Seasonal Heatwave Forecasting with Explainable Machine Learning and Remote Sensing Data

Jung-Ching Kan
jckan@kth.se

KTH Royal Institute of Technology

Marlon Vieira Passos
KTH Royal Institute of Technology

Georgia Destouni
Stockholm University

Karina Barquet
Stockholm Environment Institute (SEI)

Carla S.S. Ferreira
Polytechnic Institute of Coimbra, Applied Research Institute

Zahra Kalantari
KTH Royal Institute of Technology

Research Article

Keywords: Summer heatwaves, machine-learning models, explanatory-predictive factors, land surface factors, atmospheric climate factors, geopotential height

Posted Date: November 28th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-5267311/v1

Additional Declarations: No competing interests reported.

Abstract

Heatwaves can greatly impact societies, underscoring the need to extend current heatwave prediction lead times. This study investigated multiple machine-learning (ML) model approaches for heatwave occurrence prediction with long lead times of one to five months based on explanatory atmospheric and land surface features. Five ML classifiers were built using Google Earth Engine remote sensing datasets to predict heatwaves at national scale (Sweden) based on 16 features referring to the period of 1989–2019. Extreme Gradient Boosting performed best for lead times of one month (F1-score = 0.63, accuracy = 0.81) and four months (F1-score = 0.54, accuracy = 0.79), while K-Nearest Neighbour was best for lead times of two, three and five months (respective F1-score = 0.63, 0.65, 0.49, accuracy = 0.77, 0.79, 0.78). When applying the SHapley Additive exPlanations technique for model interpretation, land surface features emerged as more impactful heatwave predictors than atmospheric features at longer lead times. More frequent heatwave occurrence was associated with places in Sweden characterized by lower values of geopotential height, latitude, topographical slope, evaporation, precipitation and cropland area, and higher values of average temperature, mean sea level pressure, and southerly and westerly winds. The study also concretely exemplifies how use of this multi-model ML method can enhance predictions and further step-wise improve them, thereby facilitating earlier warning in support of better planning of measures to mitigate adverse heatwave impacts, up to several months ahead of their possible occurrence.

1 Introduction

A heatwave is an extreme temperature event defined as a period of excessive heat that usually lasts for one week or at least three days (Perkins-Kirkpatrick and Lewis 2020). Research into the impacts of heatwaves on human health has consistently revealed that a disproportionately high number of older individuals are affected (Xu et al. 2016; Vu et al. 2019; Rodrigues et al. 2021). The 2003 heatwave in Europe led to 70,000 excess mortalities (Robine et al. 2008) and over 13 billion euros in economic damage (United Nations Environment Programme 2003). Heatwaves can also have far-reaching cascading and compound impacts, exacerbating drought impacts on water, soil, energy and agriculture sectors (Niggli et al. 2022; Soares et al. 2023). Global predictions indicate that climate change will lead to increased intensity, frequency, and duration of heatwaves, with Europe identified as a future hotspot. The projected trend for Europe is three to four times faster than in other northern mid-latitude regions (Rousi et al. 2022).

Early warnings of heatwaves with a sufficient lead time enable timely and more effective societal responses to protect vulnerable populations and socioeconomic activities (Merz et al. 2020). State-of-the-art deterministic heatwave prediction models can predict heatwaves up to two weeks in advance (Domeisen et al. 2023), but this is often insufficient for preparing and implementing efficient mitigation measures. Predictions beyond that lead time are challenging due to the increase in complex interactions of variables over time, such as between atmospheric variables (e.g. wind speed and temperature) and land surface variables status and feedbacks (e.g. land cover and soil moisture) (Domeisen et al. 2023).

Extending heatwave predictions to a seasonal lead time would allow for a timelier and more effective reduction of risk (Weirich-Benet et al. 2023).

Machine-learning (ML) approaches can potentially increase heatwave prediction lead times as they have the advantage of identifying, rather than pre-assuming, the complex interactions between various possible explanatory-predictive factors. ML-based modelling is increasingly being applied in studies of hydro-climatic hazards, such as floods, droughts and heatwaves (Rahmati et al. 2020; Panahi et al. 2022). A study by Khan et al. (2021) achieved accurate predictions with a one-month lead time of total summer heatwave days (HWD) in Pakistan using a support vector machine. Asadollah et al. (2022) developed and compared three ML models (AdaBoost regression decision tree (ABR-DT), Random Forest and Decision Tree) for total summer HWD prediction in Iran with a three-month lead time, and found ABR-DT to be the most accurate. Weirich-Benet et al. (2023) predicted heatwaves at weekly resolution using different approaches and concluded that ML can improve sub-seasonal heatwave prediction. Straaten et al. (2022) found that explainable ML using high-dimensional remote sensing data can complement physically-based models for heatwave prediction, providing an effective alternative for such predictions with a lead time longer than two weeks. However, the development and application of ML models for heatwave research and prediction are still in their infancy compared with work on other natural hazards, such as droughts and floods (Mosavi et al. 2018; Gyaneshwar et al. 2023). Additionally, most ML/deep learning-based heatwave studies to date have considered sub-seasonal forecast lead times (less than one month), without applying explainability techniques, and the focus has been on relatively warm and/or dry climate regions such as the Middle East and Central Europe (Domeisen et al. 2023).

The aim of this study is to advance ML model development for heatwave prediction by creating a comprehensive framework for seasonal prediction with longer lead times and practical implementation. This research supports interdisciplinary efforts to mitigate climate change risks and underscores the importance of integrating scientific advances into environmental management and policy frameworks. As a practical case study, we focused on predicting summer HWD at a monthly resolution in Sweden. The specific objectives of this study were: (i) to build alternative ML models based on relevant remote sensing data for possible heatwave occurrences with lead times of one to five months, (ii) to evaluate and compare the performance of various ML models at different lead times, and (iii) to analyse the relative importance of various explanatory-predictive features (e.g. related to atmospheric climate, land surface and other relevant physical aspects) in terms of their influence on ML model outputs.

2 Materials and Methods

2.1 Study area

Sweden is located in the high-latitude region of northern Europe (55–69 °N; 11–24 °E), which is characterised by long, cold winters (Sköld Gustafsson et al. 2023b). Although major documented natural hazards in Sweden mostly comprise flooding and wildfire events (Sköld Gustafsson et al. 2023a), heatwave exposure is of increasing concern due to rising temperatures over the past two decades, e.g. mean peak temperatures in Sweden were higher in the period 1991–2020 (Fig. 1b) than in 1961–1990 (Fig. 1a). Trend analysis suggests that heavily populated areas in the southern part of the country will experience more prolonged and frequent heatwaves in the near future (Vieira Passos et al. 2024). A persistent heatwave in Sweden in the summer of 2018 resulted in around 750 cases of excess mortality (SMHI, 2020). Sweden faces a considerable risk of heatwave impacts due to its population being more acclimated to cooler climates. For example, Sweden is accustomed to building homes and organizing activities designed to withstand severe cold rather than heat (SWECO). Additionally, heatwaves are generally the most common primary hazard interacting with other natural hazards (Sköld Gustafsson et al. 2023b). Heatwaves in Sweden mostly occur in the summer months (June-August) and are defined in this study as days with the maximum temperature exceeding 27°C for at least three consecutive days, based on the threshold adopted by the Swedish Meteorological and Hydrological Institute (SMHI) for issuing high-temperature warnings (Oudin Åström et al. 2020).

2.2 Development method

The framework developed and tested in this study comprised six main steps (Fig. 2): (1) data extraction, (2) data pre-processing, (3) model training, validation, and testing, (4) model evaluation, (5) model explanation and (6) application.

2.2.1 Data extraction

A total of 21 possible explanatory-predictive physical features were selected for ML heatwave modelling, based on a literature review. These comprised seven atmospheric features (u component of wind speed (u wind), v component of wind speed (v wind), monthly mean temperature, mean sea level pressure (MSLP), specific humidity, precipitation and geopotential height), 11 land surface features (soil moisture, runoff, evaporation, surface latent heat flux, surface sensible heat flux (SSHF), soil water content at four different depths (0–7 cm, 7–28 cm, 28–100 cm and 100–289 cm, denoted soil water 1–4, respectively), land cover and slope, and three additional spatiotemporal features (heatwave occurrence in the previous month, and at various latitudes and longitudes) (Table 1).

Table 1
Features considered in this study, data source for each feature and brief description

Feature	Category	Data source	Description
u wind	Atmospheric	ERA 5	Monthly average 10 m u-component of wind speed
v wind	Atmospheric	ERA 5	Monthly average 10 m v-component of wind
Mean temperature	Atmospheric	ERA 5	Monthly average air temperature at 2 m height
Mean sea level pressure	Atmospheric	ERA 5	Monthly average sea level pressure
Specific humidity	Atmospheric	FEWS NET	Monthly average specific humidity
Precipitation	Atmospheric	ERA 5	Monthly sum of total precipitation
Geopotential height	Atmospheric	NCEP	Monthly average surface geopotential height
Soil moisture	Surface	FEWS NET	Monthly average soil moisture 0–10 cm underground
Runoff	Surface	ERA 5 land	Sum of surface runoff and subsurface runoff
Evaporation	Surface	ERA 5 land	Sum of total evaporation
Latent heat flux	Surface	ERA 5 land	Sum of latent heat exchange with surface
Sensible heat flux	Surface	FEWS NET	Sum of heat transfer between surface and atmosphere
Soil water 1	Surface	ERA 5 land	Volume of water in 0–7 cm soil layer
Soil water 2	Surface	ERA 5 land	Volume of water in 7–28 cm soil layer
Soil water 3	Surface	ERA 5 land	Volume of water in 28–100 cm soil layer
Soil water 4	Surface	ERA 5 land	Volume of water in 100–289 cm soil layer
Land cover	Surface	ESA	22 land cover classes defined with the United Nations Land Cover Classification System
Slope	Surface	USGS	
HWD in previous	Temporal		1 if heatwave in previous month, 0 otherwise

Feature	Category	Data source	Description
month			
Latitude	Spatial-geographic		
Longitude	Spatial-geographic		

The atmospheric and land surface features were retrieved from the remote sensing dataset using Google Earth Engine (GEE). GEE is a cloud computing platform that allows the acquisition of data from various sources for any specific study area worldwide, and is increasingly being used in heatwave-related studies (Zhang et al. 2022; Mamgain et al. 2023). The study period in GEE was set to 1989–2019, with a resolution of 27830 m based on data availability and quality. The preceding time feature of HWD occurrence in the previous month was only used as a possible explanatory-predictive feature in ML modelling with a one-month lead time. Figure 3 illustrates the Pearson correlation coefficient (R) between features and target, suggesting that there are no linear relationships between selected indicators and heatwave occurrence. Hence, advanced ML models could be deployed to address the nonlinear relationship. Heatwave days were identified based on daily maximum temperature. Since almost all heatwaves in Sweden occur in the summer months (June, July or August) (Sjulgård et al. 2023), only these were set as target months. Monthly heatwave occurrence was defined using binary classification, with a value of 1 allocated for one or more HWD during a month (Class 1) and a value of 0 for no HWD (Class 0).

2.2.2 Data pre-processing

The data obtained from GEE were subjected to three data pre-processing procedures, feature selection, feature scaling and data resampling to ensure high-quality, consistent, and suitable data for subsequent modelling. Data were split into training, validation and testing datasets, corresponding to data in 1989–2016, 2017–2018 and 2019, respectively.

Feature selection

Feature selection consisted of identifying an effective subset of features from the full set by removing noisy and redundant features (Tang et al. 2014). A Pearson correlation matrix of the 21 features was plotted with a heatmap to examine mutually correlated feature groups. Highly correlated features contained similar information and conveyed redundant information to the model, and therefore one feature in each highly correlated feature group was kept and the others were excluded.

Feature scaling

Feature scaling was applied to transform selected features to a common range, minimising the bias of feature value distribution (Singh and Singh 2020). This procedure is required since features present distinct value ranges in this study, and greater numerical feature values may dominate over smaller numerical feature values (Singh and Singh 2020). The minimum-maximum normalisation method was applied, where the maximum value was equal to 1 and the minimum to 0.

Data resampling

Resampling is a state-of-the-art solution to address imbalanced data (Alam et al. 2020). Based on historical data, months in which a heatwave occurred were relatively infrequent compared with months without a heatwave. This caused a class imbalance problem, leading to bias towards the majority class (Class 0) (Lin et al. 2017). Data imbalance is a common problem when dealing with real-world applications, e.g. natural hazards (Oommen et al. 2011). To balance the data without losing essential information, cluster centroids, a clustering-based under-sampling technique (Alam et al. 2020), was applied in this study. Cluster centroids replace the majority class samples with cluster centres, reducing the amount of majority class samples while maintaining the representative variations (Tsai et al. 2019). This resampling technique was only applied to the training dataset for the ML models.

2.2.3 Model training, validation and testing

Model development comprised the following steps: (1) model training with the training dataset, (2) model fine-tuning (hyperparameter tuning and threshold tuning) with the validation dataset, and (3) model testing with the testing dataset for independent performance evaluation. Lead times of one to five months were tested and the three steps above were repeated for each lead time considered. A total of 25 ML models were developed and tested in this study. Following a literature review, five popular ML classifiers (Extreme Gradient Boosting, Gaussian Naïve Bayes, K-Nearest Neighbour, Random Forest, and logistic regression) were selected based on their relevant characteristics. Brief summaries of each model class and their applications are provided below.

Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) is a highly effective and scalable algorithm derived from the ensemble of decision trees (Chen and Guestrin 2016). It has a built-in regularisation function to reduce overfitting, and parallel processing to speed up the model training process (Asadollah et al. 2022). A previous study used an XGBoost model to predict the number of heat-related ambulance calls and achieved high accuracy (Ke et al. 2023).

Gaussian Naïve Bayes

Gaussian Naïve Bayes (NB) is an efficient classifier identified as one of the top ten algorithms in data mining (Wu et al. 2008). Gaussian NB is based on Bayes' theorem, assuming the data are Gaussian-distributed and that all the features are independent (Wu et al. 2008). It has given satisfactory results in urban flood depth prediction (Wang et al. 2021).

K-Nearest Neighbour

K-Nearest Neighbour (KNN) is a generic non-parametric algorithm that does not make assumptions on data distribution (Wu et al. 2008). It does not learn the pattern, but memorises the training set, calculates the distance, and then decides based on the majority of k-nearest neighbours (Wu et al. 2008). This model has been extensively applied in drought prediction, with good results (Raja and T 2022).

Random Forest

Random Forest (RF) is an ensemble model based on averaging the results of randomised decision trees (Biau and Scornet 2016). It is one of the most widely applied models in environmental research, and has been widely tested in previous heatwave-related studies (Asadollah et al. 2022; Weirich-Benet et al. 2023).

Logistic regression

Logistic regression (LR) is a robust and flexible classification model for predicting a binary outcome, such as yes/no, when the features are continuous (Bartosik and Whittingham 2021). Logistic regression calculates a linear combination of the input features and then transforms them with a sigmoid function (Bartosik and Whittingham 2021). It is widely applied in the clinical field (Kannan et al. 2023) and hazard research, e.g. mapping of flood susceptibility and drought spatial patterns (Al-Juaidi et al. 2018; Niaz et al. 2021).

Model fine-tuning

Hyperparameters are external configuration variables of algorithms. Hyperparameter tuning consists of finding the optimal settings of hyperparameters to classify imbalanced data at algorithm level (Rosales-Pérez et al. 2023), possibly achieving significant ML model improvements (Kong et al. 2019). Random search, one of the two most widely used strategies for hyperparameter tuning, was selected for use in this study due to its advantages in terms of efficiency and flexibility (Bergstra and Bengio 2012).

Model prediction was determined probabilistically, with the normal default threshold for algorithm prediction classification set at 0.5, i.e. prediction probabilities greater than 0.5 were regarded as Class 1, while probabilities below 0.5 were regarded as Class 0 (Zou et al. 2016). However, a default threshold of 0.5 is not always suitable for imbalanced classification and leads to poor performance (Brownlee 2020). Therefore, the threshold value in this study was set as that yielding the highest F1-score in the validation dataset (see below).

2.2.4 Model evaluation

In general, the minority data class (with a heatwave in a month) was defined as positive (Class 1), while the majority class (with no heatwave in a month) was defined as negative (Class 0) (Zou et al. 2016). Two evaluation metrics, F1-score and accuracy, were selected to evaluate model performance because F1-score is a popular metric in imbalanced classification, especially when accurate prediction of heatwave occurrence (Class 1) is important (Zhang et al., 2017), and accuracy is the most commonly used metric for classification. These evaluation metrics are defined as follows (Luo et al. 2019):

$$F1 - score = \frac{2 * \frac{TP}{TP+FP} * \frac{TP}{TP+FN}}{\frac{TP}{TP+FP} + \frac{TP}{TP+FN}}$$

1

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

2

where TP (true positives) represents the number of positive predictions that are correctly predicted, TN (true negatives) is the number of negative predictions that are correctly predicted, FP is the number of false positive predictions, and FN is the number of false negative predictions.

2.2.5 Model interpretation

Machine-learning models are often considered black boxes and are increasingly being used for high-stakes prediction applications to support decision-making processes (Rudin 2019). Therefore, efforts and approaches to interpret model outputs physically are as crucial as improving predictability in order to provide scientific insights, enable model improvements and improve scientific understanding of the modelling process (Lundberg and Lee 2017). To this end, SHapley Additive exPlanations (SHAP), a state-of-the-art ML model interpretation technique based on cooperative game theory (Angelov et al. 2021), was employed. SHAP calculates the model output impact of each explanatory-predictive feature as if the different features were players in a coalition game, and the payoff of the features, referred to as the Shapley value, is a measure of their importance (Angelov et al. 2021). SHAP is a strong and insightful way to physically interpret the results of complex ML algorithms (e.g. Althoff & Destouni, 2023) and offers valuable perspectives on underlying dynamics that govern heatwave patterns.

3 Results

3.1 Feature selection

A correlation matrix heatmap illustrating the associations between the 21 possible explanatory-predictive features considered initially in this study (as outlined in Table 1) is shown in Fig. 4. Three highly correlated pairs of features were identified ($R > 0.8$): evaporation and surface latent heat flux, specific humidity and monthly mean temperature, and soil water content at different depths (1, 2, 3 4). These pairs showed a similar pattern, therefore only one of each pair was retained (evaporation, monthly mean temperature and soil water 1). The remaining 16 features made up the final optimal feature set in this study and were applied in ML modelling.

3.2 Model performance

The performance of the ML models in the testing procedure is shown in Fig. 5a. XGBoost performed best at prediction lead times of one month (F1-score = 0.63, accuracy = 0.81) and four months (F1-score = 0.54, accuracy = 0.79). KNN achieved the best predictive performance at lead times of two, three and five months (respective F1-score = 0.63, 0.65, 0.49, accuracy = 0.77, 0.79, 0.78). The RF model showed moderate performance for all lead times and was ranked second to fourth among the models, with an F1-score around 0.5 and accuracy around 0.75. The LR approach showed increasingly high ranking and predictive performance for longer lead times, ranking second for a lead time of five months (F1-score = 0.37, accuracy = 0.76). Gaussian NB showed a consistently poorer performance than the other models (F1-score < 0.5, accuracy < 0.7). Overall, model performance gradually declined with increasing prediction lead time, as indicated by a decreasing F1-score from 0.63 to 0.49 and accuracy from 0.81 to 0.74.

3.3 Feature importance

The feature importance ranking and impact value obtained using the SHAP approach for the best-performing models with lead times of one to five months is shown in Fig. 6. Geopotential height and geographic latitude emerged as key explanatory-predictive features for HWD occurrence, consistently ranking in the top five across various models and lead times. V wind is more important than u wind, and these two features generally ranked lower, appearing outside the top ten in terms of importance, which indicates a relatively minor influence on the model's predictive capability. Furthermore, soil moisture and geographic longitude exhibited enhanced importance in models over longer lead times, while evaporation was of decreased importance.

Figure 6 also shows the distribution and variability of SHAP values for each feature in the best-performing models for lead times of one to five months. The impact ranges for features became narrower and more concentrated around zero impact for longer prediction lead times, implying a weaker influence on model output at longer lead times. Overall, heatwave formation tended to be higher for lower values of geopotential height, precipitation, evaporation, latitude, and slope, and for higher values of average temperature, mean sea level pressure, u wind and v wind.

3.4 Example of application for heatwave in July 2019

Heatwave occurrence prediction maps were generated for July 2019 across Sweden utilising the best-performing model for each lead time in a series from longer to shorter lead times (Fig. 7). A large-scale heatwave did indeed occur in Sweden in July 2019, and the present predictions were made as one specific test example of the potential of the ML-based methodology framework developed in this paper to support early-warning systems. The maps in Fig. 7 present the actually observed heatwave occurrence (Fig. 7 (a)) alongside the best ML model predictions at lead times from five months to one month before the occurrence (Fig. 7 (b) to (f)).

Figure 7 illustrates the capability of the present multi-model ML methodology framework to predict heatwave occurrence with long lead times, and how this capability further improves as the lead time decreases. Specifically, the accurate heatwave predictions increase from 803 to 996 out of 1294 pixels across the whole area of Sweden.

4 Discussion

This study presents a ML methodology framework to comparatively develop, identify and extract insights from the most suitable models for heatwave occurrence prediction with seasonal lead times and global remote sensing datasets. The methodology framework could be adapted to other regions using global datasets on GEE platforms and open-source ML libraries. This study's results demonstrate that ML models can deliver a good predictive ability up to five months in advance. Ultimately, this study contributes to the strategic deployment of ML models in climate adaptation efforts, offering a methodical approach to enhancing predictive accuracy and operational readiness for heatwaves, thereby reducing their socio-economic impacts.

Among all the tested models, XGBoost and KNN showed the best performance overall across the different lead times (of one to five months) tested in this study. The superior performance of XGBoost in predicting HWD with lead times of one to three months can be attributed to its effectiveness at capturing nonlinear interactive feature effects (Zheng et al. 2021). To date KNN has been applied less to natural phenomena prediction than other popular algorithms such as RF and LR. In this study, however, KNN surprisingly outperformed those models. The good performance of KNN in this study may be because this model makes no assumption about data distribution, providing more flexibility for weather data (Badhiye et al. 2013), which may also inspire other scholars to test KNN performance in different natural hazard predictions.

Based on the feature importance analysis, two key factors—geopotential height and geographic latitude—were identified as principal contributors to heatwave formation. Land surface features, such as soil moisture, proved significantly more important for long-term predictions compared to meteorological factors like evaporation. Overall, impact values decrease with longer lead times, reflecting the increased complexity of heatwave prediction over extended periods. Notably, heatwave occurrences within the previous predicted month as an unconventional feature ranked second importance for models with a one-month lead time. This suggests that there may be large potential for unconventional heatwave factors that haven't been considered in the ML-based heatwave prediction so far, such as winter snow accumulation.

From a practical point of view, these insights not only improve understanding of heatwave formation, but also facilitate the structured preparations for such events, guiding interventions well before the heatwave occurs. By identifying key predictive features, such as geopotential height and geographic latitude, along with the significance of land surface characteristics, these models will enable disaster management agencies to develop a tiered, proactive response strategy. For instance, with the example shown in Fig. 7, agencies can predict and monitor a large-scale heatwave four to five months ahead. Between one and three months ahead, agencies can launch broad-scale awareness campaigns with increased certainty, optimise resource allocation, and implement early-stage mitigation strategies for the spatial coverage of the heatwave. These strategies could include adjustments in public transportation schedules to reduce heat exposure and the establishment of specific water-management protocols to handle increased demand during peak heat periods. In the long term, these results can inform future spatial planning, facilitating the design of houses and other buildings that are better equipped to withstand both heat and cold. For example, nature-based solutions (NbS) such as incorporating shaded areas in the summer while allowing sunlight during the winter can be effectively integrated into residential layouts (Pan et al. 2021; Barnett and Bouw 2022). Additionally, including NbS elements like pathways, water bodies, and various types of vegetated areas can further enhance resilience (Sahani et al. 2023; Ibsen et al. 2024). As the lead time shortens, the focus can shift towards fine-tuning emergency response plans, mobilising community support mechanisms and deploying targeted health services to vulnerable populations.

Machine-learning methodologies are progressively employed to enhance the accuracy of heatwave prediction, extending their ability beyond current predictability limits (Domeisen et al. 2023). These predictions are essential for issuing heatwave warnings and for the efficacy of heat-health action plans (Kotharkar and Ghosh 2022). Heatwave warning systems aim to mitigate the adverse impacts of heatwaves and enhance communication among stakeholders. According to the World Health Organization’s heat-health action plan guidelines, the adverse impacts of heatwaves can be significantly reduced through well-coordinated actions at multiple levels (Matthies et al. 2008). These measures include accurate and timely alert systems to provide timely public advisories and a heat-related health information plan about what, to whom and when to communicate (Matthies et al. 2008). In this study, the models were developed to predict the occurrence of heatwaves. For the proposed model to be utilised practically, e.g., in a multi-hazard early-warning system, follow-up research is further needed, e.g., on: (i) an impact-based prediction model and (ii) action road maps.

In an impact-based prediction model, the consequences of the heatwaves are not only related to the hazard itself, but also to the characteristics of the population in the affected area. For example, people with pre-existing health problems, socially isolated elderly people with fragile health conditions, young children, people suffering from obesity, and cardiovascular diseases (Liu et al. 2022) are particularly vulnerable to heatwaves. As an increase in the aging population is expected (SCB 2022), heatwaves are projected to have more severe impacts in the future, not only due to climate change but also as a result of rising urbanization (Hu et al. 2024). It is crucial for authorities to identify and locate vulnerable populations and areas in need of medical intervention during a heatwave within urban environments. By classifying the impact level of heatwaves—such as high-impact or low-impact—and integrating this information into the model used in this study, the accuracy of predicted risks can be significantly enhanced.

In action road maps, communication science is extremely important. Roadmaps can be created to improve societal and organisational preparedness, prevent negative consequences, and ensure the timely response to heatwave-related events and crises in Sweden. The roadmaps could be structured based on systemic mapping of best management practices in the EU and other parts of the world, their adaptability to the Swedish context, and relevant policies and regulations to support societal security in Sweden.

This study has a few limitations that could be improved and addressed in future research. To simplify the modelling process, the threshold heatwave temperature in this study was set to the same value (27°C) throughout Sweden in order to be consistent with the current national heatwave warning system. However, a single threshold value might not be suitable across all parts of Sweden, as heatwave impacts may differ considerably between northern and southern regions. Moreover, the selected explanatory-predictive features investigated in this study were based on a literature review on heatwave prediction with ML approaches, but were limited by data availability in GEE. Future research could integrate other relevant features and test for importance in seasonal heatwave prediction, e.g. cloud cover and the amount of snowfall in winter (Hansen et al. 2014; Dirmeyer et al. 2021). The study also assumed that heatwaves would occur over relatively large regional scales, with related atmospheric features measured at a relatively low spatial resolution compared with surface features. Future studies should assess whether ML model performance improves with a higher spatial resolution in data and more precise data sources, such as local weather stations.

5. Conclusions

This study bridges a research gap between physically based models that commonly enable relatively short-term heatwave prediction, and ML models for possible considerably longer-term prediction and associated earlier warning. The task of early heatwave prediction, i.e., with a seasonal lead time, is challenging due to the complex interactions between different types of variables (atmospheric, land surface and others). This study demonstrates the good ability of the present multi-model ML methodology framework, using open-source remote sensing data from GEE, to enhance the prediction of heatwaves with seasonal lead times and be implemented in, e.g., a multi-hazard early-warning system. Among the five ML models investigated, XGBoost and KNN emerged as best suited for this purpose, while Gaussian NB was found to be unsuitable, with a consistently poor performance compared to the other investigated models.

The novelty of the present study lies in its development and comparison framework of five ML models, and the concrete exemplification of the possible combined application of best-performing models in a multi-model methodology framework that facilitates better heatwave preparedness and mitigation strategies. Feature importance assessed and interpreted using the SHAP approach, and geopotential height shows good long-term predictive power of the best-performing ML models at lead times from one to five months. More generally, the results demonstrate that a multi-disciplinary approach can be effective in addressing complex coupled environmental and societal challenges in future spatial planning.

Declarations

Author Contribution

J.C.K.: Formal analysis, Methodology, Writing – original draft. M.V.P.: Methodology, Writing – review & editing. G.D.: Writing – review & editing. K.B.: Writing – review & editing. C. S.S.F.: Validation, Writing – review & editing. Z.K.: Conceptualization, Supervision, Funding acquisition, Writing – review & editing. All authors reviewed the manuscript.

Acknowledgement

We gratefully acknowledge funding from the Swedish Research Council (VR) through the projects: Science for a secure society: Hydro-climatic hazard, risk, and crisis management in Sweden (CrisAct) (grant 2021-06309); and Coupled freshwater system variations, trends and their drivers around the world (grant 2022-04672). Carla Ferreira was supported by the Portuguese Foundation for Science and Technology, through the institutional scientific employment program-contract (CEECINST/00077/2021). This study arose from a collaborative Master of Science dissertation by Li (2023) entitled 'Monthly Heatwave Prediction in Sweden Based on Machine Learning Techniques with Remote Sensing Data'.

Data Availability

All data were accessed from Google Earth Engine, a publicly available cloud-based platform, via (https://developers.google.com/earth-engine/datasets/) .

References

1. Alam TM, Shaukat K, Hameed IA, et al (2020) An Investigation of Credit Card Default Prediction in the Imbalanced Datasets. IEEE Access 8:201173–201198. https://doi.org/10.1109/ACCESS.2020.3033784
2. Al-Juaidi AEM, Nassar AM, Al-Juaidi OEM (2018) Evaluation of flood susceptibility mapping using logistic regression and GIS conditioning factors. Arab J Geosci 11:765. https://doi.org/10.1007/s12517-018-4095-0
3. Althoff D, Destouni G (2023) Global patterns in water flux partitioning: Irrigated and rainfed agriculture drives asymmetrical flux to vegetation over runoff. One Earth 6:1246–1257. https://doi.org/10.1016/j.oneear.2023.08.002
4. Angelov PP, Soares EA, Jiang R, et al (2021) Explainable artificial intelligence: an analytical review. WIREs Data Min Knowl Discov 11:e1424. https://doi.org/10.1002/widm.1424
5. Asadollah SBHS, Khan N, Sharafati A, et al (2022) Prediction of heat waves using meteorological variables in diverse regions of Iran with advanced machine learning models. Stoch Environ Res Risk Assess 36:1959–1974. https://doi.org/10.1007/s00477-021-02103-z

6. Barnett J, Bouw M (2022) Managing the Climate Crisis: Designing and Building for Floods, Heat, Drought, and Wildfire. Island Press
7. Bartosik A, Whittingham H (2021) Chapter 7 - Evaluating safety and toxicity. In: Ashenden SK (ed) The Era of Artificial Intelligence, Machine Learning, and Data Science in the Pharmaceutical Industry. Academic Press, pp 119–137
8. Bergstra J, Bengio Y (2012) Random Search for Hyper-Parameter Optimization
9. Biau G, Scornet E (2016) A random forest guided tour. TEST 25:197–227. https://doi.org/10.1007/s11749-016-0481-7
10. Brownlee J (2020) Imbalanced Classification with Python: Better Metrics, Balance Skewed Classes, Cost-Sensitive Learning. Machine Learning Mastery
11. Chen T, Guestrin C (2016) XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, New York, NY, USA, pp 785–794
12. Dirmeyer PA, Balsamo G, Blyth EM, et al (2021) Land-Atmosphere Interactions Exacerbated the Drought and Heatwave Over Northern Europe During Summer 2018. AGU Adv 2:e2020AV000283. https://doi.org/10.1029/2020AV000283
13. Domeisen DIV, Eltahir EAB, Fischer EM, et al (2023) Prediction and projection of heatwaves. Nat Rev Earth Environ 4:36–50. https://doi.org/10.1038/s43017-022-00371-z
14. Gorelick N, Hancher M, Dixon M, et al (2017) Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens Environ 202:18–27. https://doi.org/10.1016/j.rse.2017.06.031
15. Gyaneshwar A, Mishra A, Chadha U, et al (2023) A Contemporary Review on Deep Learning Models for Drought Prediction. Sustainability 15:6160. https://doi.org/10.3390/su15076160
16. Hansen BB, Isaksen K, Benestad RE, et al (2014) Warmer and wetter winters: characteristics and implications of an extreme weather event in the High Arctic. Environ Res Lett 9:114021. https://doi.org/10.1088/1748-9326/9/11/114021
17. Hu X, Cao J, Qian Y, et al (2024) Extreme heat events in mainland China from 1981 to 2015: Spatial patterns, temporal trends, and urbanization impacts. Sustain Cities Soc 100:104999. https://doi.org/10.1016/j.scs.2023.104999
18. Ibsen PC, Crawford BR, Corro LM, et al (2024) Urban tree cover provides consistent mitigation of extreme heat in arid but not humid cities. Sustain Cities Soc 113:105677. https://doi.org/10.1016/j.scs.2024.105677
19. Kannan S, Subbaram K, Faiyazuddin Md (2023) Chapter 17 - Artificial intelligence in vaccine development: Significance and challenges ahead. In: Philip A, Shahiwala A, Rashid M, Faiyazuddin Md (eds) A Handbook of Artificial Intelligence in Drug Delivery. Academic Press, pp 467–486
20. Ke D, Takahashi K, Takakura J, et al (2023) Effects of heatwave features on machine-learning-based heat-related ambulance calls prediction models in Japan. Sci Total Environ 873:162283. https://doi.org/10.1016/j.scitotenv.2023.162283

21. Khan N, Shahid S, Ismail TB, Behlil F (2021) Prediction of heat waves over Pakistan using support vector machine algorithm in the context of climate change. Stoch Environ Res Risk Assess 35:1335–1353. https://doi.org/10.1007/s00477-020-01963-1
22. Kong J, Kowalczyk W, Nguyen DA, et al (2019) Hyperparameter Optimisation for Improving Classification under Class Imbalance. In: 2019 IEEE Symposium Series on Computational Intelligence (SSCI). pp 3072–3078
23. Kotharkar R, Ghosh A (2022) Progress in extreme heat management and warning systems: A systematic review of heat-health action plans (1995-2020). Sustain Cities Soc 76:103487. https://doi.org/10.1016/j.scs.2021.103487
24. Li Z (2023) Monthly heatwave prediction in Sweden based on Machine Learning techniques with remote sensing data
25. Lin W-C, Tsai C-F, Hu Y-H, Jhang J-S (2017) Clustering-based undersampling in class-imbalanced data. Inf Sci 409–410:17–26. https://doi.org/10.1016/j.ins.2017.05.008
26. Liu J, Varghese BM, Hansen A, et al (2022) Heat exposure and cardiovascular health outcomes: a systematic review and meta-analysis. Lancet Planet Health 6:e484–e495. https://doi.org/10.1016/S2542-5196(22)00117-6
27. Lundberg SM, Lee S-I (2017) A Unified Approach to Interpreting Model Predictions. In: Advances in Neural Information Processing Systems. Curran Associates, Inc.
28. Luo H, Pan X, Wang Q, et al (2019) Logistic Regression and Random Forest for Effective Imbalanced Classification. In: 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC). pp 916–917
29. Mamgain S, Gupta K, Roy A, et al (2023) LONG-TERM THERMAL ANOMALY DETECTION AND MAPPING AT PIXEL LEVEL USING A GOOGLE EARTH ENGINE TOOL. Int Arch Photogramm Remote Sens Spat Inf Sci XLVIII-M-3–2023:147–153. https://doi.org/10.5194/isprs-archives-XLVIII-M-3-2023-147-2023
30. Matthies F, Bickler G, Marín NC, Hales S (2008) Heat-health Action Plans: Guidance. World Health Organization
31. Merz B, Kuhlicke C, Kunz M, et al (2020) Impact Forecasting to Support Emergency Management of Natural Hazards. Rev Geophys 58:e2020RG000704. https://doi.org/10.1029/2020RG000704
32. Mosavi A, Ozturk P, Chau K (2018) Flood Prediction Using Machine Learning Models: Literature Review. Water 10:1536. https://doi.org/10.3390/w10111536
33. Niaz R, Zhang X, Iqbal N, et al (2021) Logistic Regression Analysis for Spatial Patterns of Drought Persistence. Complexity 2021:e3724919. https://doi.org/10.1155/2021/3724919
34. Niggli L, Huggel C, Muccione V, et al (2022) Towards improved understanding of cascading and interconnected risks from concurrent weather extremes: Analysis of historical heat and drought extreme events. PLOS Clim 1:e0000057. https://doi.org/10.1371/journal.pclm.0000057
35. Oommen T, Baise LG, Vogel RM (2011) Sampling Bias and Class Imbalance in Maximum-likelihood Logistic Regression. Math Geosci 43:99–120. https://doi.org/10.1007/s11004-010-9311-8

36. Oudin Åström D, Åström C, Forsberg B, et al (2020) Heat wave–related mortality in Sweden: A case-crossover study investigating effect modification by neighbourhood deprivation. Scand J Public Health 48:428–435. https://doi.org/10.1177/1403494818801615
37. Pan H, Page J, Cong C, et al (2021) How ecosystems services drive urban growth: Integrating nature-based solutions. Anthropocene 35:100297. https://doi.org/10.1016/j.ancene.2021.100297
38. Panahi M, Rahmati O, Kalantari Z, et al (2022) Large-scale dynamic flood monitoring in an arid-zone floodplain using SAR data and hybrid machine-learning models. J Hydrol 611:128001. https://doi.org/10.1016/j.jhydrol.2022.128001
39. Perkins-Kirkpatrick SE, Lewis SC (2020) Increasing trends in regional heatwaves. Nat Commun 11:3357. https://doi.org/10.1038/s41467-020-16970-7
40. Rahmati O, Falah F, Dayal KS, et al (2020) Machine learning approaches for spatial modeling of agricultural droughts in the south-east region of Queensland Australia. Sci Total Environ 699:134230. https://doi.org/10.1016/j.scitotenv.2019.134230
41. Raja A, .T G (2022) Drought Prediction and Validation for Desert Region using Machine Learning Methods. Int J Adv Comput Sci Appl 13:47–53. https://doi.org/10.14569/IJACSA.2022.0130707
42. Robine J-M, Cheung SLK, Le Roy S, et al (2008) Death toll exceeded 70,000 in Europe during the summer of 2003. C R Biol 331:171–178. https://doi.org/10.1016/j.crvi.2007.12.001
43. Rodrigues M, Santana P, Rocha A (2021) Modelling of Temperature-Attributable Mortality among the Elderly in Lisbon Metropolitan Area, Portugal: A Contribution to Local Strategy for Effective Prevention Plans. J Urban Health Bull N Y Acad Med 98:516–531. https://doi.org/10.1007/s11524-021-00536-z
44. Rosales-Pérez A, García S, Herrera F (2023) Handling Imbalanced Classification Problems With Support Vector Machines via Evolutionary Bilevel Optimization. IEEE Trans Cybern 53:4735–4747. https://doi.org/10.1109/TCYB.2022.3163974
45. Rousi E, Kornhuber K, Beobide-Arsuaga G, et al (2022) Accelerated western European heatwave trends linked to more-persistent double jets over Eurasia. Nat Commun 13:3851. https://doi.org/10.1038/s41467-022-31432-y
46. Rudin C (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 1:206–215. https://doi.org/10.1038/s42256-019-0048-x
47. Sahani J, Kumar P, Debele SE (2023) Efficacy assessment of green-blue nature-based solutions against environmental heat mitigation. Environ Int 179:108187. https://doi.org/10.1016/j.envint.2023.108187
48. S.Badhiye S, U. Sambhe N, N. Chatur P (2013) KNN Technique for Analysis and Prediction of Temperature and Humidity Data. Int J Comput Appl 61:7–13. https://doi.org/10.5120/9994-4847
49. SCB (2022) After age 60. A description of older people in Sweden
50. Singh D, Singh B (2020) Investigating the impact of data normalization on classification performance. Appl Soft Comput 97:105524. https://doi.org/10.1016/j.asoc.2019.105524

51. Sjulgård H, Keller T, Garland G, Colombi T (2023) Relationships between weather and yield anomalies vary with crop type and latitude in Sweden. Agric Syst 211:103757. https://doi.org/10.1016/j.agsy.2023.103757
52. Sköld Gustafsson V, Andersson Granberg T, Pilemalm S, Waldemarsson M (2023a) Identifying decision support needs for emergency response to multiple natural hazards: an activity theory approach. Nat Hazards. https://doi.org/10.1007/s11069-023-06305-2
53. Sköld Gustafsson V, Hjerpe M, Strandberg G (2023b) Construction of a national natural hazard interaction framework: The case of Sweden. iScience 26:106501. https://doi.org/10.1016/j.isci.2023.106501
54. SMHI (2020) Health effects of heat stress. https://www.smhi.se/en/research/research-departments/meteorology/heat-and-air-quality-in-cities/health-effects-of-heat-stress-1.167391. Accessed 15 Feb 2024
55. SMHI (2023) Värmebölja. https://www.smhi.se/kunskapsbanken/meteorologi/temperatur/varmebolja-1.22372. Accessed 15 Feb 2024
56. Soares PR, Harrison MT, Kalantari Z, et al (2023) Drought effects on soil organic carbon under different agricultural systems. Environ Res Commun 5:112001. https://doi.org/10.1088/2515-7620/ad04f5
57. Straaten C van, Whan K, Coumou D, et al (2022) Using Explainable Machine Learning Forecasts to Discover Subseasonal Drivers of High Summer Temperatures in Western and Central Europe. Mon Weather Rev 150:1115–1134. https://doi.org/10.1175/MWR-D-21-0201.1
58. SWECO Building heatwave resilience in European cities
59. Tang J, Alelyani S, Liu H (2014) Feature Selection for Classification: A Review
60. Tsai C-F, Lin W-C, Hu Y-H, Yao G-T (2019) Under-sampling class imbalanced datasets by combining clustering analysis and instance selection. Inf Sci 477:47–54. https://doi.org/10.1016/j.ins.2018.10.029
61. United Nations Environment Programme (2003) Impacts of Summer 2003 Heat Wave in Europe - Environment Alert Bulletin 2. https://wedocs.unep.org/xmlui/handle/20.500.11822/40942. Accessed 15 Feb 2024
62. Vieira Passos M, Kan J-C, Destouni G, et al (2024) Identifying regional hotspots of heatwaves, droughts, floods, and their co-occurrences. Stoch Environ Res Risk Assess. https://doi.org/10.1007/s00477-024-02783-3
63. Vu A, Rutherford S, Phung D (2019) Heat Health Prevention Measures and Adaptation in Older Populations—A Systematic Review. Int J Environ Res Public Health 16:4370. https://doi.org/10.3390/ijerph16224370
64. Wang H, Wang H, Wu Z, Zhou Y (2021) Using Multi-Factor Analysis to Predict Urban Flood Depth Based on Naive Bayes. Water 13:432. https://doi.org/10.3390/w13040432

65. Weirich-Benet E, Pyrina M, Jiménez-Esteve B, et al (2023) Subseasonal Prediction of Central European Summer Heatwaves with Linear and Random Forest Machine Learning Models. Artif Intell Earth Syst 2:. https://doi.org/10.1175/AIES-D-22-0038.1
66. Wu X, Kumar V, Ross Quinlan J, et al (2008) Top 10 algorithms in data mining. Knowl Inf Syst 14:1–37. https://doi.org/10.1007/s10115-007-0114-2
67. Xu Z, FitzGerald G, Guo Y, et al (2016) Impact of heatwave on mortality under different heatwave definitions: A systematic review and meta-analysis. Environ Int 89–90:193–203. https://doi.org/10.1016/j.envint.2016.02.007
68. Zhang C, Wang G, Zhou Y, Jiang J (2017) A New Approach for Imbalanced Data Classification Based on Minimize Loss Learning. In: 2017 IEEE Second International Conference on Data Science in Cyberspace (DSC). pp 82–87
69. Zhang M, Yang X, Cleverly J, et al (2022) Heat wave tracker: A multi-method, multi-source heat wave measurement toolkit based on Google Earth Engine. Environ Model Softw 147:105255. https://doi.org/10.1016/j.envsoft.2021.105255
70. Zheng Z, Zhao L, Oleson KW (2021) Large model structural uncertainty in global projections of urban heat waves. Nat Commun 12:3736. https://doi.org/10.1038/s41467-021-24113-9
71. Zou Q, Xie S, Lin Z, et al (2016) Finding the Best Classification Threshold in Imbalanced Classification. Big Data Res 5:2–8. https://doi.org/10.1016/j.bdr.2015.12.001

Figures

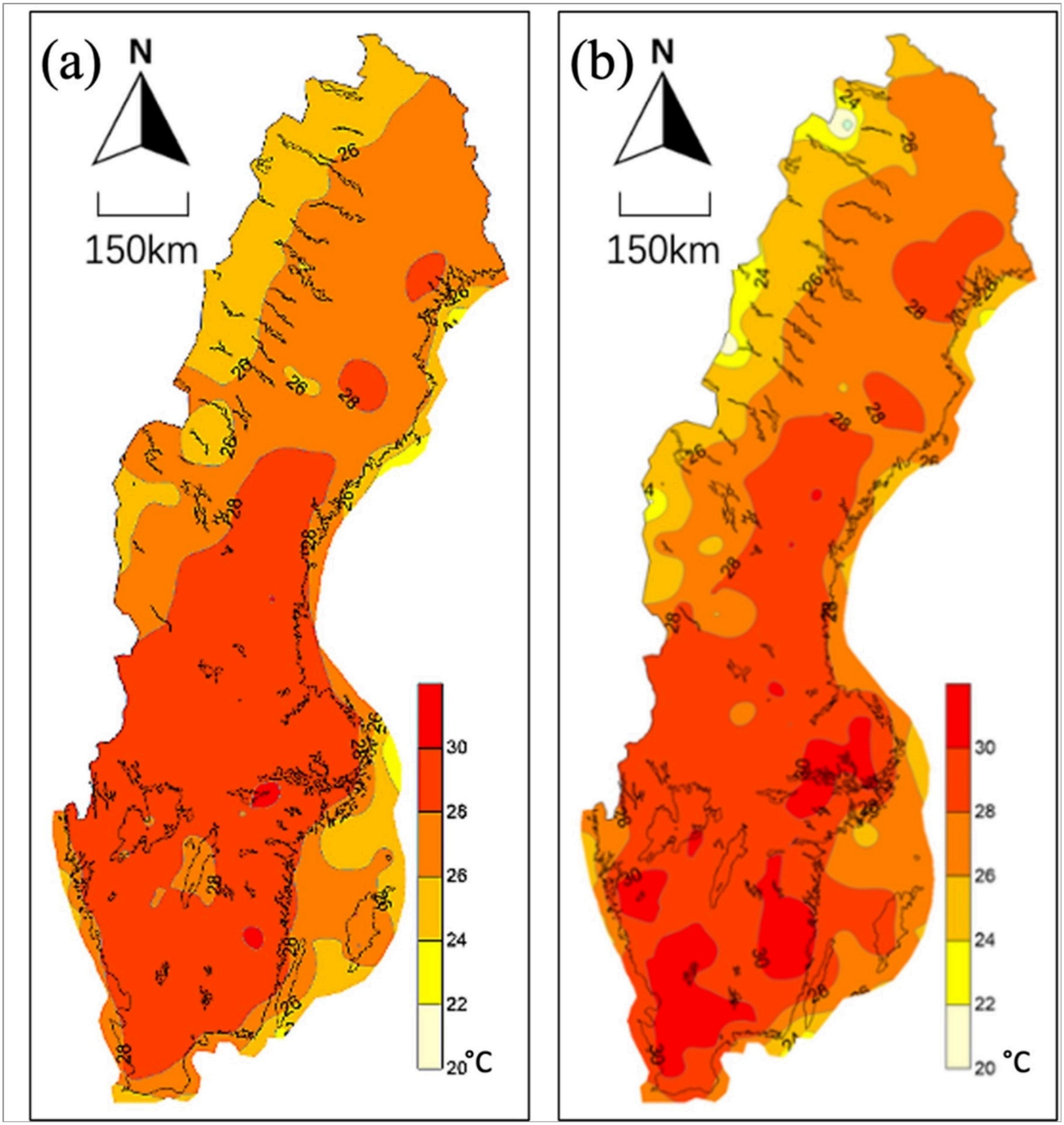

Figure 1

Maps of Sweden showing the average value of the years’ highest temperatures during the periods (a) 1961-1990; (b) 1991-2020 (adapted from SMHI, 2023).

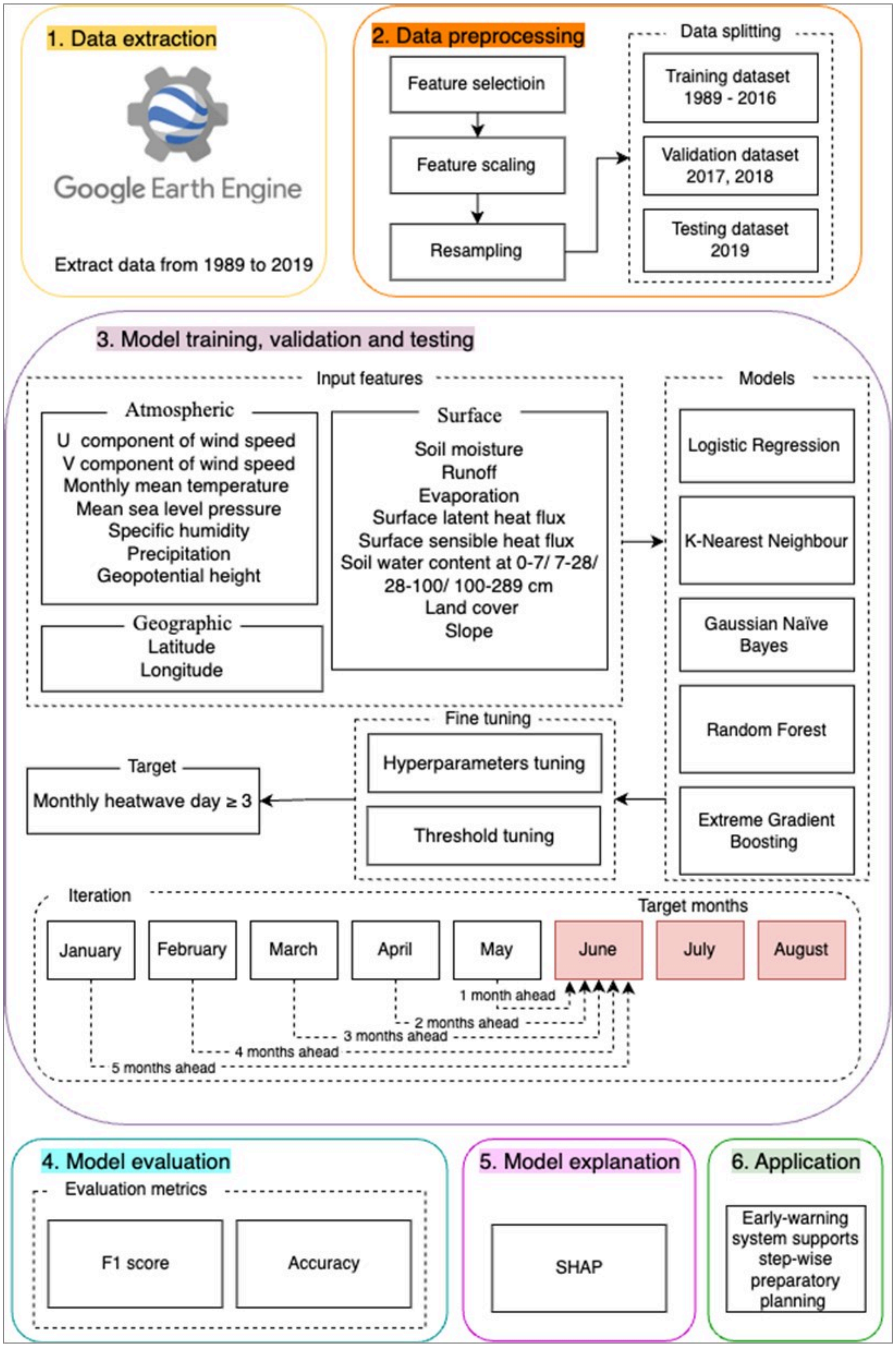

Figure 2

Flowchart of the framework developed and used for heatwave prediction in Sweden.

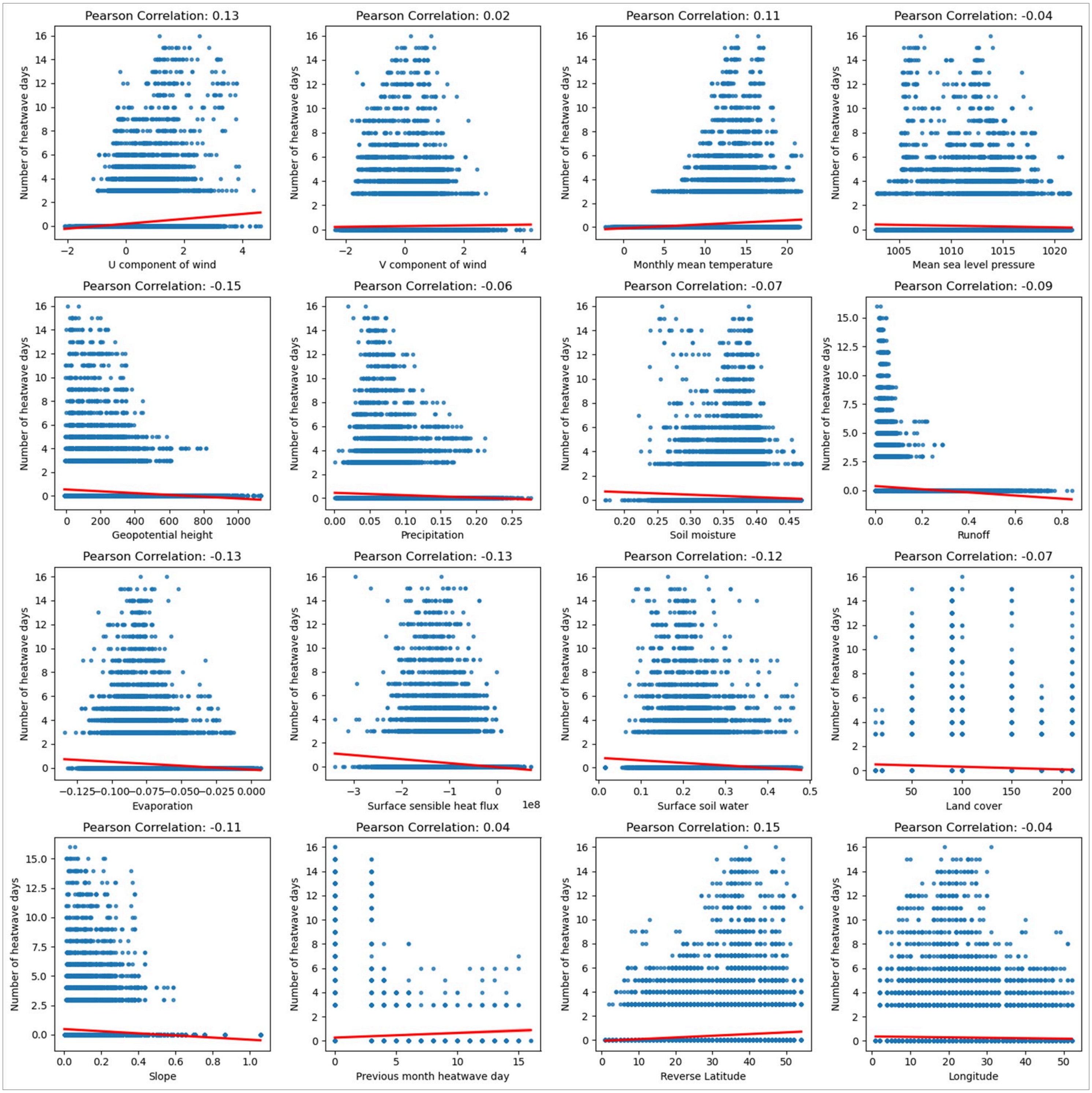

Figure 3

Scatterplots of selected variables and the number of heatwave days in the target months (June, July and August) between 1989 and 2019. Low R values suggest that there is no linear relationship between the features and target. The red colour represents the regression line.

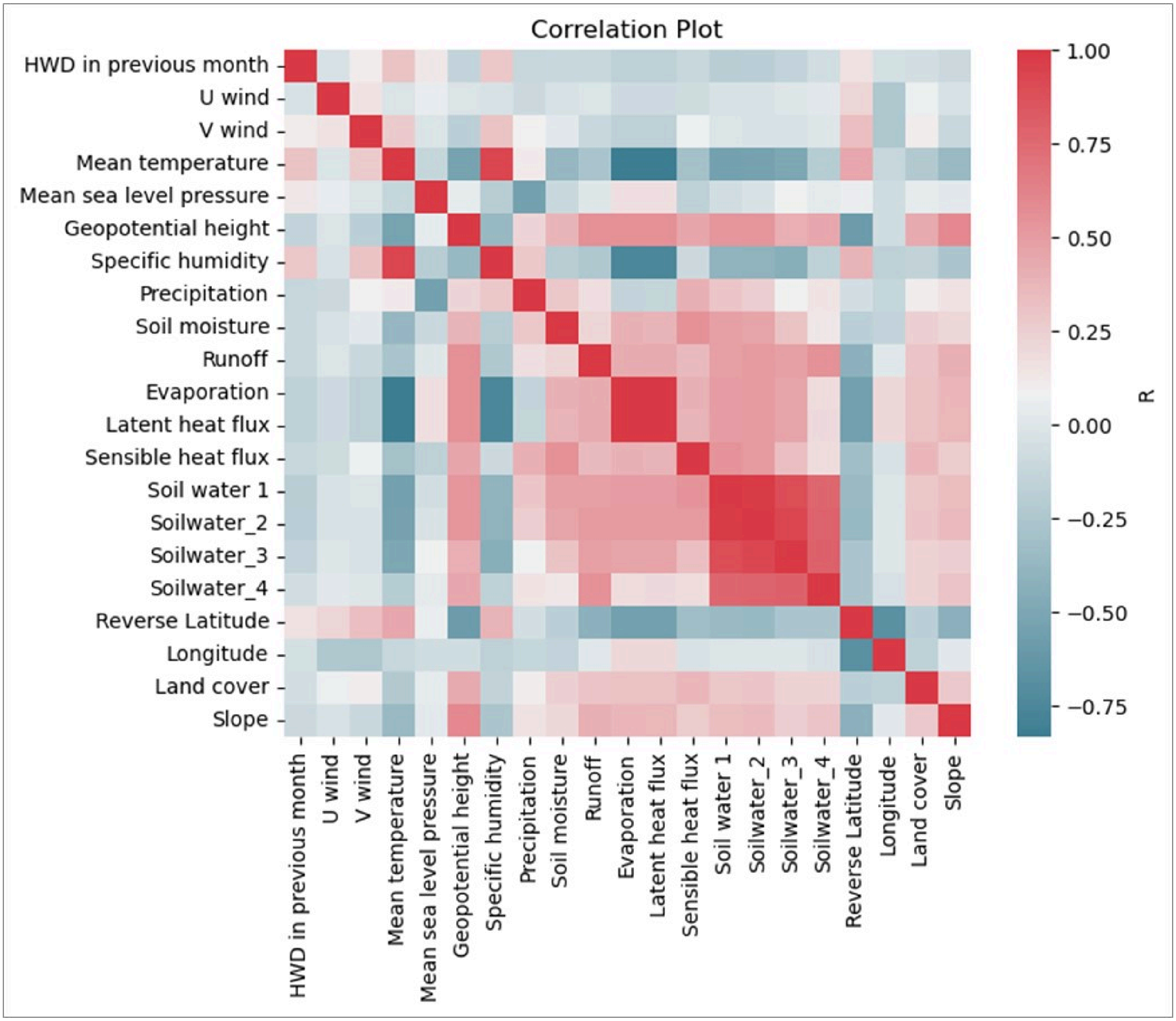

Figure 4

Correlation matrix heatmap of possible relevant explanatory-predictive features for heatwave prediction investigated in this study. Red indicates features with strong positive correlations, white no correlation, and blue negative correlations. See Table 1 for feature descriptions.

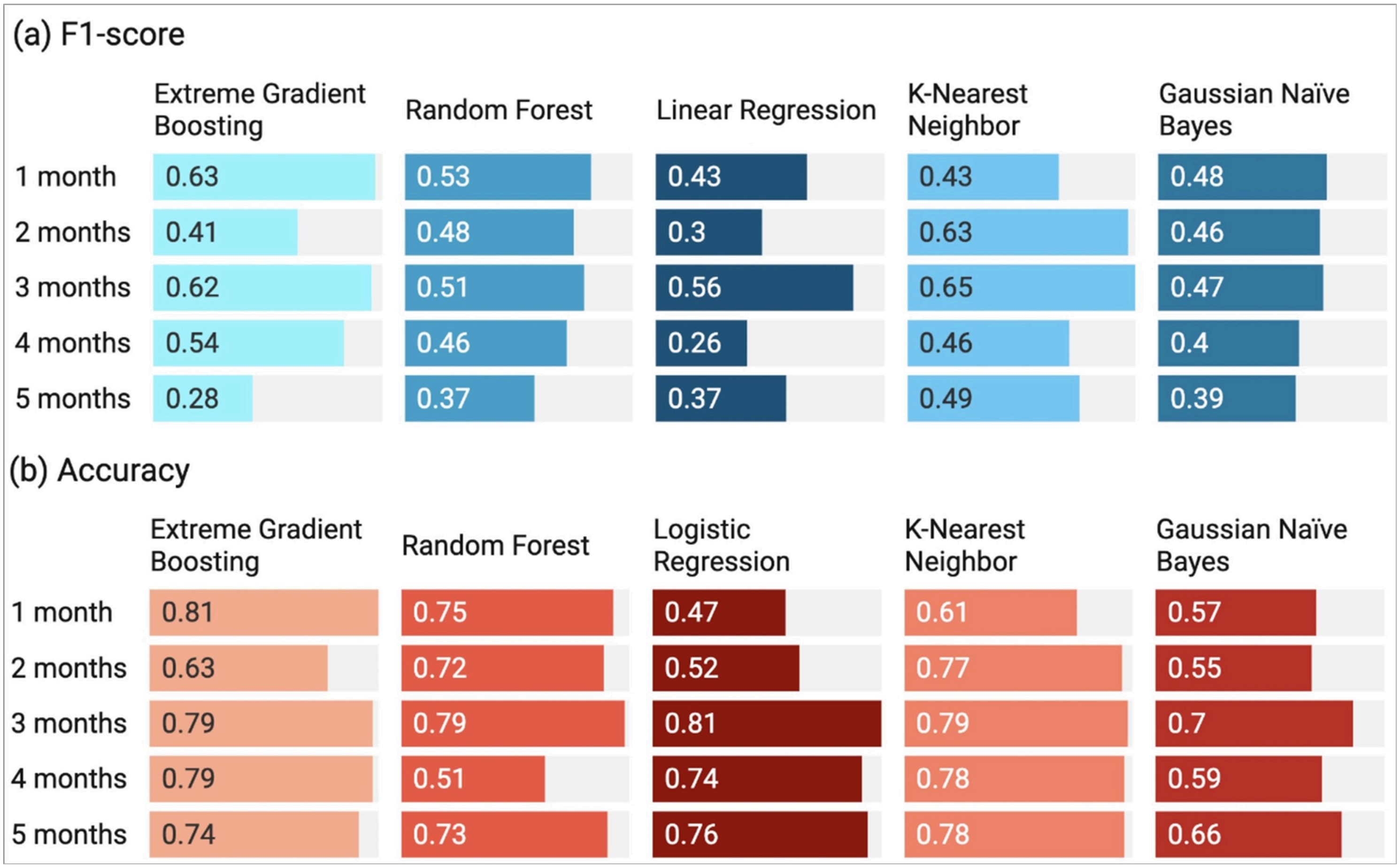

Figure 5

Performance of the five machine-learning models in the prediction of heatwave days in Sweden with lead times of one to five months: (a) F1-score and (b) accuracy

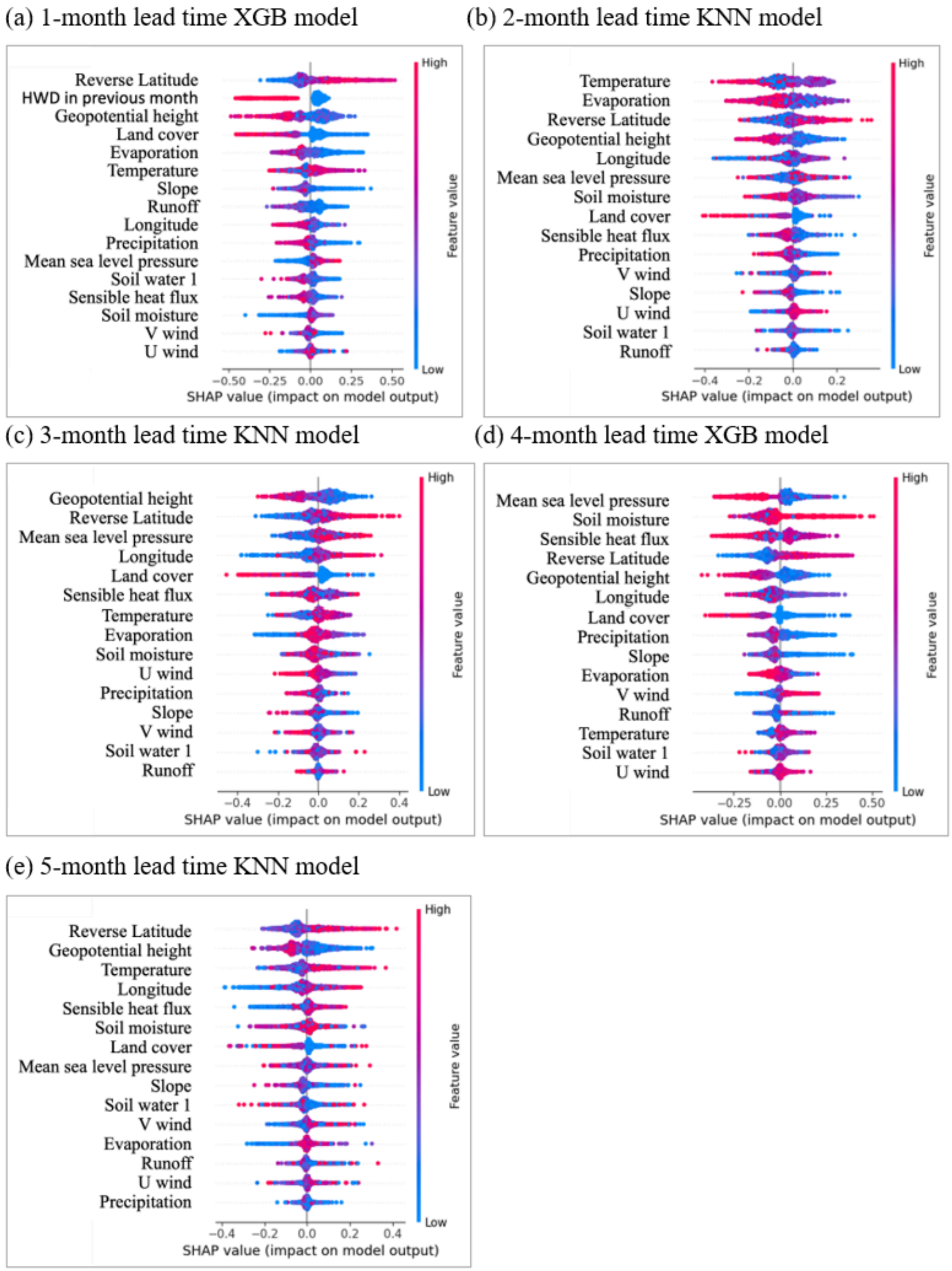

Figure 6

SHapley Additive exPlanations (SHAP) values for each feature in the best-performing models in prediction with lead times of one to five months, where blue represents lower and red represents higher feature values, and positive and negative SHAP values (x-axis) imply a higher and lower probability of heatwave occurrence, respectively.

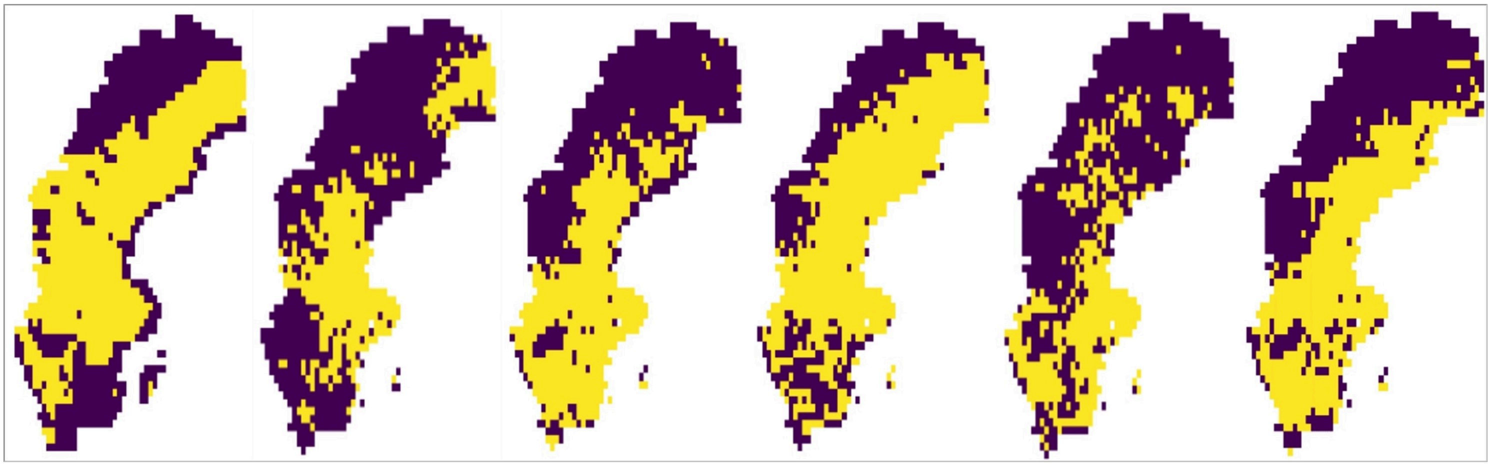

Figure 7

Example of heatwave occurrence observed in July 2019 (a) and associated predictions at 1-month (b) lead time with XGB model, at 2- (c) and 3-months (d) lead time with KNN model, 4-month (e) lead time using XGB model and 5-month (f) lead time based on KNN model. The yellow area indicates the presence of a heatwave, while the purple area indicates regions without a heatwave.