Machine learning models to predict myocardial infarctions from past climatic and environmental conditions

Lennart Marien[1], Mahyar Valizadeh[2,*], Wolfgang zu Castell[3], Christine Nam[1], Diana Rechid[1], Alexandra Schneider[4], Christine Meisinger[5,6], Jakob Linseisen[5,6], Kathrin Wolf[4,**], and Laurens M Bouwer[1,*,**]

[1]Climate Service Center Germany (GERICS), Helmholtz-Zentrum Hereon, Fischertwiete 1, 20095, Hamburg, Germany
[2]Strategy and Digitalization, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstraße 1, 85764, Neuherberg
[3]Department Geoinformation, Helmholtz Centre Potsdam - GFZ German Research Centre for Geosciences, Telegrafenberg, 14473, Potsdam, Germany
[4]Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstraße 1, 85764, Neuherberg, Germany
[5]Chair of Epidemiology, University of Augsburg, University Hospital Augsburg, Stenglinstr. 2, 86156 Augsburg
[6]Clinical Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstr. 1, 85754 Neuherberg
[*]These authors contributed equally to this work.
[**]These authors share last authorship.
Correspondence: Laurens M. Bouwer (laurens.bouwer@hereon.de)

Abstract. Myocardial infarctions (MI) are a major cause of death worldwide, and both high and low temperatures (i.e., heat and cold) may increase the risk of MI. The relationship between health impacts and climate is complex and influenced by a multitude of climatic, environmental, socio-demographic, and behavioral factors. Here, we present a Machine Learning (ML) approach for predicting MI events based on multiple environmental and demographic variables. We derived data on MI events from the KORA MI registry dataset for Augsburg, Germany between 1998 and 2015. Multivariable predictors include weather and climate, air pollution (PM_{10}, NO, NO_2, SO_2, and O_3), surrounding vegetation, as well as demographic data. We tested the following ML regression algorithms: Decision Tree, Random Forest, Multi-layer Perceptron, Gradient Boosting and Ridge Regression. The models are able to predict the total annual number of MI reasonably well (adjusted $R^2 = 0.62$–0.71). Inter-annual variations and long-term trends are captured. Across models the most important predictors are air pollution and daily temperatures. Variables not related to environmental conditions, such as demographics need to be considered as well. This ML approach provides a promising basis to model future MI under changing environmental conditions, as projected by scenarios for climate and other environmental changes.

1 Introduction

Myocardial infarctions (MI), are a major cause of cardiovascular related mortality and morbidity. The estimated prevalence of MI worldwide in 2015 was close to 16 million, with 33,000 years lived with disability attributed to the condition (Vos et al., 2016). In light of ageing western societies as well as ongoing environmental and climatic changes, which have been identified

as important risk factors, MI is likely to remain a considerable burden to health systems in the future (e.g., Khraishah et al., 2022). It is therefore paramount to deepen the understanding of the complex interplay between environmental and other risk factors and their effect on MI, and to estimate their expected future development.

20   Epidemiological research has shown that both high and low air temperatures (i.e., extreme cold and heat) temperature extremes can play an important role in triggering acute MI (Chen et al., 2019; Wolf et al., 2009; Sun et al., 2018). This is especially apparent in winter when most of the MI events are observed. Most previous studies (e.g. with registry data) have reported significant cold effects on MI occurrence (e.g., The Eurowinter Group, 1997; Schwartz et al., 2004; Wolf et al., 2009; Bhaskaran et al., 2010) whereas fewer studies have observed increased risk of MI triggered by heat exposures so far (e.g., 25   Bhaskaran et al., 2012; Madrigano et al., 2013; Chen et al., 2019). Severe periods of heat as encountered during heat waves are likely to occur with higher frequency, intensity, and duration due to anthropogenic climate change, even if limited to warming levels between 1.5° and 2° (Sieck et al., 2020). Increasing levels of urbanisation entail higher levels of exposure to heat as well, due to the urban heat island effect (e.g., Feng et al., 2014; Zhang et al., 2009). Air pollution is another environmental factor known to potentially trigger MI after periods of intense short-term exposure (e.g., Peters et al., 2004; Mustafić et al., 2012) but 30   also to increase the risk in association with elevated long-term exposure (Cesaroni et al., 2014; Wolf et al., 2021; Rajagopalan et al., 2018). Moreover, the elderly are particularly vulnerable to MI, exacerbating the potential adverse effects in light of the demographic ageing expected in developed countries, such as in Germany (Schmidt et al., 2013; Rai et al., 2019).

A key issue in understanding current and future health impacts is the inclusion of a multitude of processes and circumstances that influence the health outcomes (Roth, 2020), in quantitative models. For MI, these include the occurrence of high and 35   low temperature events, air quality, the presence of water bodies and vegetation and characteristics of the built environment. Although the relevance of humidity for MI has not been confirmed (e.g., Schwartz et al., 2004), it is often included when studying human health impacts (Davis et al., 2016). For instance, high temperatures are often perceived as more stressful under very humid conditions. Hot and strongly saturated air carries less oxygen and interferes with transpiration as main mechanism of cooling the human body (Havenith, 2005). Therefore, the same temperature can be perceived more straining if humidity is 40   high as well. Changes in the exposed population, such as their age, their health status and underlying diseases are important as well. Therefore, future health risks from climate change cannot only be estimated from changes in (extreme) weather, but it is critically important to account also for all these other relevant factors (Vanos et al., 2020). Finally, health interventions such as heat health action plans and improved healthcare have been shown to reduce health risks from extreme temperatures (see for instance Achebak et al., 2019). But also policies related to climate change, such as reduced traffic emissions, are expected to 45   lead to a reduction in disease burden (Laverty et al., 2021).

For more reliable estimations of potential future risks, multiple variables must be incorporated into prediction models. In addition, several of the relations between environmental and other factors, and health outcome are only partially known. This is where data-driven approaches are particularly useful, as they can provide accurate estimations of complex processes, taking up many variables and also account for complex and non-linear relations. Machine Learning (ML) approaches are now being 50   tested widely for environmental studies (Reichstein et al., 2019), and they are also increasingly used to estimate social and economic impacts of environmental extremes such as floods and windstorms (Merz et al., 2013; Wagenaar et al., 2017, 2021).

ML however, has only recently been applied to health impact modeling. Several studies have employed statistical methods as well as ML to predict infectious diseases, such as malaria transmissions (Zinszer et al., 2012; Sewe et al., 2017). Zhang et al. (2014) studied heat-related mortality, and identified relevant temperature and humidity variables using Random Forests. Other studies applied ML to evaluate the risk for MI, or to predict acute MI based on data such as patient history, blood markers, or electrocardiogram, but lack an environmental dimension (e.g., Tamarappoo et al., 2021; Commandeur et al., 2020).

In this study we employ several ML algorithms in a data-driven setting, using a range of meteorological, environmental, demographic and health variables on preceding days. We estimate the importance of the predictive variables in the models. We also assess the effects on different sub-groups, depending on location (urban/rural) as these may exhibit different vulnerabilities (Gabriel and Endlicher, 2011), and patient characteristics (age, smoking, and diabetes). The ML models that are presented can be used to estimate future health outcomes, using a set of scenarios for changes in climatic, environmental and demographic variables. Instead of using an approach based on time series modelling (see e.g., Armstrong, 2006; Chen et al., 2018), we employ multivariate ML regression models. These models do not require the presupposition of a known exposure-response relationship. Also, our study is aimed towards developing models to make long-term projections at climate-timescales (30 years). At such timescales underlying statistical properties may change gradually which would not be reflected by any pre-scribed exposure-response function based on historic or current data. Contrary to other studies, we also do not account for seasonal effects. Instead, we solely rely on a data-driven approach in which we make no a priori assumptions about the relationship between features and the health outcome. While this does not allow for an explicit decomposition of the time series into, e.g., trend, seasonality and random effects, it might generalize better when applied to an ensemble of climate simulations in which the statistics of the features may have changed drastically compared to the historical training data.

We expect that none of the risk factors that are included in our models is strong enough to directly trigger MI in an otherwise healthy person. Instead, these environmental and demographic factors must be assumed to increase the statistical likelihood of vulnerability to MI over longer periods of time. Many of the risk factors that we cover in this study can modify this individual likelihood of suffering from MI. In light of this, we do not expect for the models to be able to accurately provide predictions on a daily basis. However, our research motivation is to eventually estimate the long-term tendencies in MI due to climate change. We therefore decided to aggregate our model results on an annual basis. This should allow for some of the inherent randomness to average out and allow a more statistical view on MI occurrence over annual and interannual timescales.

In Section 2, we present the methods used to develop the ML models. In Section 3, we describe the input data for our data-driven approach. In Section 4 the results of the simulations and their performance are given. In Section 5 we discuss the results and give an outlook for using the models to project future MI events, and finally in Section 6 we provide the conclusions.

2 Methods

In this section, we present the approach to modelling the occurrence of MI events from a large variety of data and discuss the ML methods that were applied. We also consider correlations among the features and describe how we selected suitable parameters for the ML algorithms.

2.1 A supervised learning problem for MI events

ML models can comprise of classification or regression based algorithms. In this study, we focus solely on regression methods. The registry data is case-only, i.e., by design each participant is bound to have an MI.

The target variable in our case is the time series of daily events of MI observed in the study region. In addition, the co-occurring environmental variables that have a plausible causal relation to this target variable are collected and used as predictors in the training process. We use the scikit-learn package for performing the calculations (see Pedregosa et al., 2011; Pedregosa, F. et al.). The figures use colors chosen with disability-friendliness in mind (Crameri et al., 2020).

For any given day d let y_d be the number of MI events and $x_{i,d}$ the value of the i-th predictive variable on that day (e.g., daily maximum temperature or daily mean PM_{10}). To work with standard regression algorithms, a fixed number of features must be selected and together with the target value y_d be provided as training input. The variables $x_{i,d}$ represent a time series and therefore only a subset of them should be selected as a feature of the regression problem, namely the conditions on the day of prediction. Past conditions, however, might also have an influence on current events, both long and short term. The sliding window method allows for this by selecting the features with a lag n, referred to as the window size. The merits of allowing for shorter or longer memories are difficult to estimate. For instance, the effects of extremely high temperatures on MI are generally expected to be short-term (Breitner et al., 2014), ranging from immediate effects to up to three weeks lag. The vector of features, i.e., the training (or test) instance on day d, is then given as:

$$x_d = (x_{1,d-n+1}, x_{1,d-n+2}, \ldots, x_{1,d}, x_{2,d-n+1}, \ldots, x_{2,d}, x_{m,d-n+1}, \ldots, x_{m,d})$$

where n is the windows size and m the number of variables. Each predictive variable then yields n features and the total number of features for this problem is $n \cdot m$. Accumulating the x_d and y_d for all days into a matrix X and a vector y yields input that can directly be used with the scikit-learn regression algorithms. We applied the five ML methods and associated scikit-learn classes, listed in Table 1 with their abbreviations as used in the remainder of this paper. Note that some features such as the slowly changing demographic variables, were not subject to the sliding window and instead simply used the value on the day of prediction. For this study, after testing different lags between 1 and 21 days, we exclusively used a lag of $n = 3$ days as this resulted in the best overall scores. However, in order to account for possibly longer lasting (see Sun et al., 2018) cold effects, we added a predictor using the 21-day rolling mean of the minimum temperature.

Note that throughout this paper, we use the terms predictor and feature in an interchangeable manner, namely to refer to the features of the supervised learning problem derived above: the vector X and its components.

We also added a random feature to be able to use its importance as a benchmark. Predictors less important than the random feature can be assumed to be irrelevant. Finally, we added three time variables, namely the day of the week, the day of the year and the current month.

Regression Method	Abbreviation	scikit-learn Class	Version
Decision Tree	DTR	sklearn.tree.DecisionTreeRegressor	0.23.2
Random Forest	RF	sklearn.ensemble.RandomForestRegressor	0.23.2
Gradient Boosting	GBR	sklearn.ensemble.GradientBoostingRegressor	0.23.2
Ridge Regression	RR	sklearn.linear_model.Ridge	0.23.2
Multi-layer Perceptron	MLP	sklearn.neural_network.MLPRegressor	0.23.2

Table 1. Regression methods used and associated scikit-learn classes.

2.2 Scaling and random split

Different magnitudes of the features can have adverse effects as the results could be biased towards those variables given in nominally large units relative to others. To avoid this, we apply the sklearn.preprocessing.StandardScaler class to the input, resulting in features that are centered around 0 with unit variance. Second, we withhold parts of the data from the training to have independent data instances for validation. We apply sklearn.model_selection.train_test_split with shuffle, resulting in a random 75%/25% split of the data in training and test portions. The 25% of data not used for training the algorithms are used for validation. Splitting the data randomly means that the underlying time series lose their natural temporal order. This has implications when visualising and interpreting model results that we will cover in a later section, but it reduces the likelihood of autocorrelations (e.g., seasonal signals) present in the time series that could result in overoptimistic predictions. In order to split the data randomly, the random number generator has to be initialised with a seed. We found that different random seeds can result in significantly different results. To avoid reporting results that are strongly dependent on the chosen seed, we repeated all calculations with 100 randomly selected seeds. The result with the R^2-score closest to the average score of the ensemble was then selected as a representative example of model capability. Moreover, as the dependency on the random seed is likely related to unbalanced splits, we employed a simple stratification strategy. The data is stratified along the number of MI occurrences observed, i.e., data points with the same number of MI are split among test and training in a representative way. This is especially important for rare events, such as 5 or more MI in one day. The dependency on the random seed was substantially reduced in this way, but significant differences between different seeds could still be observed.

2.3 Feature Importance

It is useful to evaluate the relative importance of different features, i.e., to measure the contribution a given feature makes to the overall prediction. In this study, we use the built-in variable importance capabilities provided by scikit-learn package, yielding a number between 0 and 1 for each feature. The sum of all individual contributions is always equal to 1. For RR we simply relate the magnitude of the trained weights (coefficients) of the model to their associated predictors. Here, care must be taken to consider the relative magnitudes of the predictors, but this has been addressed in our study by scaling the input data. For DT, RF and GBR the importance is based on the normalized total impurity decrease, i.e., a measure of the quality of splits associated with a given feature, aggregated across the whole tree or the ensemble of trees respectively. For MLP no

variable importance is provided by scikit-learn and we therefore constrained this part of the analysis to the four aforementioned algorithms.

2.4 Feature reduction

Correlated features can lead to an overemphasis of their influence on the target variable. This can be counteracted by choosing only one of the correlated features, usually the one that has the strongest correlation with the target variable. In our case, we aimed to include as many variables as possible that could reasonably have an effect on MI. The downside is that some features, for instance maximum, minimum, and mean temperature, are highly correlated on a daily basis. A visualisation of the correlation between the predictors used in this study is shown in appendix Figure A8. To address this issue, we tested the option of transforming the data to a smaller feature space using principal component analysis (PCA). The resulting principal components are uncorrelated to each other and the risk of introducing spurious or overly strong relationships into the training data is reduced while retaining most of the original information. We used sklearn.decomposition.PCA and opted to retain at least 98% of the variance. Having the principal components as optional features allowed us to compare predictions with PCA to estimate the potential adverse effects of correlations present in our data. The results using the PCA data (not shown here) did not improve, suggesting that using the original set of features does not introduce spurious relations. Moreover, using PCA leads to a reduction of interpretability, as the principal components are linear combinations of the original features, without a clear relation to the original variables.

2.5 Hyperparameter optimisation

The ability of the ML algorithms listed in Table 1 to produce accurate predictions is dependent on the selection of appropriate hyperparameters. These parameters generally control specific aspects of the underlying methods, such as the maximum depth of a decision tree, the number of neurons in a layer, or the strength of regularisation. With regularisation, a penalty is added as model complexity increases, which helps to avoid overfitting. In this study, we used the sklearn.model_selection.GridSearchCV class to optimise hyperparameters over predefined parameter spaces with 5-fold cross validation. We used the adjusted R^2 as the governing score to make decisions on optimal parameters. The parameter set with the best overall score is selected. Using cross-validation allows to produce more robust generalisation error estimates without having to reserve a dedicated cross-validation set that would not be available for training. Moreover, by using folds based only on 75% of the training data, no information from the remaining 25% data is used for optimising the models and validation through parameter selection.

Due to substantial computational expense, we only optimised over rather sparse parameter spaces and a limited number of the available parameters. Table 2 shows a list of the selected hyperparameters for all the methods used as well as their optimised values. To speed up the calculations we used the Intel® extension package for scikit-learn, called scikit-learn-intelex.

Hyperparameters	random_state	max_depth	max_features	min_sample_leaf	min_sample_split	n_estimators	alpha	learning_rate	max_iter	activation	parallel
DTR	**429**	**10**	**log2**	**64**	**2**	-	-	-	-	-	no
RF	**371**	**5**	**log2**	**2**	**5**	**100**	-	-	-	-	yes
GBR	**357**	**3**	**log2**	**64**	**2**	**100**	-	-	-	-	no
RR	**0**	-	-	-	-	-	**5000.0**	-	-	-	yes
MLP	**419**	-	-	-	-	-	**5.0**	**constant**	**2500**	**logistic**	yes

Table 2. Optimised hyperparmeters marked with grey cells

3 Data

The dataset used in this study is highly heterogeneous along many dimensions, with differences ranging from file format, metadata conventions, spatial coverage (e.g., regional, local) and resolution, to temporal frequency (e.g., daily, monthly, annual) and representation (e.g., raster, polygon and point data). In this section, we give an overview of the data used in this study and describe the workflow applied to homogenise and prepare these. Table 3 lists all environmental and demographic predictive variables that were used for this study in addition to the MI data, as well as the source datasets and associated references.

3.1 KORA MI registry

The health dataset for our study is the KORA/MONICA MI Registry (see Tunstall-Pedoe et al., 1994; Holle et al., 2005), comprising records of MI events that occurred within the study region from 1985 to 2015. These data were collected at the hospitals in the Augsburg region. Each record contains the date of the MI occurrence, age and sex of the patient. Depending on availability, complementary information is given, such as the patients' residential county (Landkreis), their body mass index (BMI), smoking status, and preexisting conditions such as diabetes. Although no detailed information is provided on the location of the patient during an MI event, they can be assigned to either the urban (City of Augsburg) or one of the two rural counties (Landkreise) of the study region (Landkreis Augsburg and Aichach-Friedberg). As pointed out earlier, the individual patient-specific data could not be used as predictive data due to the nature of the regression approach, which aims to predict the gross number of MI in the population. It is, however, possible to use these data to confine investigations to subgroups, e.g., to inhabitants of either urban or rural areas, and also to the elderly, or to smokers, albeit at the cost of being limited to a smaller subset of the overall data. In total the number of recorded MI is $n = 34,618$. Until 2008 the study was limited to participants of up to 74 years of age, with $n = 30,081$ records total in that category. Figure 1 shows the aggregated number of MI per year and the mean annual cycle for the population aged under 75. The yearly maximum in MI is observed during the winter months, whereas the summer time shows the lowest occurrences. To generate the ground truth for our regression problem, we counted the total daily number of MI observed in the KORA study and used the resultant time series as input for the ML algorithms.

Variable	Abbreviation	Unit	Time period	Resolution	Dataset	Reference
MI events (DONSET)	MI	#	1985-2015	District, daily	KORA	Helmholtz Zentrum München
Maximum temperature	TXK	°C	1985-2015	Station, daily	DWD	Deutscher Wetterdienst (DWD)
Mean temperature	TMK	°C	1985-2015	Station, daily	DWD	Deutscher Wetterdienst (DWD)
Minimum temperature	TNK	°C	1985-2015	Station, daily	DWD	Deutscher Wetterdienst (DWD)
21-Day rolling mean minimum temperature	TNK21D	°C	1985-2015	Grid, daily	This study	–
Relative humidity	UPM	%	1985-2015	Station, daily	DWD	Deutscher Wetterdienst (DWD)
Apparent temperature	ATMK	°C	1985-2015	Grid, daily	This study	–
Maximum apparent temperature	ATXK	°C	1985-2015	Grid, daily	This study	–
Minimum apparent temperature	ATNK	°C	1985-2015	Grid, daily	This study	–
Vegetation index	NDVI	-	1998-2015	1 km2, 10-daily	NDVI v2	Copernicus Global Land Service (CGLS)
Nitrogen oxide	NO	ppm	1993-2015	Station, daily	LÜB	Bayerische Landesamt für Umwelt
Nitrogen dioxide	NO_2	ppm	1993-2015	Station, daily	LÜB	Bayerische Landesamt für Umwelt
Sulfur dioxide	SO_2	ppm	1980-2015	Station, daily	LÜB	Bayerische Landesamt für Umwelt
Ozone	O_3	ppm	1990-2015	Station, daily	LÜB	Bayerische Landesamt für Umwelt
Particulate matter	PM_{10}	ppm	1980-2015	Station, daily	LÜB	Bayerische Landesamt für Umwelt
Male population (total)	mtotal	#	1985-2015	District, annual	LfStat	Bayerisches Landesamt für Statistik
Female population (total)	ftotal	#	1985-2015	District, annual	LfStat	Bayerisches Landesamt für Statistik
Male/female population age under 1	u1m, u1f	#	1985-2015	District, annual	LfStat	Bayerisches Landesamt für Statistik
Male/female population age 1 to 4	1t4m, 1t4f	#	1985-2015	District, annual	LfStat	Bayerisches Landesamt für Statistik
Male/female population age 5 to 9	5t9m, 5t9f	#	1985-2015	District, annual	LfStat	Bayerisches Landesamt für Statistik
Male/female population age 10 to 14	10t14m, 10t14f	#	1985-2015	District, annual	LfStat	Bayerisches Landesamt für Statistik
Male/female population age 15 to 19	15t19m, 15t19f	#	1985-2015	District, annual	LfStat	Bayerisches Landesamt für Statistik
Male/female population age 20 to 24	20t24m, 20t24f	#	1985-2015	District, annual	LfStat	Bayerisches Landesamt für Statistik
Male/female population age 25 to 29	25t29m, 25t29f	#	1985-2015	District, annual	LfStat	Bayerisches Landesamt für Statistik
Male/female population age 30 to 34	30t34m, 30t34f	#	1985-2015	District, annual	LfStat	Bayerisches Landesamt für Statistik
Male/female population age 35 to 39	35t39m, 35t39f	#	1985-2015	District, annual	LfStat	Bayerisches Landesamt für Statistik
Male/female population age 40 to 44	40t44m, 40t44f	#	1985-2015	District, annual	LfStat	Bayerisches Landesamt für Statistik
Male/female population age 45 to 49	45t49m, 45t49f	#	1985-2015	District, annual	LfStat	Bayerisches Landesamt für Statistik
Male/female population age 50 to 54	50t54m, 50t54f	#	1985-2015	District, annual	LfStat	Bayerisches Landesamt für Statistik
Male/female population age 55 to 59	55t59m, 55t59f	#	1985-2015	District, annual	LfStat	Bayerisches Landesamt für Statistik
Male/female population age 60 to 64	60t64m, 60t64f	#	1985-2015	District, annual	LfStat	Bayerisches Landesamt für Statistik
Male/female population age 65 to 74	65t74m, 65t74f	#	1985-2015	District, annual	LfStat	Bayerisches Landesamt für Statistik
Male/female population age over 75	o75m, o75f	#	1985-2015	District, annual	LfStat	Bayerisches Landesamt für Statistik
Random variable	RND	fractional	1985-2015	Study area, daily	This study	–
Day of the week	DOW	1–7	1985-2015	Study area, daily	This study	–
Month of the year	MOY	1–12	1985-2015	Study area, daily	This study	–
Day of the year	DOY	1–366	1985-2015	Study area, daily	This study	–

Table 3. Overview of predictive variables, source datasets and their origin.
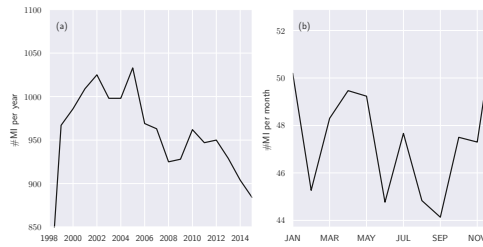
Figure 1. Number of annual MI (a) and mean annual cycle (b) in people aged under 75 from 1985 to 2015 for the study region (city of Augsburg and counties Aichach-Friedberg and Augsburg).

3.2 Air temperature and humidity

Air temperature close to the ground is the most important factor to consider as the most direct measure of human exposure to heat and cold. The relatively small spatial scale of the study region (1998 km^2) puts high demand on the data in terms of spatial resolution and accuracy. At the same time, daily environmental data are required for our approach.

We opted to derive a 1x1 km grid for the study period between 1985 and 2015 from daily data of 22 DWD stations in the vicinity of Augsburg and its neighboring districts. To this end, we applied universal Kriging with linear drift to the daily values at the temperature stations shown in Figure 2. The resulting gridded datasets (minimum, maximum and mean temperature) were aggregated to the counties comprising the study region. This relatively simple approach proved to be accurate enough to obtain realistic aggregated daily time series for the study region, as shown by the reasonable predictions in this paper.

We also include humidity features in the models to gauge their relative importance. Relative humidity was also gathered from DWD and we applied the same Kriging procedure for spatial interpolation, as used for temperature. To account for possible effects of perceived heat stress expressed by simultaneous high humidity and high temperatures we included apparent temperature. Measures of apparent temperatures relate a given temperature to the ambient humidity to account for the perceived temperature differences between dry and humid conditions. The specifics of the computation can be found in the Appendix A1.

In a next step, the data was aggregated for the three different counties within the model region, the urban and the two rural areas by computing weighted area means. The resulting daily time series can be readily used as input to the ML models, as described in Section 2.

3.3 Air quality

Air pollution is usually a complex mixture but several particulate and gaseous pollutants can be considered in investigating its effects on MI (e.g., Chen et al., 2018; Bourdrel et al., 2017; **?**). From the "Bavarian Air Hygiene State Monitoring System" (LÜB) database (Bayerische Landesamt für Umwelt) we collected data on PM$_{10}$, NO, NO$_2$, Ozone (O$_3$), and SO$_2$ concentrations at multiple stations across Bavaria at daily resolution.

Table A1 in the Appendix gives an overview of the selected measuring stations and their urban or rural categories, the corresponding pollutants data and their availability. Figure 2 gives an overview of the selected temperature and air quality measurement stations. We determined the aggregated daily means by calculating the mean values of the aforementioned stations, taking into consideration their location proximity to the city centers, traffic-loaded inner-city streets, on industrial areas, on the outskirts, or the large-scale background pollution.

The map shows that there are only few air quality stations within the study region (five blue circles, and an archived station with red border circle). Since not all stations have been always active during our study period, we use merely the active stations. However, if none of the regularly used station in the counties had recorded data on a given day, especially for the surrounding counties, alternative stations (light blue dots in Figure 2) with equal proximity settings from outside the study region were used as replacements for the calculations. This has been achieved through an acceptable 10-15 percent error criterion for the monthly value of alternative stations compared to the calculated monthly mean value of the county over a span of time provided by the monitoring system. The calculated monthly mean time series have been provided in the appendix figure A9.

3.4 Vegetation

The Normalised Difference Vegetation Index (NDVI) is an indicator of the greenness of the natural vegetation and other vegetation types such as agriculture, parks and gardens. It is widely used for ecosystems monitoring. In this study NDVI also is used as a proxy for shade as well as potential local cooling effect of vegetation by absorbing sunlight and through evapotranspiration. The NDVI_v2_1km database of the Copernicus Global Land Service (CGLS) vegetation products is freely available at a 1x1 km spatial resolution starting on April 1998 measured every ten days. We extracted the NDVI for our region and used a cubic spline interpolation to upscale the temporal resolution from 10 days to daily values. Given the very gradual rate of change in vegetation cover and consequently the NDVI, we assume this interpolation does not produce large errors. Note that due to lack of availability of NDVI data before 1998, training and testing of the algorithms had to be confined to the time between April 1998 and December 2015.

3.5 Demographics

The absolute number of MI does not only depend on various environmental risk factors but also on the size and characteristics of the population. Disregarding other factors, any change in the absolute number of inhabitants would produce a similar change in the number of cases of MI as well. Moreover, both age and sex are strongly correlated with health outcomes in general, and specifically so for MI. Given trends of increasing urbanisation, rural depopulation, and an ageing society, it is important to account for changes both in number of inhabitants and age stratification of the population over time. In addition, domestic migration reflected in relative changes between urban and rural parts, leading to differential changes in exposure to environmental hazards in the Augsburg region, can be important as well. We collected data from the Bavarian Office of Statistics that comprise annual values for the total number of inhabitants for each of the three counties, as well as the distribution of sex and age in the population from 1985 to 2015. Overall, 17 different age groups are accounted for as listed in Table 3. Since the algorithms require daily input values, a linear interpolation was applied to estimate the development within a given year.

Figure 2. Air quality (blue) and temperature (orange) stations in the region of interest (ROI) around Augsburg.

4   Results

4.1   Weekly predictions of MI events

Our models produce daily predictions of MI events based on the environmental and demographic features within the given window size. We found that the models are not able to reproduce the daily variability of MI with sufficient accuracy. As an example, we show the daily predictions aggregated to 7-day intervals to increase visibility in Figure 3. The resultant scores are given in Table 4 for both training and validation respectively.

Although the seven-day predictions suggest some skill for the training period, for the testing period the models do not predict 7-day variations (or day-to-day predictions) accurately enough for practical purposes. The predictions are too close to the mean and lack the variability displayed by the observations. An overview of average mean, standard deviation, minimum and maximum daily predictions across models and for each subgroup considered is given in Table A8. This is likely related to
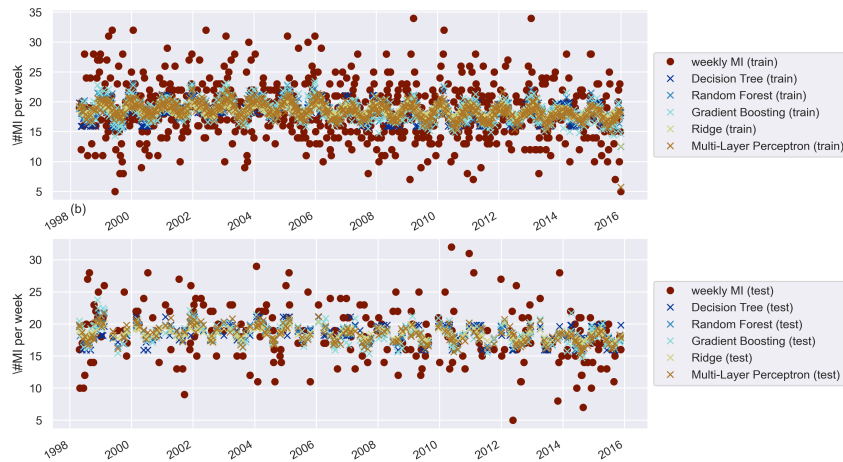
Figure 3. Daily predictions aggregated to 7-day intervals for all models. Shows predicted (crossed) and observed MI (dotted) for the training (top) and test (bottom) sets.

Method	DTR		RF		GBR		RR		MLP	
Score	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Adj R^2	0.02	-0.22	0.24	-0.19	0.29	-0.26	-0.01	-0.21	0.04	-0.21
R^2	0.1	0.06	0.3	0.08	0.35	0.03	0.06	0.07	0.11	0.07
Max-Error	15.8	14.17	13.58	14.26	13.63	15.58	14.95	13.82	15.36	14.03
RMSE	4.44	4.38	3.91	4.33	3.77	4.44	4.52	4.36	4.41	4.37
BIC	2419.29	979.57	2241.72	973.88	2194.26	986.19	2444.85	977.7	2409.56	978.11

Table 4. Training and test scores for 7-day aggregated daily predictions.

randomness as well as risk factors that affect MI events that were not considered in the models. For instance, the temperature or air quality predictors may not sufficiently capture actual local circumstances, but also information about the built environment and other conditions that cannot be easily accounted for is missing.

4.2 Annual predictions of MI events

260 Figure 4 shows the model performances on both the training and test sets as well as the actually observed MI as a reference for the five ML models given in Table 1. After training the models and performing the daily prediction on the test set, the results were aggregated to annual sums. By aggregating the model results to an annual basis, some of the inherent randomness is averaged out. Based on the annualised prediction results and time series of observed MI, the performance scores were derived (see Table 5). The training scores demonstrate that the ML models are able to predict the year-to-year variations quite well, with
265 adjusted R^2 scores between 0.87 and 0.94. The performance on the test dataset is relevant for assessing the generalisation error

for previously unseen data. In contrast to the training data, the results on the test set are less but still reasonably accurate, with adjusted R^2 scores between 0.62 and 0.71, showing that inter-annual variations and long-term trends are largely captured. The RR and MLP models exhibit the best performance, showing that both well-tuned linear models as well as neural networks are able to simulate the relations between environmental conditions and MI events. The DTR shows the lowest overall performance by comparison.
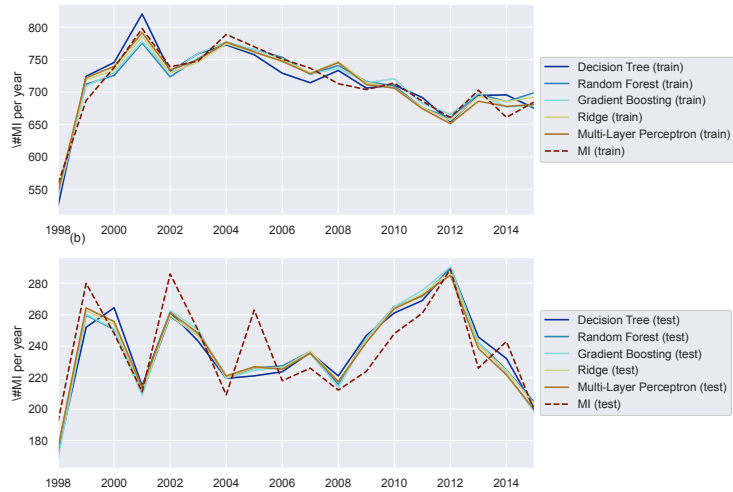
Figure 4. Annually aggregated predictions of MI in the general population for all models. Predicted (solid) and observed MI (dashed) for training (a) and test (b) sets.

Method	DTR		RF		GBR		RR		MLP	
Score	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Adj R^2	0.87	0.62	0.92	0.68	0.94	0.66	0.92	0.7	0.93	0.71
R^2	0.87	0.64	0.93	0.69	0.94	0.67	0.92	0.71	0.93	0.72
Max-Error	37.37	41.91	28.09	36.83	23.89	38.33	33.96	36.59	35.21	36.11
RMSE	19.14	17.4	14.71	16.04	12.76	16.48	15.13	15.63	14.47	15.19
BIC	262.34	258.91	252.87	255.99	247.74	256.96	253.89	255.04	252.29	254.03

Table 5. Training and test scores on annual basis for the general population.

4.3 Feature importance

In Figure 5 we show a condensed rendition of the feature importance where related variables have been grouped together of each model; except for the MLP which does not support feature importance within the scikit-learn framework. Note that variables subject to the sliding window were aggregated over the window length of three days to improve readability. Moreover, features related to time such as the current month number and the day of the week were also aggregated to a single group. More detailed plots retaining the differentiation of all features and window days can be found in the Appendix (see Figures A6 and A7). The latter Figure also shows that many of the original demographic features carry little to no weight. We therefore reduced the granularity of the demographic data to the age groups $0-29, 30-49, 50-74$ and > 75, generally yielding improved results.

While the performance of the models differs, some trends can be observed. Overall, the single most important group is air quality, closely followed by temperature, demographic and time related predictors. Humidity as well as NDVI exhibit the lowest explanatory power. NDVI is ranked very closely to the random feature by all models.
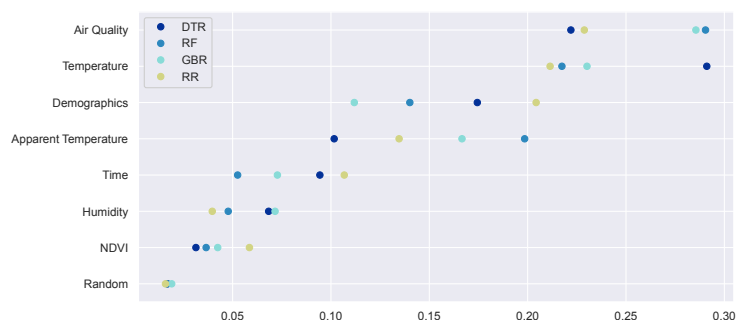
Figure 5. Aggregated feature importance for predicting MI for the general population. Related features have been grouped thematically. Larger values indicate higher importance and per model the sum over all features equals one.

Compared to the environmental features that display strong daily variation the demographic predictors are subject to slow, gradual change only. We therefore also conducted this experiment with all demographic features turned off. The results are shown in Table A7. As evidenced by the reduction in scores (adj. R^2 reduced from 0.67 to 0.62 on average) the demographic predictors still make a relevant contribution to the overall result despite the lower temporal resolution of the input data.

4.4 Subgroup analysis

The models were also applied to subgroups of the population, albeit at the expense of a reduced amount of available training data (see Table 6) for an overview. For this analysis we selected a total of five subgroups: the urban (Augsburg city) and rural population (two adjacent counties) respectively, the elderly (people aged between 60 and 74), patients with diabetes, and active smokers. The data was reduced to include only participants with the associated attribute. The training procedure was then repeated as detailed for the general case on the resulting subsets. As expected, the validation scores dropped considerably for all subgroups, likely a consequence of reduced amounts of training data. We refer to the Appendix for detailed results, but

for the urban and rural subgroups adjusted R^2 scores between 0.35 and 0.6 were observed in validation (see Tables A2 and A3). Both subgroups, being of almost equal size, performed comparably well, with the urban population exhibiting slightly lower scores however.

Population	n			Age			
	all	female	male	min	max	mean	stdev
General	17134	4330	12804	25.0	74.0	62.34	9.42
Urban	8424	2248	6176	25.0	74.0	62.46	9.46
Rural	8710	2082	6628	25.0	74.0	62.22	9.39
Elderly	11470	3314	8156	60.0	74.0	67.94	4.26
Diabetic	5451	1521	3930	26.0	74.0	64.44	8.07
Smoker	3800	793	3007	26.0	74.0	55.72	9.76

Table 6. Overview of the number of cases as well as age and sex distribution for the different study populations considered.

The validation results for the elderly population (see Figure A3 and Table A4) are more accurate (adjusted R^2 between 0.53 and 0.65) than for the urban and rural populations, although the number of training samples is much higher in both of those cases.

The results for patients with diabetes are shown in Figure A4 and Table A5. As observed with the elderly, the scores for patients with diabetes (adjusted R^2 between 0.28 and 0.61) are comparable to those of the (much bigger) rural and urban subgroups, except for DTR which resulted in a substantially reduced score.

The results for the smoking population are shown in Figure A5 and the scores are given in Table A6. For this group adjusted R^2 validation scores drop to around 0.42 on average, indicating a less accurate fit than for all the other subgroups. This is consistent with the smoker group being the smallest of the explored subgroups resulting in the lowest amount of training data as well.

Overall, the skill of the models is clearly reduced when limited to subsets of the overall data. The decrease in performance, however, is quite different between subgroups, especially when taking into account their relative sizes. A particularly interesting question is whether the variable importance for any one subgroup changes substantially in comparison to the general population. Figure 6 shows the difference in variable importance for each of the subgroups in relation to that of the general population. To aid readability related features have been grouped again. Considerable differences between subgroups, models and feature groups can be observed. For instance, most models agree that demographics, humidity and NDVI are particularly important for predicting urban MI, while giving less weight to the temperature related features. The importance of time related indicators reduced consistently over the general population. In some cases the importance of the random feature is also reduced, indicating increased robustness of the results.

For the rural population the results suggest slightly increased relevance NDVI compared to the overall population. Temperature and air quality mostly align with the results for the general population. The demographic indicators are less relevant when compared to the general case, as are the time related features.

For the elderly, the models are mostly undecided on air quality, with a slight tendency towards increased importance. The weight of the demographic features is emphasized in comparison to the general case. Less importance is also attributed to (apparent) temperature and humidity.

For patients with diabetes, the models mostly agree that demographic features, NDVI and air quality are more important in predicting MI for this group in comparison to the general population. On the other hand, (apparent) temperature, humidity and time related features are ranked lower.

Lastly, for the group of active smokers the models mostly suggest an increased importance of air quality as well as demographic features for the prediction of MI. Humidity as well as (apparent) temperature and time-related features are overall considered less important.
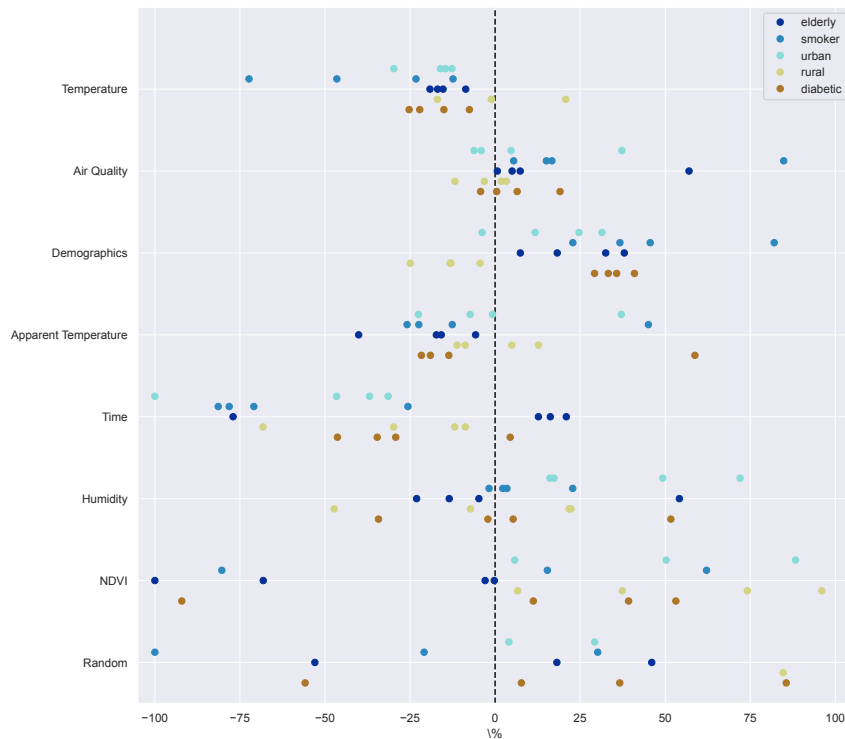
Figure 6. Change of feature importance for predicting MI in percent relative to the general population for every subgroup considered. Related features have been grouped thematically.

5 Discussion

To our knowledge, this is the first study building and testing ML models that include more than only weather variables (such as Zhang2009 for heat mortality) for predicting MI prevalence. The developed ML models have varying skill in predicting MI. At

the daily to 7-day time scales, randomness seems too large to produce meaningful predictions. However, when predictions are aggregated to annual sums, the models are very well capable of reproducing the inter-annual variability of observed MI, as well as the long-term trends, also for the validation datasets. This is comparable to the performance of methods used for predicting malaria incidence (e.g., Sewe et al., 2017). In terms of performance scores the models achieve very similar outcomes both in training and validation (see Table 5), indicating some robustness of the predictions. More qualitative differences emerge, however, when investigating feature importance. There are substantial differences between the ML models in terms of some features (Figure 5). Most models rank air quality variations and temperatures among the most important features, but a large spread between models can be observed. This indicates at least some inherent uncertainty.

Classical epidemiological approaches like general linear or additive models are mostly used for explaining the direction and corresponding uncertainty of associations between environmental risk factors and health outcomes, thereby adjusting for potential confounding factors. In case of potential non-linearities, the shape of the exposure-response curve is usually modeled as a smooth function. However, the models are limited in case of high correlation and/or high-dimensional interactions between the covariates. The suggested ML approaches can (partly) handle these issues and offer the possibility to compare the importance and predictive performance of a multitude of environmental predictors.

The training scores in many cases are close to the maximum, with adjusted R^2 values greater than 0.85. This may be indicative of overfitting, possibly opening room for improving further on the generalisation by applying stronger regularisation. While the models were adjusted by optimising the hyperparameters, not all possible parameter values have been explored. For instance, in the case of the tree-based models pruning is an effective way to reduce overfitting, which was not applied here. For the MLP and RR models regularising parameters were explicitly included in the optimisation, but possibly the ranges were not wide enough to achieve the best trade-off between training and validation.

The model results are sensitive to the selection of the random seed that is used in making the initial train-test split. We found that changes in the random seed routinely had greater impact than the choice of hyperparameters. One way of dealing with this would be to also include this random seed in the optimisation process. Currently, only the random seeds used for randomly selecting the folds in cross-validation and in initialising the regressors are optimised. In light of the strong influence of the initial split, however, we opted to instead test over a range of possible seeds and select the results closest to the average performance of the models, not to overstate our results. The sensitivity to the initial split may indicate a lack of data, but is likely mostly due to unbalanced splits. We reduced this sensitivity by employing a simple but effective stratification strategy. This reduced the variation across seeds, but does not entirely resolve the issue. Possibly, more intricate stratification approaches may reduce the dependency even further.

We were able to indicate differences between different geographical regions, i.e., urban and rural populations. For instance, humidity, demographics and NDVI become more important predictors for the urban population, compared to the overall population, at the expense of (apparent) temperature. The models could be further improved by increasing the spatial representation, as the environmental predictors also would support this. Increased spatial representation would also allow for additional exposure metrics to be established and more predictors, such as those related to building structures, their insulation and energy efficiency.

An additional area of possible improvement is the environmental data. Some variables such as the NDVI and the air quality indicators were not fully available for the period between 1985 and 2015, effectively limiting analysis to the period from 1998 to 2015. It is also possible to reduce the bias in the station data (temperature, air quality), for instance by using more sophisticated interpolation methods or additional data sources such as from remote sensing.

All ML models results consistently demonstrated the importance of the air quality variables. Climate impact studies, especially related to MI, might therefore benefit from carefully analysing possible future developments of these variables. Electrification of traffic, reduction of fossil fuel and related changes might yield substantial improvements in air quality in the future. Instead of just focusing on projected changes in, e.g., temperature and humidity, scenarios for air quality need to be considered as well.

Current data availability from climate modelling, and demographic and environmental scenario development provide many opportunities to use the developed ML models from our research for projecting future health risks. Ensembles of regional climate models provide climate projections with the highest spatial resolution. For the study region, EURO-CORDEX simulations (Jacob et al., 2014, 2020) can be considered as they provide the largest ensemble at a high spatial (0.11°, i.e., 12 km) and temporal (daily) resolution climate simulations that are available today. Several of the predictors used in this study could be derived from the EURO-CORDEX ensemble, namely temperatures and in many cases relative humidity, as well as dew-point temperatures. Alternatively, an ensemble of convection permitting decadal regional climate simulations at 3km, both for historic and future conditions, has been created within CORDEX FPS (e.g., Ban et al., 2021). Using an ensemble of near-future (2035-2065) climate model simulations allow for scenario uncertainty, internal climate variability, and climate model uncertainty to be assessed (Hawkins and Sutton, 2011) when comparing the changes in MI to the reference historical simulations.

Demographic predictions until 2039 for the study region at county level can be obtained from the Bayerisches Landesamt für Statistik. Longer-term projections up until the year 2060, albeit contingent on different socio-economic scenarios and at the level of the federal state of Bavaria, could be obtained from the Statistisches Bundesamt and be used to estimate the local projected demographics in the study area. These projections would provide a robust basis to estimate potential developments of the local population in the near-future.

For vegetation changes, as represented by NDVI, it can be reasonably assumed that the potential for increased greenness in the inner city is limited. Likewise, the potential for substantial effects from added green in the rural surroundings of Augsburg is low, as it is already ubiquitous there. We therefore believe that moderate up- or downscaling of NDVI patterns observed in the past and present may suffice to yield suitable estimates of possible future developments, such as adaptation measures of increasing vegetation to reduce the urban heat island effect.

Air quality projections are related to the emission scenarios used by global climate models. For the CMIP6 climate models, estimates of regional surface air quality are available at the global model scale Turnock et al. (2020). These projections could be used to scale the observed daily air quality observations, but more exhaustive and local projection data would be preferred. To date, however, regional climate models do not feature the necessary complex chemical models to accurately model the

transport, dispersion and diffusion of pollutants. A pragmatic up- or downscaling of the observed patterns from the global to the local level currently appears as the most convenient approach.

6 Conclusions

We have developed an approach for predicting MI events using multi-variable ML methods, based on environmental and demographic data. Given that health outcomes depend on a multitude of factors, we applied a data-driven approach to establish relevant relationships. We acquired data on MI events from the KORA MI registry in Augsburg, Germany, as well as weather, environmental and demographic data from various sources to create a meaningful and consistent daily time series of the predictive features and the target variable.

Starting from these time series, a supervised learning problem for MI was formulated, accounting for lagged effects. Five different regression algorithms were trained on this data, based on random $75/25$ train-test splits for the period between April 1998 and December 2015. Various hyperparameters were used to optimise the performance of the algorithms, based on 5-fold cross validation with respect to the R^2 scores.

Applying the trained models on the unseen test data allowed an estimation of the generalisation error of the models. We found that the daily or weekly results do not meaningful and accurate predictions of MI events. We found that the annually aggregated predictions agree well with the observed MI events, accurately reflecting observed trends and inter-annual variability of MI. The match between observations and the model predictions is supported by the observed validation scores, with adjusted R^2 scores ranging between 0.62 and 0.71. Overall, the models displayed comparable skill, but the Ridge Regression (RR) and Multi-layer Perceptron (MLP) models slightly outperformed the tree-based methods. The least accurate results were produced by the Decision Tree (DTR) model. The feature importance showed that despite similar overall scores, the relative weight can vary substantially between the models. This emphasised the necessity to consider ensembles of models, as it allows to gauge the model spread and estimate inherent uncertainty. In this study, air quality tends to be the most important feature to predict MI, closely followed by temperature, demographics, and apparent temperature. We also applied the models to various vulnerable subgroups, such as the elderly or patients with diabetes, resulting in only slightly reduced skill scores due to the reduced amounts of training data.

Possibilities to improve the current approach are manifold, including increasing the variety and quality of the predictor data. Further analysis of the data, including accounting for trends over time, may further increase robustness of the results to prevent the attribution of exogenous effects not considered in the model to the existing features. Also, different ML approaches could be explored, such as density estimation and Bayesian methods, yielding estimates of relative risk of different groups to suffer MI. Such estimates could be more readily compared with commonly used epidemiological models than the regression models presented here. Overall, the models' capacity to give reasonable estimates of possible future developments of MI based on the predictive features appears robust. In a next step, the trained models can be applied to scenarios of future climatic, environmental and demographic conditions. This will allow estimating future changes in MI taking into account climatic, as well as other environmental and demographic factors expanding on limitations of earlier studies. These changes could also

include further improvements in air quality, or increased 'greening' of urban environments with vegetation. Such estimates will enable to gauge the sensitivity of the complex health-environment interactions, and benefits of proposed environmental and health interventions in urban areas.

Appendix A

A1 Derivation of apparent temperature

We have computed the apparent temperature according to:

$$T_a = -2.653 + 0.994T + 0.01537T_d^2$$

where T_a is the apparent temperature, T the near-surface mean temperature and T_d the near surface dewpoint temperature (see Davis et al., 2016). Dewpoint temperature, however, was not available for this study. To facilitate estimating the apparent temperature we therefore first derived another humidity related quantity: vapour pressure. Applying again universal Kriging with linear drift, we arrived at 1x1 km gridded data for vapour pressure, applying the Magnus formula to estimate the dewpoint temperature:

$$T_d = \frac{b \cdot v}{a - v}$$

where $a = 7.5$, $b = 237.3$, $v = \log_{10}\left(\frac{p_v}{6.1078}\right)$ with p_v the vapour pressure.

Applying these formulas to the gridded temperature and humidity data derived before yields a 1x1 km grid for apparent temperature. Note that the formulas were independently applied to mean, maximum and minimum temperature. Subsequent aggregation over the model region then completed the preparation of apparent temperature as input feature.

A2 Detailed subgroup and feature importance results
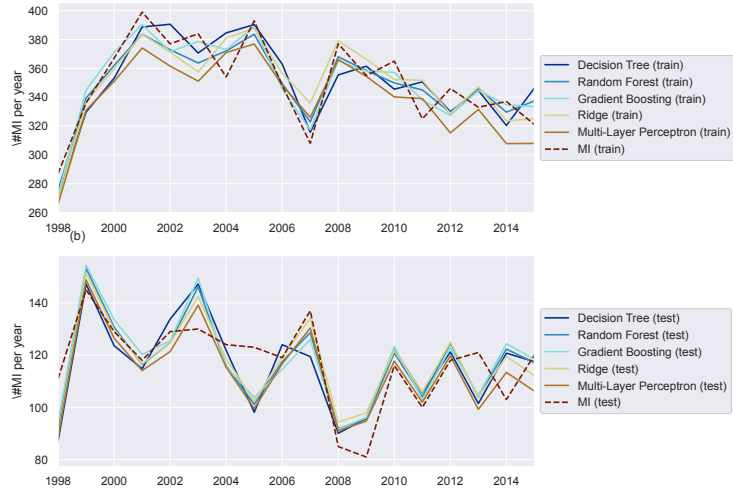
Figure A1. Annually aggregated predictions of MI in the urban population for all models. Predicted (solid) and observed MI (dashed) for training (a) and test (b) sets.
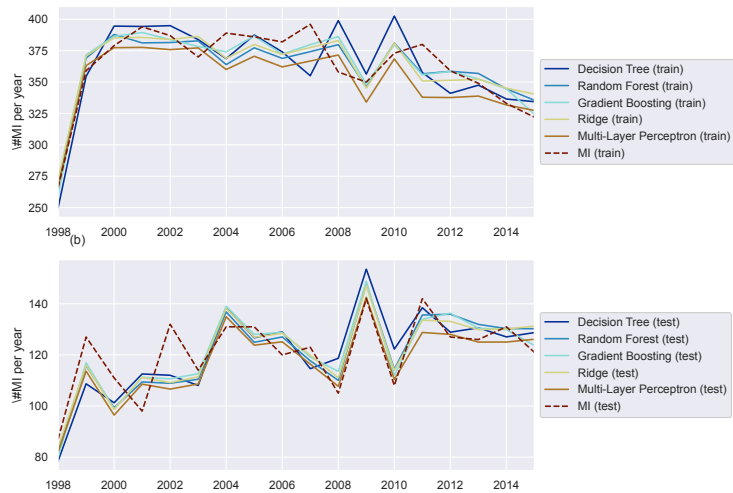
Figure A2. Annually aggregated predictions of MI in the rural population for all models. Predicted (solid) and observed MI (dashed) for training (a) and test (b) sets.
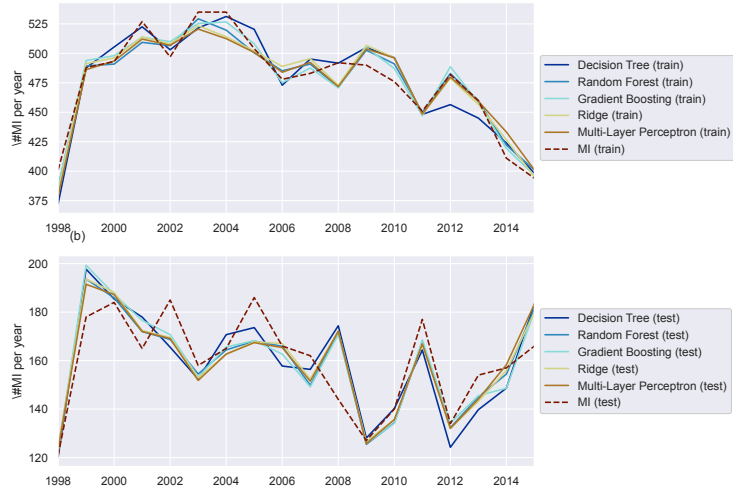
Figure A3. Annually aggregated predictions of MI in the elderly population for all models. Predicted (solid) and observed MI (dashed) for training (a) and test (b) sets.
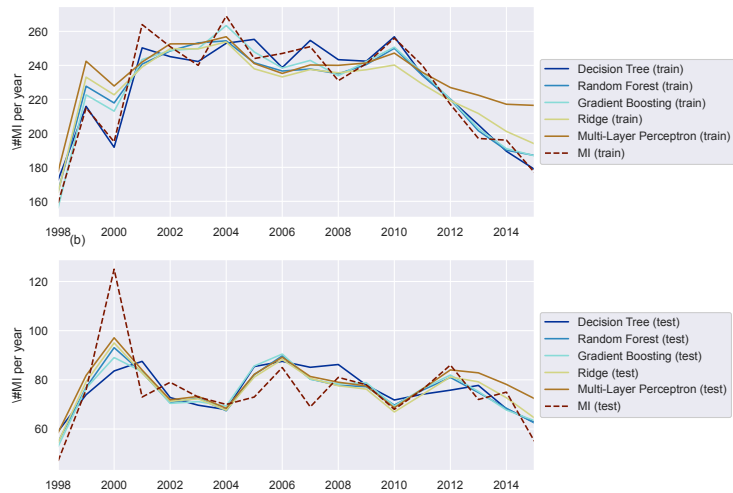
Figure A4. Annually aggregated predictions of MI in the diabetic population for all models. Predicted (solid) and observed MI (dashed) for training (a) and test (b) sets.
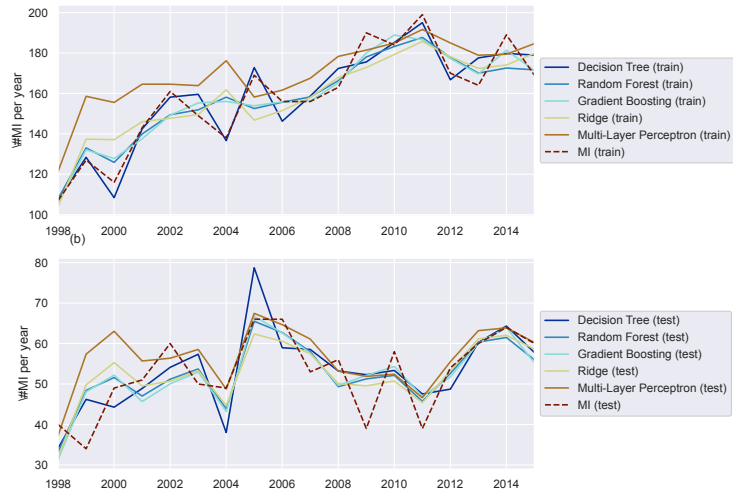
Figure A5. Annually aggregated predictions of MI in the smoking population for all models. Predicted (solid) and observed MI (dashed) for training (a) and test (b) sets.
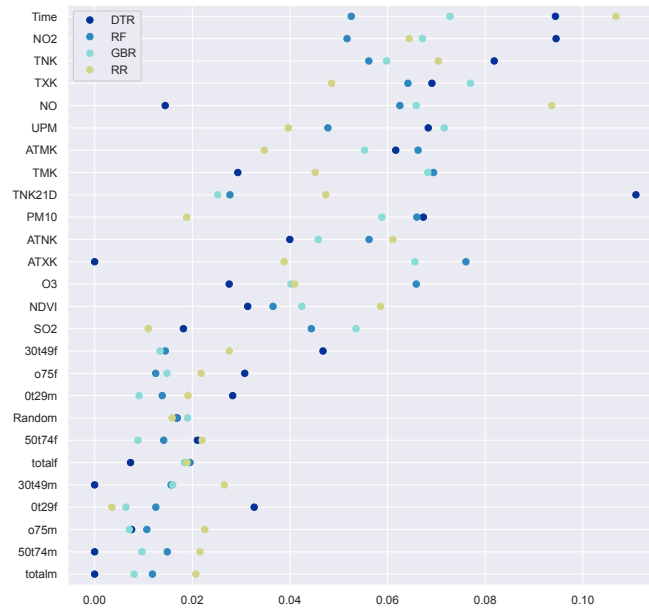
Figure A6. Feature importance for predicting MI in the general population. Larger values indicate higher importance and per model the sum over all features equals one.
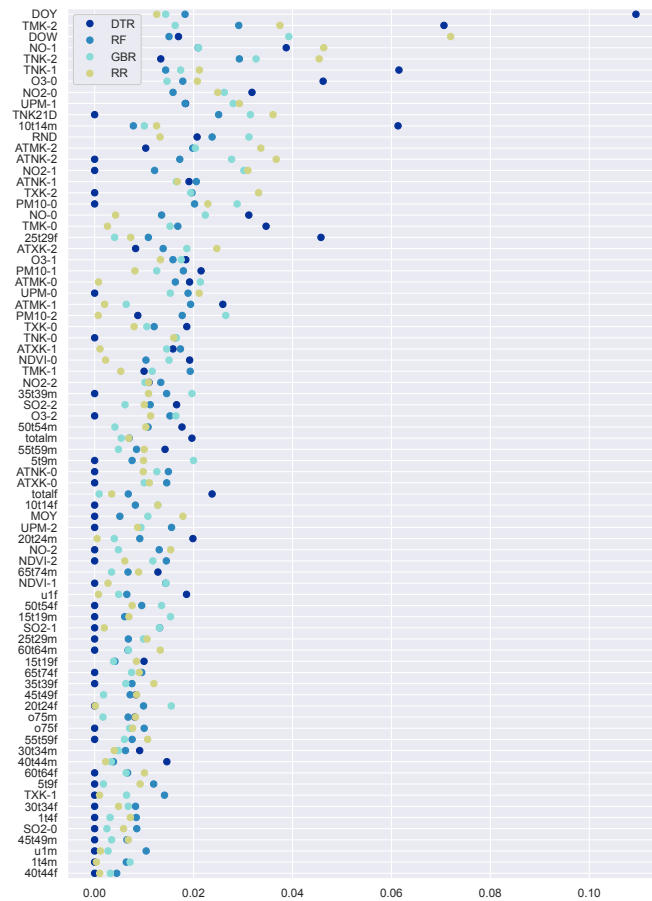
Figure A7. Feature importance for predicting MI in the general population when the full set of demographic features is used. Where applicable the lag in days (0, 1 or 2) is indicated. Larger values indicate higher importance and per model the sum over all features equals one.
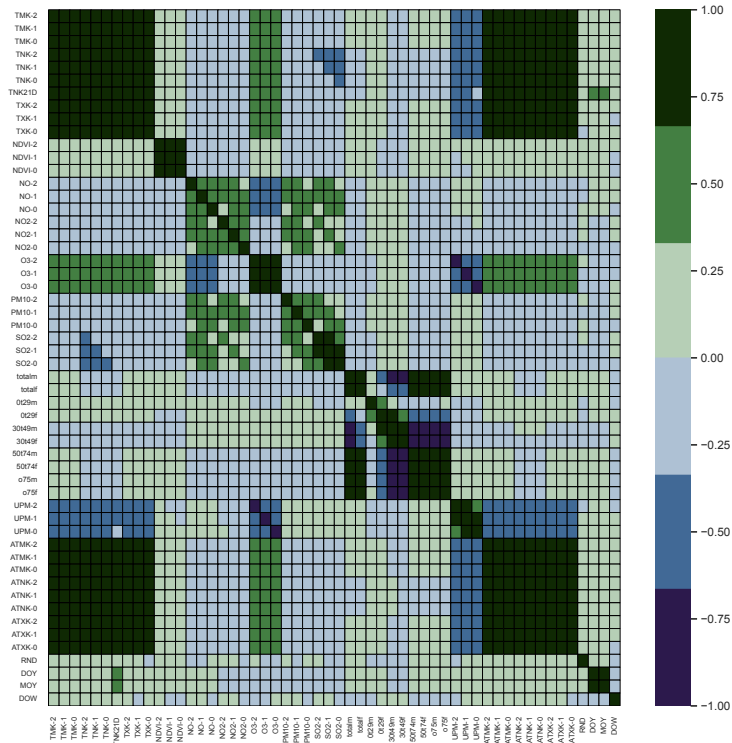
Figure A8. Feature correlation matrix differentiated by lag in days (0, 1 or 2) and with the full set of demographic features.
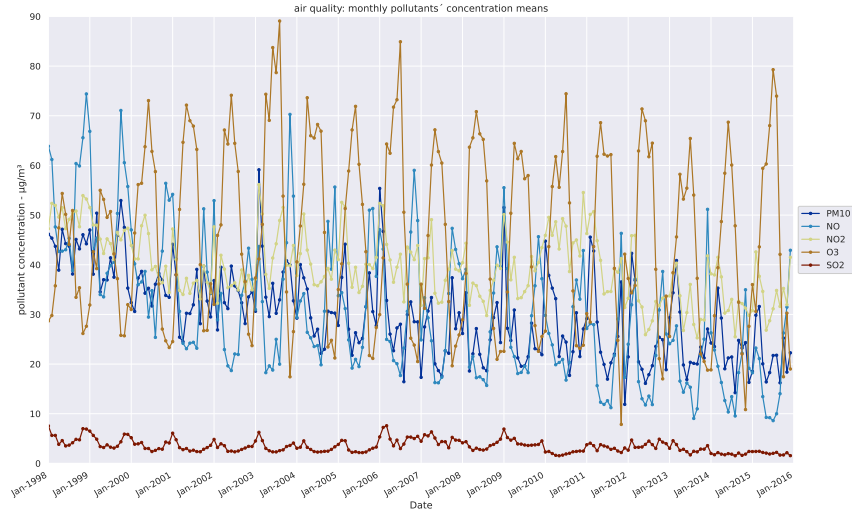
Figure A9. Monthly mean time series for all air quality features.

Stations:	Augsburg/Königsplatz	Augsburg/BourgesPlatz	Augsburg/LfU	Augsburg/Karlstraße	Augsburg/Haunstetten	Andechs/Rothenfeld	Kempten/Westendstraße
	Time period(resolution)	Time period(resolution)	Time period(resolution)	Time period(resolution)	Time period(resolution)	Time period(resolution)	Time period(resolution)
Urban-Rural categories:	Urban					Rural	
PM_{10}	1980-1985 (daily) 1986-2004(3hr) 2005-2018(1hr)	1986-2004(3hr) 2005-2018(1hr) 2013-2014 missing	2000-2004(3hr) 2005-2018(1hr) 2010-2016 missing	2003-2004(3hr) 2005-2018(1hr)	-	2003-2004(3hr) 2005-2018(1hr)	1980-1985 (daily) 1986-2004(3hr) 2005-2015(1hr) 2015-2018 missing
$PM_{2.5}$	-	2008-2018(1hr)	2008-2018(1hr) 2010-2016 missing	-	-	2012-2018(1hr)	2014-2018(1hr)
NO	1980-2018(1hr)	1986-2018(1hr)	2000-2018(1hr) 2012-2013 missing	2003-2018(1hr)	-	2003-2018(1hr)	1993-2018(1hr)
NO_2	1980-2018(1hr)	1986-2018(1hr)	2000-2018(1hr) 2012-2013 missing	2003-2018(1hr)	-	2003-2018(1hr)	1993-2018(1hr)
Ozone (O_3)	1980-1985(1hr)	2012-2018(1hr)	2000-2018(1hr)	-	1985-1999(1hr)	2003-2018(1hr)	1990-2018(1hr)
CO	1980-1999(1hr) only 2018(1hr)	1987-1999(1hr)	only 2018(1hr)	only 2018(1hr)	1980-1999(1hr)	-	1980-1999(1hr)
SO_2	1980-2018(1hr)	1986-2018(1hr)	2000-2018(1hr) 2012-2013 missing	2003-2018(1hr)	-	2003-2018(1hr)	1993-2018(1hr)
BTX	1980-2018(1hr)	1986-2018(1hr)	2000-2018(1hr) 2012-2013 missing	2003-2018(1hr)	-	2003-2018(1hr)	1993-2018(1hr)

Table A1. Air quality stations data availability and categories

Method	DTR		RF		GBR		RR		MLP	
Score	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Adj R^2	0.67	0.35	0.8	0.48	0.87	0.45	0.69	0.5	0.57	0.47
R^2	0.67	0.37	0.81	0.49	0.87	0.47	0.69	0.51	0.58	0.49
Max-Error	30.59	24.88	20.29	21.83	19.03	21.39	27.91	21.33	32.91	23.44
RMSE	16.69	12.7	12.84	11.38	10.39	11.68	16.12	11.15	18.93	11.43
BIC	257.41	247.59	247.97	243.61	240.34	244.55	256.17	242.9	261.94	243.78

Table A2. Training and test scores on annual basis for the urban population.

Method	DTR		RF		GBR		RR		MLP	
Score	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Adj R^2	0.62	0.43	0.8	0.59	0.85	0.6	0.79	0.59	0.67	0.55
R^2	0.62	0.45	0.8	0.6	0.85	0.61	0.79	0.6	0.68	0.57
Max-Error	40.97	19.97	25.12	23.09	28.21	21.38	29.16	22.84	42.1	25.38
RMSE	19.04	10.57	13.72	8.97	11.84	8.87	14.1	9.02	17.67	9.39
BIC	262.15	240.98	250.35	235.05	245.06	234.64	251.35	235.26	259.47	236.72

Table A3. Training and test scores on annual basis for the rural population.

Method	DTR		RF		GBR		RR		MLP	
Score	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Adj R^2	0.88	0.53	0.93	0.65	0.94	0.61	0.9	0.65	0.89	0.64
R^2	0.88	0.54	0.93	0.66	0.94	0.62	0.9	0.66	0.89	0.65
Max-Error	28.76	30.37	20.63	27.66	21.31	26.74	20.93	28.16	22.54	28.13
RMSE	13.7	12.94	10.54	11.21	9.93	11.75	12.6	11.19	13.39	11.34
BIC	250.3	248.24	240.87	243.1	238.74	244.78	247.29	243.03	249.48	243.51

Table A4. Training and test scores on annual basis for the elderly population.

Method	DTR		RF		GBR		RR		MLP	
Score	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Adj R^2	0.93	0.28	0.87	0.59	0.92	0.5	0.79	0.61	0.63	0.54
R^2	0.93	0.31	0.87	0.6	0.92	0.51	0.79	0.63	0.63	0.56
Max-Error	16.0	41.4	23.24	31.92	20.59	35.94	27.67	29.99	39.45	27.9
RMSE	8.16	12.55	11.06	9.53	8.58	10.5	13.79	9.22	18.49	10.03
BIC	231.64	247.15	242.61	237.24	233.44	240.73	250.55	236.04	261.1	239.08

Table A5. Training and test scores on annual basis for the diabetic population.

Method	DTR		RF		GBR		RR		MLP	
Score	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Adj R^2	0.91	0.37	0.85	0.53	0.87	0.48	0.74	0.5	0.41	0.22
R^2	0.91	0.39	0.85	0.55	0.87	0.5	0.74	0.51	0.41	0.25
Max-Error	14.43	13.3	20.14	14.42	18.04	14.08	23.77	15.78	39.55	23.37
RMSE	7.42	7.35	9.5	6.37	8.69	6.68	12.54	6.6	18.79	8.18
BIC	228.25	227.91	237.14	222.72	233.91	224.45	247.11	224.0	261.69	231.74

Table A6. Training and test scores on annual basis for the smoker population.

Method	DTR		RF		GBR		RR		MLP	
Score	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Adj R^2	0.83	0.67	0.81	0.63	0.9	0.63	0.74	0.59	0.74	0.57
R^2	0.83	0.68	0.81	0.63	0.9	0.64	0.74	0.6	0.74	0.58
Max-Error	39.97	34.42	47.65	38.81	28.72	38.79	56.42	40.22	53.93	40.96
RMSE	22.01	16.35	23.38	17.42	16.81	17.22	27.51	18.18	27.34	18.59
BIC	238.47	227.76	240.65	230.04	228.77	229.64	246.5	231.58	246.28	232.39

Table A7. Training and test scores on annual basis for the general population with demographic features turned off.

Subgroup	Mean		Stdev		Min		Max	
Score	Train	Test	Train	Test	Train	Test	Train	Test
General Population	2.64	2.63	0.26	0.26	1.91	1.92	3.64	3.58
Urban	1.29	1.29	0.15	0.15	0.91	0.93	1.99	1.84
Rural	1.34	1.33	0.13	0.13	0.94	1.0	1.86	1.76
Elderly	1.77	1.76	0.21	0.21	1.15	1.23	2.77	2.51
Diabetic	0.85	0.85	0.13	0.12	0.5	0.53	1.4	1.28
Smoker	0.59	0.59	0.09	0.09	0.38	0.37	0.99	0.85

Table A8. Daily mean, standard deviation, minimum and maximum predictions across models for each subgroup.

Code and data availability. The data used in this paper are available from third party sources. The principle MI registry data is available from the KORA data set, and can be applied for at HMGU here: https://epi.helmholtz-muenchen.de/. Other data sources (environmental and demographic data) are available from the sources quoted in the paper (Table 3). The code used for the ML models and data pre- and post-processing is available on request from the authors.

Author contributions. L.M., W.Z.C., and L.M.B. designed the research, L.M. and M.V. set up the models, collected the data, and did the simulations, L.M., M.V., and L.M.B. analysed the results and wrote the paper. AS, CM, JL and KW provided the KORA MI data. All authors contributed to the interpretation of results, reviewed the paper, and approved the final draft, and accepted responsibility to submit for publication.

Competing interests. The authors have no competing interests.

Acknowledgements. We thank our colleague Gaby Langendijk, and Francesco Sera and another anonymous referee, for their comments and suggestions that helped to substantially improve this paper. This study is part of the Digital Earth project, supported by the Innovation and Networking Fund of the Helmholtz Association (funding code ZT-0025).

We would like to thank all members of the Helmholtz Zentrum München, Institute of Epidemiology, and the Chair of Epidemiology at the University Hospital of Augsburg, who were involved in the planning and conduct of the study. Steering partners of the MI Registry, Augsburg, include the Chair of Epidemiology at the University Hospital of Augsburg and the Department of Internal Medicine I, Cardiology, University Hospital of Augsburg. Many thanks for their support go to the local health departments, the office-based physicians and the clinicians of the hospitals within the study area. Finally, we express our appreciation to all study participants.

This work was supported by the Helmholtz Zentrum München, German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education, Science, Research and Technology and by the State of Bavaria and the German Federal Ministry of Health. This research received also supported from the Faculty of Medicine, University of Augsburg, and the University Hospital of Augsburg, Germany. Since the year 2000, the collection of MI data has been co-financed by the German Federal Ministry of Health to provide population-based MI morbidity data for the official German Health Report (see www.gbe-bund.de).

References

Achebak, H., Devolder, D., and Ballester, J.: Trends in temperature-related age-specific and sex-specific mortality from cardiovascular diseases in Spain: a national time-series analysis, The Lancet Planetary Health, 3, e297–e306, https://doi.org/10.1016/S2542-5196(19)30090-7, 2019.

Armstrong, B.: Models for the Relationship Between Ambient Temperature and Daily Mortality, Epidemiology, 17, 624–631, https://doi.org/10.1097/01.ede.0000239732.50999.8f, 2006.

Ban, N., Caillaud, C., Coppola, E., et al.: The first multi-model ensemble of regional climate simulations at kilometer-scale resolution, part I: evaluation of precipitation, Climate Dynamics, 57, 275–302, https://doi.org/10.1007/s00382-021-05708-w, 2021.

Bayerische Landesamt für Umwelt: Lufthygienische Landesüberwachungssystem Bayern (LÜB), https://www.lfu.bayern.de/luft/immissionsmessungen/messwertarchiv/index.htm.

Bayerisches Landesamt für Statistik: GENESIS Datenbank, https://www.statistikdaten.bayern.de/genesis/online/.

Bhaskaran, K., Hajat, S., Haines, A., Herrett, E., Wilkinson, P., and Smeeth, L.: Short term effects of temperature on risk of myocardial infarction in England and Wales: time series regression analysis of the Myocardial Ischaemia National Audit Project (MINAP) registry, BMJ, 341, c3823, https://doi.org/10.1136/bmj.c3823, publisher: British Medical Journal Publishing Group Section: Research, 2010.

Bhaskaran, K., Armstrong, B., Hajat, S., Haines, A., Wilkinson, P., and Smeeth, L.: Heat and risk of myocardial infarction: hourly level case-crossover analysis of MINAP database, BMJ (Clinical research ed.), 345, e8050, https://doi.org/10.1136/bmj.e8050, 2012.

Bourdrel, T., Bind, M.-A., Béjot, Y., Morel, O., and Argacha, J.-F.: Cardiovascular effects of air pollution, Archives of Cardiovascular Diseases, 110, 634–642, https://doi.org/10.1016/j.acvd.2017.05.003, 2017.

Breitner, S., Wolf, K., Devlin, R. B., Diaz-Sanchez, D., Peters, A., and Schneider, A.: Short-term effects of air temperature on mortality and effect modification by air pollution in three cities of Bavaria, Germany: A time-series analysis, Science of The Total Environment, 485-486, 49–61, https://doi.org/10.1016/j.scitotenv.2014.03.048, 2014.

Cesaroni, G., Forastiere, F., Stafoggia, M., Andersen, Z. J., Badaloni, C., Beelen, R., Caracciolo, B., de Faire, U., Erbel, R., Eriksen, K. T., Fratiglioni, L., Galassi, C., Hampel, R., Heier, M., Hennig, F., Hilding, A., Hoffmann, B., Houthuijs, D., Jöckel, K.-H., Korek, M., Lanki, T., Leander, K., Magnusson, P. K. E., Migliore, E., Ostenson, C.-G., Overvad, K., Pedersen, N. L., J, J. P., Penell, J., Pershagen, G., Pyko, A., Raaschou-Nielsen, O., Ranzi, A., Ricceri, F., Sacerdote, C., Salomaa, V., Swart, W., Turunen, A. W., Vineis, P., Weinmayr, G., Wolf, K., de Hoogh, K., Hoek, G., Brunekreef, B., and Peters, A.: Long term exposure to ambient air pollution and incidence of acute coronary events: prospective cohort study and meta-analysis in 11 European cohorts from the ESCAPE Project, BMJ (Clinical research ed.), 348, f7412, https://doi.org/10.1136/bmj.f7412, 2014.

Chen, K., Wolf, K., Breitner, S., Gasparrini, A., Stafoggia, M., Samoli, E., Andersen, Z. J., Bero-Bedada, G., Bellander, T., Hennig, F., Jacquemin, B., Pekkanen, J., Hampel, R., Cyrys, J., Peters, A., and Schneider, A.: Two-way effect modifications of air pollution and air temperature on total natural and cardiovascular mortality in eight European urban areas, Environment International, 116, 186–196, https://doi.org/10.1016/j.envint.2018.04.021, 2018.

Chen, K., Breitner, S., Wolf, K., Hampel, R., Meisinger, C., Heier, M., von Scheidt, W., Kuch, B., Peters, A., Schneider, A., for the KORA Study Group, Peters, A., Schulz, H., Schwettmann, L., Leidl, R., Heier, M., and Strauch, K.: Temporal variations in the triggering of myocardial infarction by air temperature in Augsburg, Germany, 1987–2014, 40, 1600–1608, https://doi.org/10.1093/eurheartj/ehz116, 2019.

Commandeur, F., Slomka, P. J., Goeller, M., Chen, X., Cadet, S., Razipour, A., McElhinney, P., Gransar, H., Cantu, S., Miller, R. J. H., Rozanski, A., Achenbach, S., Tamarappoo, B. K., Berman, D. S., and Dey, D.: Machine learning to predict the long-term risk of myocardial infarction and cardiac death based on clinical risk, coronary calcium, and epicardial adipose tissue: a prospective study, Cardiovascular Research, 116, 2216–2225, https://doi.org/10.1093/cvr/cvz321, 2020.

Copernicus Global Land Service (CGLS): NDVI 1km V2.2 Global, https://land.copernicus.vgt.vito.be/geonetwork/srv/eng/catalog.search#/metadata/urn:cgls:global:ndvi_v2_1km.

Crameri, F., Shephard, G. E., and Heron, P. J.: The misuse of colour in science communication, 11, 5444, https://doi.org/10.1038/s41467-020-19160-7, 2020.

Davis, R. E., McGregor, G. R., and Enfield, K. B.: Humidity: A review and primer on atmospheric moisture and human health, Environmental Research, 144, 106–116, https://doi.org/10.1016/j.envres.2015.10.014, 2016.

Deutscher Wetterdienst (DWD): DWD Open Data Server, https://opendata.dwd.de/.

Feng, H., Zhao, X., Chen, F., and Wu, L.: Using land use change trajectories to quantify the effects of urbanization on urban heat island, Advances in Space Research, 53, 463–473, https://doi.org/https://doi.org/10.1016/j.asr.2013.11.028, 2014.

Gabriel, K. M. A. and Endlicher, W. R.: Urban and rural mortality rates during heat waves in Berlin and Brandenburg, Germany, Environmental Pollution, 159, 2044–2050, https://doi.org/10.1016/j.envpol.2011.01.016, 2011.

Havenith, G.: Temperature Regulation, Heat Balance and Climatic Stress, in: Extreme Weather Events and Public Health Responses, edited by Kirch, W., Bertollini, R., and Menne, B., pp. 69–80, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/3-540-28862-7_7, 2005.

Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in projections of regional precipitation change, Climate Dynamics, 37, 407–418, https://doi.org/10.1007/s00382-010-0810-6, 2011.

Helmholtz Zentrum München: KORA Datenbank (KORA.PASST), https://epi.helmholtz-muenchen.de/.

Holle, R., Happich, M., Löwel, H., and Wichmann, H.: KORA - A Research Platform for Population Based Health Research, Das Gesundheitswesen, 67, 19–25, https://doi.org/10.1055/s-2005-858235, 2005.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.-F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., and Yiou, P.: EURO-CORDEX: new high-resolution climate change projections for European impact research, Regional Environmental Change, 14, 563–578, https://doi.org/10.1007/s10113-013-0499-2, 2014.

Jacob, D., Teichmann, C., Sobolowski, S., Katragkou, E., Anders, I., Belda, M., Benestad, R., Boberg, F., Buonomo, E., Cardoso, R. M., Casanueva, A., Christensen, O. B., Christensen, J. H., Coppola, E., De Cruz, L., Davin, E. L., Dobler, A., Domínguez, M., Fealy, R., Fernandez, J., Gaertner, M. A., García-Díez, M., Giorgi, F., Gobiet, A., Goergen, K., Gómez-Navarro, J. J., Alemán, J. J. G., Gutiérrez, C., Gutiérrez, J. M., Güttler, I., Haensler, A., Halenka, T., Jerez, S., Jiménez-Guerrero, P., Jones, R. G., Keuler, K., Kjellström, E., Knist, S., Kotlarski, S., Maraun, D., van Meijgaard, E., Mercogliano, P., Montávez, J. P., Navarra, A., Nikulin, G., de Noblet-Ducoudré, N., Panitz, H.-J., Pfeifer, S., Piazza, M., Pichelli, E., Pietikäinen, J.-P., Prein, A. F., Preuschmann, S., Rechid, D., Rockel, B., Romera, R., Sánchez, E., Sieck, K., Soares, P. M. M., Somot, S., Srnec, L., Sørland, S. L., Termonia, P., Truhetz, H., Vautard, R., Warrach-Sagi, K., and Wulfmeyer, V.: Regional climate downscaling over Europe: perspectives from the EURO-CORDEX community, Regional Environmental Change, 20, 51, https://doi.org/10.1007/s10113-020-01606-9, 2020.

Khraishah, H., Alahmad, B., Ostergard, R. L., AlAshqar, A., Albaghdadi, M., Vellanki, N., Chowdhury, M. M., Al-Kindi, S. G., Zanobetti, A., Gasparrini, A., and Rajagopalan, S.: Climate change and cardiovascular disease: implications for global health, Nature Reviews Cardiology, pp. 1–15, https://doi.org/10.1038/s41569-022-00720-x, publisher: Nature Publishing Group, 2022.

545 Laverty, A. A., Goodman, A., and Aldred, R.: Low traffic neighbourhoods and population health, BMJ, 372, n443, https://doi.org/10.1136/bmj.n443, publisher: British Medical Journal Publishing Group Section: Editorial, 2021.

Madrigano, J., Mittleman, M. A., Baccarelli, A., Goldberg, R., Melly, S., von Klot, S., and Schwartz, J.: Temperature, myocardial infarction, and mortality: effect modification by individual- and area-level characteristics, Epidemiology (Cambridge, Mass.), 24, 439–446, https://doi.org/10.1097/EDE.0b013e3182878397, 2013.

Merz, B., Kreibich, H., and Lall, U.: Multi-variate flood damage assessment: a tree-based data-mining approach, Natural Hazards and Earth
550 System Sciences, 13, 53–64, https://doi.org/10.5194/nhess-13-53-2013, 2013.

Mustafić, H., Jabre, P., Caussin, C., Murad, M. H., Escolano, S., Tafflet, M., Périer, M.-C., Marijon, E., Vernerey, D., Empana, J.-P., and Jouven, X.: Main Air Pollutants and Myocardial Infarction: A Systematic Review and Meta-analysis, JAMA, 307, 713–721, https://doi.org/10.1001/jama.2012.126, 2012.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V.,
555 Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825–2830, 2011.

Pedregosa, F. et al.: Scikit-learn: software repository, https://scikit-learn.org/stable/.

Peters, A., von Klot, S., Heier, M., Trentinaglia, I., Hörmann, A., Wichmann, H. E., Löwel, H., and Cooperative Health Research in the Region of Augsburg Study Group: Exposure to traffic and the onset of myocardial infarction, The New England Journal of Medicine, 351,
560 1721–1730, https://doi.org/10.1056/NEJMoa040203, 2004.

Rai, M., Breitner, S., Wolf, K., Peters, A., Schneider, A., and Chen, K.: Impact of climate and population change on temperature-related mortality burden in Bavaria, Germany, Environmental Research Letters, 14, 124 080, https://doi.org/10.1088/1748-9326/ab5ca6, publisher: IOP Publishing, 2019.

Rajagopalan, S., Al-Kindi, S. G., and Brook, R. D.: Air Pollution and Cardiovascular Disease: JACC State-of-the-Art Review, Journal of the
565 American College of Cardiology, 72, 2054–2070, https://doi.org/10.1016/j.jacc.2018.07.099, 2018.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195–204, https://doi.org/10.1038/s41586-019-0912-1, 2019.

Roth, G. A.: Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019, 76, 40, 2020.

Schmidt, S., Hendricks, V., Griebenow, R., and Riedel, R.: Demographic change and its impact on the health-care budget for heart failure
570 inpatients in Germany during 1995–2025, Herz, 38, 862–867, https://doi.org/10.1007/s00059-013-3955-3, 2013.

Schwartz, J., Samet, J. M., and Patz, J. A.: Hospital Admissions for Heart Disease: The Effects of Temperature and Humidity, Epidemiology, 15, 755–761, https://doi.org/10.1097/01.ede.0000134875.15919.0f, 2004.

Sewe, M. O., Tozan, Y., Ahlm, C., and Rocklöv, J.: Using remote sensing environmental data to forecast malaria incidence at a rural district hospital in Western Kenya, Scientific Reports, 7, 2589, https://doi.org/10.1038/s41598-017-02560-z, 2017.

575 Sieck, K., Nam, C., Bouwer, L. M., Rechid, D., and Jacob, D.: Weather extremes over Europe under 1.5 °C and 2.0 °C global warming from HAPPI regional climate ensemble simulations, https://doi.org/10.5194/esd-2020-4, 2020.

Statistisches Bundesamt: 14. Koordinierte Bevölkerungsvorausberechnung - Basis 2018, https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Bevoelkerungsvorausberechnung/aktualisierung-bevoelkerungsvorausberechnung.html.

Sun, Z., Chen, C., Xu, D., and Li, T.: Effects of ambient temperature on myocardial infarction: A systematic review and meta-analysis, Environmental Pollution (Barking, Essex: 1987), 241, 1106–1114, https://doi.org/10.1016/j.envpol.2018.06.045, 2018.

Tamarappoo, B. K., Lin, A., Commandeur, F., McElhinney, P. A., Cadet, S., Goeller, M., Razipour, A., Chen, X., Gransar, H., Cantu, S., Miller, R. J., Achenbach, S., Friedman, J., Hayes, S., Thomson, L., Wong, N. D., Rozanski, A., Slomka, P. J., Berman, D. S., and Dey, D.: Machine learning integration of circulating and imaging biomarkers for explainable patient-specific prediction of cardiac events: A prospective study, Atherosclerosis, 318, 76–82, https://doi.org/10.1016/j.atherosclerosis.2020.11.008, 2021.

The Eurowinter Group: Cold exposure and winter mortality from ischaemic heart disease, cerebrovascular disease, respiratory disease, and all causes in warm and cold regions of Europe, The Lancet, 349, 1341–1346, https://doi.org/10.1016/S0140-6736(96)12338-2, 1997.

Tunstall-Pedoe, H., Kuulasmaa, K., Amouyel, P., Arveiler, D., Rajakangas, A. M., and Pajak, A.: Myocardial infarction and coronary deaths in the World Health Organization MONICA Project. Registration procedures, event rates, and case-fatality rates in 38 populations from 21 countries in four continents., Circulation, 90, 583–612, https://doi.org/10.1161/01.CIR.90.1.583, 1994.

Turnock, S. T., Allen, R. J., Andrews, M., Bauer, S. E., Deushi, M., Emmons, L., Good, P., Horowitz, L., John, J. G., Michou, M., Nabat, P., Naik, V., Neubauer, D., O'Connor, F. M., Olivié, D., Oshima, N., Schulz, M., Sellar, A., Shim, S., Takemura, T., Tilmes, S., Tsigaridis, K., Wu, T., and Zhang, J.: Historical and future changes in air pollutants from CMIP6 models, Atmospheric Chemistry and Physics, 20, 14 547–14 579, https://doi.org/10.5194/acp-20-14547-2020, publisher: Copernicus GmbH, 2020.

Vanos, J. K., Baldwin, J. W., Jay, O., and Ebi, K. L.: Simplicity lacks robustness when projecting heat-health outcomes in a changing climate, Nature Communications, 11, 6079, https://doi.org/10.1038/s41467-020-19994-1, 2020.

Vos et al.: Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015, The Lancet, 388, 1545–1602, https://doi.org/10.1016/S0140-6736(16)31678-6, 2016.

Wagenaar, D., de Jong, J., and Bouwer, L. M.: Multi-variable flood damage modelling with limited data using supervised learning approaches, Natural Hazards and Earth System Sciences, 17, 1683–1696, https://doi.org/10.5194/nhess-17-1683-2017, 2017.

Wagenaar, D., Hermawan, T., Homberg, M. J. C., Aerts, J. C. J. H., Kreibich, H., Moel, H., and Bouwer, L. M.: Improved Transferability of Data-Driven Damage Models Through Sample Selection Bias Correction, Risk Analysis, 41, 37–55, https://doi.org/10.1111/risa.13575, 2021.

Wolf, K., Schneider, A., Breitner, S., von Klot, S., Meisinger, C., Cyrys, J., Hymer, H., Wichmann, H. E., and Peters, A.: Air Temperature and the Occurrence of Myocardial Infarction in Augsburg, Germany, Circulation, 120, 735–742, https://doi.org/10.1161/CIRCULATIONAHA.108.815860, 2009.

Wolf, K., Hoffmann, B., Andersen, Z. J., Atkinson, R. W., Bauwelinck, M., Bellander, T., Brandt, J., Brunekreef, B., Cesaroni, G., Chen, J., Faire, U. d., Hoogh, K. d., Fecht, D., Forastiere, F., Gulliver, J., Hertel, O., Hvidtfeldt, U. A., Janssen, N. A. H., Jørgensen, J. T., Katsouyanni, K., Ketzel, M., Klompmaker, J. O., Lager, A., Liu, S., MacDonald, C. J., Magnusson, P. K. E., Mehta, A. J., Nagel, G., Oftedal, B., Pedersen, N. L., Pershagen, G., Raaschou-Nielsen, O., Renzi, M., Rizzuto, D., Rodopoulou, S., Samoli, E., Schouw, Y. T. v. d., Schramm, S., Schwarze, P., Sigsgaard, T., Sørensen, M., Stafoggia, M., Strak, M., Tjønneland, A., Verschuren, W. M. M., Vienneau, D., Weinmayr, G., Hoek, G., Peters, A., and Ljungman, P. L. S.: Long-term exposure to low-level ambient air pollution and incidence of stroke and coronary heart disease: a pooled analysis of six European cohorts within the ELAPSE project, The Lancet Planetary Health, 5, e620–e632, https://doi.org/10.1016/S2542-5196(21)00195-9, publisher: Elsevier, 2021.

Zhang, D.-L., Shou, Y.-X., and Dickerson, R. R.: Upstream urbanization exacerbates urban heat island effects, Geophysical Research Letters, 36, https://doi.org/10.1029/2009GL041082, 2009.

Zhang, K., Li, Y., Schwartz, J. D., and O'Neill, M. S.: What weather variables are important in predicting heat-related mortality? A new application of statistical learning methods, Environmental Research, 132, 350–359, https://doi.org/10.1016/j.envres.2014.04.004, 2014.

Zinszer, K., Verma, A. D., Charland, K., Brewer, T. F., Brownstein, J. S., Sun, Z., and Buckeridge, D. L.: A scoping review of malaria
620   forecasting: past work and future directions, BMJ Open, 2, e001 992, https://doi.org/10.1136/bmjopen-2012-001992, 2012.