, including air pollution data, that provide promising methodological pathways for generating hyper-local climate data (Najafi et al., 2026; Snaiki and Merabtine, 2025). Finally, future research must also focus on the assessment of extreme heat effects by specific cause of death and age cohorts, allowing for a better understanding of risk profiles.

5. Conclusions

This study establishes a methodological paradigm for climate-health surveillance by demonstrating the synergistic integration of statistical inference with machine learning validation. By coupling the structural interpretability of GLMMs with the predictive precision of GPBoost, we overcame the historical dichotomy between explanation and prediction. The convergence of these distinct modeling frameworks, despite their fundamental algorithmic differences, confirms that the 11.29 % to 14.7% estimated increase in heat-related mortality is a robust physiological signal, not a methodological artifact.

Our analysis highlights the importance of moving beyond national aggregates for surveillance. With nearly half of the mortality variance attributable to local heterogeneity, reliance on country-level averages systematically obscures local vulnerabilities. The GPBoost analysis adds a critical layer of reliability to this finding by identifying a "conservative lower bound" of risk, distinguishing between the theoretical risk captured by statistical models and the realized risk signals present in historical data.

These findings support a transition toward more spatially resolved public health systems. As climate change intensifies, the ability to downscale risk assessment from the national to the municipal level is essential for effective adaptation. The framework presented here offers a replicable, rigorous approach for this purpose, providing health authorities with the granular evidence needed to target interventions where they are most needed, without requiring a complete overhaul of existing epidemiological methods.

Data availability

Mortality data are available from Statistics Portugal upon reasonable request. Climate data are publicly available from the Copernicus Climate Data Store (EOBS).

CRediT authorship contribution statement

Maria Miguel Oliveira: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis. **Ana Margarida Alho:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Ana Patrícia Oliveira:** Writing – review & editing, Visualization, Supervision, Project administration, Methodology, Funding acquisition, Data curation. **Paulo Nogueira:** Writing – review & editing, Writing – original draft, 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.

Acknowledgements

This work was funded by the Portuguese Foundation for Science and Technology (FCT) under the AI4HeatHealth project (Reference: 2024.07710.IACDC). The authors thank Statistics Portugal (INE) for providing access to mortality data.

References

Athey, S., Imbens, G.W., 2019. Machine learning methods that economists should know about. Annu. Rev. Econom. 11, 685–725. https://doi.org/10.1146/annurev-economics-080217-053433.

Ballester, J., Quijal-Zamorano, M., Méndez Turrubiates, R.F., Pegenaute, F., Herrmann, F.R., Robine, J.M., Basagaña, X., Tonne, C., Antó, J.M., Achebak, H., 2023. Heat-related mortality in Europe during the summer of 2022. Nat. Med. 29, 1857–1866. https://doi.org/10.1038/s41591-023-02419-z.

Barceló, M.A., Saez, M., 2025. Assessing excess mortality and heat-attributable risk during the summer of 2022 in Catalonia, Spain: a Bayesian spatiotemporal analysis. J. Geogr. Syst. https://doi.org/10.1007/s10109-025-00475-2.

Breiman, L., 2001. Statistical modeling: the two cultures (with comments and a rejoinder by the author). Stat. Sci. 16. https://doi.org/10.1214/ss/1009213726.

Chen, S., Yu, J., Chamouni, S., Wang, Y., Li, Y., 2024. Integrating machine learning and artificial intelligence in life-course epidemiology: pathways to innovative public health solutions. BMC Med. 22, 354. https://doi.org/10.1186/s12916-024-03566-x.

Clark, L.P., Zilber, D., Schmitt, C., Fargo, D.C., Reif, D.M., Motsinger-Reif, A.A., Messier, K.P., 2025. A review of geospatial exposure models and approaches for health data integration. J. Expo. Sci. Environ. Epidemiol. 35, 131–148. https://doi.org/10.1038/s41370-024-00712-8.

Cornes, R.C., van der Schrier, G., van den Besselaar, E.J.M., Jones, P.D., 2018. An ensemble version of the E-OBS temperature and precipitation data sets. J. Geophys. Res.: Atmosp. 123, 9391–9409. https://doi.org/10.1029/2017JD028200.

Direção-Geral do Território, 2025. Carta Administrativa Oficial de Portugal - CAOP 2024.

Hamilton, A.J., Strauss, A.T., Martinez, D.A., Hinson, J.S., Levin, S., Lin, G., Klein, E.Y., 2021. Machine learning and artificial intelligence: applications in healthcare epidemiology. Antimicrob. Stewardship Healthc. Epidemiol. 1, e28. https://doi.org/10.1017/ash.2021.192.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning. Springer New York, New York, NY. https://doi.org/10.1007/978-0-387-84858-7.

Intergovernmental Panel on Climate Change (IPCC), 2022. Climate change 2022: Impacts, Adaptation and Vulnerability: Contribution of The Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, New York, NY. https://doi.org/10.1017/9781009325844.

ISciences, L.L.C., 2023. exactextract: fast and accurate zonal statistics.

Janoš, T., Quijal-Zamorano, M., Shartova, N., Gallo, E., Méndez Turrubiates, R.F., Denisse Beltrán Barrón, N., Peyrusse, F., Ballester, J., 2025. Heat-related mortality in Europe during 2024 and health emergency forecasting to reduce preventable deaths. Nat. Med. 31, 4065–4074. https://doi.org/10.1038/s41591-025-03954-7.

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

Leite, A., Santos, A.J., Silva, S., Nunes, B., Mexia, R., Rodrigues, A.P., 2020. Assessing the use and understanding of the Portuguese heat–health warning system (ÍCARO). J. Public Health 42, 395–402. https://doi.org/10.1093/pubmed/fdaa029.

Levy, J.J., Bobak, C.A., Nasir-Moin, M., Veziroglu, E.M., Palisoul, S.M., Barney, R.E., Salas, L.A., Christensen, B.C., Tsongalis, G.J., Vaickus, L.J., 2022. Mixed effects machine learning models for colon cancer metastasis prediction using spatially localized immuno-oncology markers. Pac. Symp. Biocomput. 27, 175–186.

Lloyd, S.J., Quijal-Zamorano, M., Achebak, H., Hajat, S., Muttarak, R., Striessnig, E., Ballester, J., 2023. The direct and indirect influences of interrelated regional-level sociodemographic factors on heat-attributable mortality in Europe: insights for adaptation strategies. Environ. Health Perspect. 131. https://doi.org/10.1289/EHP11766.

Masselot, P., Mistry, M., Vanoli, J., Schneider, R., Iungman, T., Garcia-Leon, D., Ciscar, J.-C., Feyen, L., Orru, H., Urban, A., Breitner, S., Huber, V., Schneider, A., Samoli, E., Stafoggia, M., De'Donato, F., Rao, S., Armstrong, B., Nieuwenhuijsen, M., Vicedo-Cabrera, A.M., Gasparrini, A., Achilleos, S., Kyselý, J., Indermitte, E., Jaakkola, J.J.K., Ryti, N., Pascal, M., Katsouyanni, K., Analitis, A., Goodman, P., Zeka, A., Michelozzi, P., Houthuijs, D., Ameling, C., Rao, S., das Neves Pereira da Silva, S., Madureira, J., Holobaca, I.-H., Tobias, A., Íñiguez, C., Forsberg, B., Åström, C., Ragettli, M.S., Analitis, A., Katsouyanni, K., Surname, F.n., Zafeiratou, S., Vazquez Fernandez, L., Monteiro, A., Rai, M., Zhang, S., Aunan, K., 2023. Excess mortality attributed to heat and cold: a health impact assessment study in 854 cities in Europe. Lancet Planet. Health 7, e271–e281. https://doi.org/10.1016/S2542-5196(23)00023-2.

Molnar, Christoph, 2025. Interpretable Machine Learning : A Guide for Making Black Box Models Explainable. ChristophMolnar.

Morais, L., Lopes, A., Nogueira, P., 2021. Human health outcomes at the neighbourhood scale implications: elderly's heat-related cardiorespiratory mortality and its influencing factors. Sci. Total Environ. 760, 144036. https://doi.org/10.1016/j.scitotenv.2020.144036.

Morais, L., Lopes, A., Nogueira, P., 2020. Which heatwave measure has higher predictive power to prevent health risks related to heat: EHF or GATO IV? – evidence from modelling Lisbon mortality data from 1980 to 2016. Weather Clim. Extrem. 30, 100287. https://doi.org/10.1016/j.wace.2020.100287.

Najafi, H., Lagerwall, G.L., Obeysekera, J., Liu, J., 2026. Machine learning in climate downscaling: a critical review of methodologies, physical consistency, and operational applications. Water 18, 271. https://doi.org/10.3390/w18020271.

Nishi, A., Inoue, K., 2025. What epidemiologists can do in the era of machine learning and artificial intelligence. J. Epidemiol. 35, JE20240412. https://doi.org/10.2188/jea.JE20240412.

Nogueira, P., 2005. Examples of heat health warning systems: Lisbon's ÍCARO's surveillance system, Summer of 2003. Extreme Weather Events and Public Health Responses. Springer-Verlag, Berlin/Heidelberg, pp. 141–159. https://doi.org/10.1007/3-540-28862-7_14.

Nogueira, P., Paixão, E., 2008. Models for mortality associated with heatwaves: update of the Portuguese heat health warning system. Int. J. Climatol. 28, 545–562. https://doi.org/10.1002/joc.1546.

Nogueira, P.J., Nunes, B., Dias, C.M., Falcão, J.M., 1999. Um sistema de vigilância e alerta de ondas de calor com efeitos na mortalidade : O índice de Ícaro. Rev. Portug. Saúde Públ. 79–84.

Python Software Foundation, 2024. Python.

Rodríguez-González, A., Zanin, M., Menasalvas-Ruiz, E., 2019. Public health and epidemiology informatics: can artificial intelligence help future global challenges? An overview of antimicrobial resistance and impact of climate change in disease epidemiology. Yearb. Med. Inform. 28, 224–231. https://doi.org/10.1055/s-0039-1677910.

Sarker, R., Roknuzzaman, A.S.M., Nazmunnahar, Shahriar, M., Hossain, M.J., Islam, M. R., 2023. The WHO has declared the end of pandemic phase of COVID-19: way to come back in the normal life. Health Sci. Rep. 6. https://doi.org/10.1002/hsr2.1544.

Sigrist, F., 2023. Latent Gaussian model boosting. IEEE Trans. Pattern Anal. Mach. Intell. 45, 1894–1905. https://doi.org/10.1109/TPAMI.2022.3168152.

Sigrist, F., 2022. Gaussian process boosting. J. Mach. Learn. Res. 23, 1–46.

Silva, C.S., Fernandes, D.V., Gomes, R., Costa, F.P., Pinto, L., Scuri, S., Brito, A., Nunes, B., Silva, S.P., 2024. CLIMAEXTREMO: a new risk indicator for the health risk to building occupants during extreme weather events in Portugal. Sustainability 16, 5171. https://doi.org/10.3390/su16125171.

Snaiki, R., Merabtine, A., 2025. Recent advances on machine learning techniques for urban heat island applications: a review and new horizons. Sustain. Cities Soc. 134, 106943. https://doi.org/10.1016/j.scs.2025.106943.

Sokhansanj, B.A., Rosen, G.L., 2022. Predicting COVID-19 disease severity from SARS-CoV-2 spike protein sequence by mixed effects machine learning. Comput. Biol. Med. 149, 105969. https://doi.org/10.1016/j.compbiomed.2022.105969.

Solano, D., Solans, M., Perafita, X., Ruiz-Comellas, A., Saez, M., Barceló, M.A., 2025. A spatiotemporal Bayesian hierarchical model of heat-related mortality in Catalonia, Spain (2012–2022): the role of environmental and socioeconomic modifiers.

United Nations Office for Disaster Risk Reduction (UNDRR), 2024. Portugal heatwave, 2022 - forensic analysis. UN Global Assessment Report on Disaster Risk Reduction.

Wang, J., Nikolaou, N., an der Heiden, M., Irrgang, C., 2024. High-resolution modeling and projection of heat-related mortality in Germany under climate change. Commun. Med. 4, 206. https://doi.org/10.1038/s43856-024-00643-3.

Wong, S., Flegg, J.A., Golding, N., Kandanaarachchi, S., 2023. Comparison of new computational methods for spatial modelling of malaria. Malar. J. 22, 356. https://doi.org/10.1186/s12936-023-04760-7.

Zhang, S., Breitner, S., Rai, M., Nikolaou, N., Stafoggia, M., de' Donato, F., Samoli, E., Zafeiratou, S., Katsouyanni, K., Rao, S., Palomares, A.D.-L., Gasparrini, A., Masselot, P., Aunan, K., Peters, A., Schneider, A., 2023. Assessment of short-term heat effects on cardiovascular mortality and vulnerability factors using small area data in Europe. Environ. Int. 179, 108154. https://doi.org/10.1016/j.envint.2023.108154.

Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A., Smith, G.M., 2009. Mixed effects models and extensions in ecology with R. Springer New York, New York, NY. https://doi.org/10.1007/978-0-387-87458-6.