



OPEN

Modeling health outcomes of air pollution in the Middle East by using support vector machines and neural networks

Ayesha¹, Muhammad Noor-ul-Amin², Olayan Albalawi³, Nadia Mushtaq⁴,
Emad E. Mahmoud⁵, Uzma Yasmeen^{1,6} & Muhammad Nabi⁷✉

This study investigates the impact of air pollution on health outcomes in Middle Eastern countries, a region facing severe environmental challenges. As such, these are important in an effort to add up to policy-level as well as interventional changes that can be put in practice in the area of public health. Numeration analysis and association with health parameters was carried out by using Analytical tools such as, AIR Data, ARIMA, ANN, SVM and Exponential smoothing. Amongst the models, Support Vector Machine came again on top, with high accuracy yielding Mean Absolute Percentage Error of approximately 1%. Mortality of Air pollution in Qat from the case of Mortality of Air Pollution in Qatar is 959 while Auto regressive Integrated Moving average is 11.096, Exponential Smoothing 9.892 and Artificial Neural Networks are the source of inspiration for the development of this paper 4.61. The above perceptions indicate that there is need to adapt modeling strategies depending on the context and establish that it is possible to implement ML models in public health planning basket. This paper publishes the methodological frameworks for the purpose of modeling and analysis of the EHDs and serves as policy prescription for the policy makers to intending to reduce the effects of air borne pollution on health.

Keywords Disability-adjusted life years, Air pollution, Health outcome, Mortalities, Middle East

Middle East region there has been economic development and expansion of built up areas, as well as raised level of industrial activity and motorised traffic. While such developments have helped in the fight and growth of the regions development story the quality of air pollution has over time threatened the health of the people. It becomes easier to deal with health risks posed by pollution, if the relation between air quality and human welfare is properly understood in the Middle East. This research paper aims at undertaking the following exploratory research question in an attempt to establish the correlation between various types of pollution and public health in the Middle East region. Specifically, it intend to forecast and analyse the impact of pollution in relation to health drawing attention to the critical decade of 2020–2030. The research explores the complex aspects of this issue by examining various pollutants, their origins, and the socio-economic factors that affect population vulnerability in the region. Pollution, in this context, refers to the emission of substances or energy into the atmosphere, leading to harmful impacts on the environment, including risks to human health, damage to ecosystems, and interference with environmental functions and other legitimate uses. This definition is provided by the European Environmental Agency^{1,2}. Land pollution, as defined by the Texas Disposal System, pertains to the degradation of the surfaces of the Earth's land, encompassing the above ground levels and underground level³. The contamination of water sources by substances renders the water unsuitable for various activities such as cooking, drinking, swimming, cleaning, and more. This information is provided by Harvard T.H. Chan School of Public Health⁴. Air pollution is a perilous form of environmental contamination. As outlined by the National Institute of Environmental Health Sciences, it is described as a combination of harmful substances originating from both human-made and natural sources (National Institute of Environmental Health Science, NIH⁵.

¹The University of Lahore, Lahore, Pakistan. ²Department of Statistics, COMSATS University Islamabad-Lahore Campus, Lahore, Pakistan. ³Department of Statistics, Faculty of Science, University of Tabuk, Tabuk, Saudi Arabia. ⁴Forman Christian College (A Chartered University), Lahore, Pakistan. ⁵Department of Mathematics and Statistics, Collage of Science, Taif University, P.O. Box 11099, 21944 Taif, Saudi Arabia. ⁶Brock University, St. Catharines, Canada. ⁷Khost Mechanics Institute, Khost, Afghanistan. ✉email: m.nabi.com1990@gmail.com

Nature releases hazardous substances into the air, including gases produced from the decay of organic matter in soils, gases and ashes from volcanic explosions, and fumes from wildfires, often ignited by human activities. This information is sourced from the National Institute of Health Sciences⁶.

Ambient Particulate Matter Pollution, abbreviated as PMP, comprises specific solid and liquid particles present in the air, including dirt, soot, smoke, and dust. These particles do not originate from a single source but rather form a mixture of diverse particles. They vary in composition, size, and shape, collectively found in the air. HAP is produced through the ignition of household fuels, resulting in indoor air pollution and adding to the overall burden of outdoor air pollution. This insight is provided by the World Health Organization⁷. As stated by the “Pan American Health Organization (PAHO)”, the ineffective burning of fuels stands out as one of the primary factors contributing to Disability-Adjusted Life Years (DALYs) and fatalities worldwide, Pan American Health Organization⁸. Ozone, a gas composed of three oxygen atoms, exists naturally and is also produced through human activities. It is highly reactive and is found in both the troposphere and stratosphere, with the majority located in the stratosphere. The upper stratosphere contains a protective layer of ozone that shields the Earth from harmful ultraviolet radiations emitted by the sun. An atmosphere contaminated by pollution poses a threat to everyone, with infants being particularly vulnerable. The number of infant deaths resulting from air pollution is staggering, contributing significantly to global child mortality rates annually. This issue is especially prominent in underdeveloped and developing countries. The primary culprits behind these life-threatening diseases are air pollution, whether indoors or outdoors. Notably, even in developed nations such as China, the toll of premature mortalities due to air pollution remains alarmingly high⁹. In this analysis, they assess the effectiveness of employing different models¹⁰. MLM has been useful to conduct in-depth studies on the suitability of various existing architectures for diverse forecasting tasks¹¹. A comprehensive examination is conducted, examining all air pollutant factors by correlating these elements from top to bottom¹². The SARIMA model is employed to project future concentrations of PM2.5, anticipating an increase in the number of PM2.5 particles in the upcoming year. The forecast provides both the minimum and maximum predictions, ranging over 100 $\mu\text{g}/\text{m}^3$. In previous literature, the Autoregressive Integrated Moving Average (ARIMA) model has been utilized to predict the prevalence of diseases due to its ability and ease in elucidating dataset^{13,16–18}. Holt's introduced Winter exponential smoothing (ES) that has found application in modeling and forecasting various issues, including but not limited to electricity consumption¹⁶. Artificial Neural Networks (ANN) were utilized to predict the daily cases of COVID-19 in various countries, including China, Iran, Italy, Japan, South Africa, Singapore, and the USA¹⁵. In forecasting the outbreak of COVID-19, Artificial Neural Networks (ANN) have been sparingly employed to predict death and recovery cases¹⁹. The study utilizes an ANN to accurately predict COVID-19 confirmed cases and deaths, demonstrating the efficacy of the ANN model in forecasting future trends and emphasizing the importance of stringent control measures to curb the pandemic's spread²⁰. ANN proved to be dependable in forecasting increases in timber prices, yet encountered challenges when faced with unforeseen surges. Research indicated their superiority over traditional methods, highlighting the necessity for enhancements, such as incorporating exogenous variables and optimizing model structures, to achieve more accurate predictions of timber prices²¹. The research, which conducted a comparison of statistical and machine learning models (ARIMA, SARIMA, NAR, LSTM) based on Mean Absolute Percentage Error (MAPE), demonstrated that, in general, open-loop models outperformed closed-loop ones, although exceptions were observed at specific stations. While there wasn't a clear optimal method identified for open loops, machine learning techniques, particularly Long Short-Term Memory (LSTM) and Neural Autoregression (NAR-NN), exhibited superior performance compared to statistical methods in closed-loop setups. The study underscored a notable correlation between MAPE and Relative Standard Deviation (RSD) for both loop types²². The study utilized artificial neural networks for mortality forecasting in the Kurdistan Region, citing their superior predictive accuracy. It underscored the importance of data quality and recommended the adoption of the KRG-HIS program to enhance health data collection. The study highlighted that widespread implementation of such a system could add to better health planning and more accurate mortality forecasts²³.

This study provides a complete comparative analysis of classical statistical models and advanced machine learning techniques for predicting air pollution levels and the health impacts linked with them. The study recognizes the strengths and weaknesses of each approach by evaluating the performance of models such as ARIMA, ES, SVM, and ANN. This benchmarking is decisive for selecting the most appropriate model for specific forecasting tasks. The outcomes offer valuable insights into model performance, practical guidance for model selection and support for informed decision-making in public health and policy. This study contributes to environmental health research by highlighting the potential of advanced machine learning techniques to improve predictive accuracy and reliability.

Data source and methodology

This research employed information sourced from “the Institute of Health Metrics and Evaluation GBD” websites (1990–2019) focusing on eight Middle East Arab States: Syrian Arab Republic, Iraq, Kuwait, United Arab Emirates, Iran, Lebanon, Saudi Arabia, and Qatar.

The research employs the aforementioned dataset, covering the years 1990–2019 for Middle Eastern countries, consisting of 240 observations (eight countries with 30 values each). In applying machine learning techniques, 80% of the data (1990–2013) was used for model training, while the remaining 20% (2013–2019) was reserved for model testing. Forecasting plots were created to visually represent the predicted values generated by the models. The Mean Absolute Percentage Error (MAPE) was utilized for model validation. Both traditional statistical methods and machine learning algorithms, including ARIMA, Exponential Smoothing (ES), Support Vector Machines (SVM), and Neural Networks (NN), were employed in constructing these forecasting models.

ARIMA model

The ARIMA modeling approach with Box Jenkin Methodology offers the conventional framework for time series analysis. Its strengths lie in flexibility and automatic model selection through functions like `auto.arima()` in R and the interpretability of its components. The steps for ARIMA model are illustrated in Fig. 1.

The ARIMA is denoted as “ARIMA(p, d, q)”. The expression is as follows:

$$\hat{Y}'_t = \mu + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{(t-q)} + \varepsilon_t \quad (1)$$

where, d : the integrated value, p : the order of autoregressive process, μ : mean, q : the order of moving average process, ϕ : the autoregressive operator, y_t : the time series, ε : the random error, y'_t : the first derivative of time series, θ : the moving average operator.

Exponential smoothing method

The Single Exponential Smoothing (ES) method was utilized to generate forecasts for a time series dataset, selected due to the absence of any seasonal patterns and the lack of a significant upward or downward trend in the data. This approach provides a way to estimate the current level of the time series using a single smoothing parameter, α , which determines the weight assigned to the most recent observation. The value of α varies between 0 and 1. The method uses exponentially decreasing weights and also incorporates weights moving average; the calculation can be explained with the help of the following formula.

$$s_t = \alpha x_t + (1 - \alpha)s_{t-1} = s_{t-1} + \alpha(x_t - s_{t-1}) \quad (2)$$

t the time, s_t : c : the quantity of interest, y_t : the target statistic, α : the smoothing factor, s_{t-1} : the previous smoothed statistic. In the course of our analysis, we chose Holt's Exponential Smoothing (ES) method because in the observed data samples no periodic relations were identified. The goal of employing Holt's Exponential Smoothing in our analysis is to capture and understand the fundamental patterns in the data, ultimately leading to more precise predictions and better-informed decision-making: the time, s_t : the smoothed statistic, and α : the smoothing factor, s_{t-1} : the previous smoothed statistic. However, there is conflict in the data; therefore, in the current analysis, we opted for Holt's Exponential Smoothing (ES) method since there was no evidence of a seasonality present in the data. The objective to apply Holt's Exponential Smoothing as the method of our analysis will allow to unearth the underlying characteristics of the patterns and generate more accurate forecasts and effective decisions. The formula is expressed as: $s_1 = x_1$ and $b_1 = x_1 - x_0$ for $t > 1$,

$$s_t = \alpha x_t + (1 - \alpha)(s_{t-1} + b_{t-1}) \quad (3)$$

$$\beta_t = \beta(s_t - s_{t-1}) + (1 - \beta)b_{t-1} \quad (4)$$

here, t : the time, α : the smoothing factor, b_t : the trend best estimate at time t , s_{t-1} : the previous smoothed statistic, s_t : the smoothed statistic, β : the trend smoothing factor; $0 < \beta < 1$

Support Vector Machine Regression

The ordinary Support Vector Machine (SVM) Regression is a popular technique in machine learning best used on time series data. SVMs are particularly good at working with large and complex data and have following characteristics: Time series data analysis. One of the significant advantages is that they do not tend to overfit because they are not influenced by outliers: they work with support vectors—the most significant points. This characteristic renders SVMs flexible and able to generalize if given new data, this boost up the accuracy in different applications.

SVR was implemented using a linear kernel which is defined as $(K(x, x') = x^T x')$. This choice of kernel allows for effective modeling of linear relationships within the data. Alternative kernels such as polynomial or radial basis function (RBF) could be employed for complex data. The regularization parameter (C) was set to 1.0 that helps balance the trade-off between minimizing training error and maintaining model simplicity to

BOX JENKIN METHODOLOGY

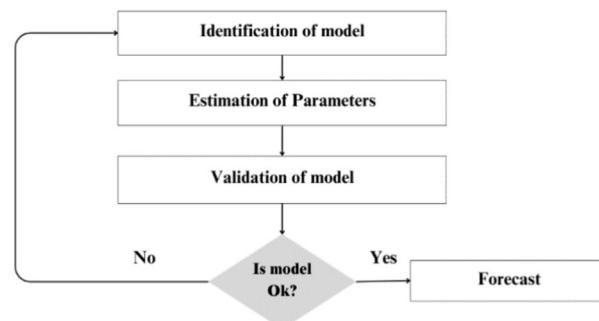


Fig. 1. Box–Jenkin methodology.

prevent overfitting. The ϵ -insensitive loss function was applied to ignore errors within a margin ϵ . It focuses only on deviations that exceed this threshold. The SVR objective function is expressed to minimize the succeeding objective function:

$$\text{Objective: } \frac{1}{2} |w|^2 + C \sum_{i=1}^n \xi_i$$

subject to the constraints:

$$y_i - (w^T \phi(x_i) + b) \leq \epsilon + \xi_i$$

$$(w^T \phi(x_i) + b) - y_i \leq \epsilon + \xi_i$$

where ξ_i are slack variables, w represents the weight vector, $(\phi(x_i))$ denotes the kernel-transformed feature space, and b is the bias term. Predictions are obtained using the function:

$$f(x) = \sum_{i=1}^n \alpha_i K(x_i, x) + b$$

where (α_i) are the Lagrange multipliers from the training phase and $(K(x_i, x))$ is the kernel function applied to the test input (x) . Predictions were made on the test set to assess the model's performance, and a forward-looking forecasting loop was incorporated. This loop generates predictions for future years based on the last three observations in the processed data. The Fig. 2 presented the structure of SVM model.

Artificial neural networks

Artificial neural networks is a sequence of algorithms inspired by the cognitive processes of the human brain. They are at the lead of this predictive revolution. These networks showcase an amazing ability to approximate all functions without requiring extensive knowledge about the data's particulars. ANNs are extensively applied in time series forecasting. Neural networks play a crucial role in shaping the future of predictions and form the cornerstone of deep learning, a subfield of machine learning inspired by the intricate structure of the human brain. These networks as shown in figure three, take data through them, and are capable of independent training to detect particular patterns. This unique ability allows them to predict the output for the new similar data sets, emulate the brain's multiple steps to analyze the information. The methodology adopted in the present study can be described in the following manner:

Data preprocessing

The first resolute within the undertaken methodology is the preliminary analysis of the time series to ensure their compatibility with the input data for the neural network model. Next to this, the data is first ordered chronologically to preserve the sequential structure which is required when undertaking time series analysis. This alters the range of the target variable in order that it may be normally distributed. This normalization process is performed through reducing the mean to zero and standard deviation to one of the data it is used to improve the stability and the performance of the neural network while training.

Dataset construction

They are expressed in the form of data for training a neural network based on the sequences of historical observation expanded in time. This involves a creation of tuples each of which has the past time steps as an input together with the value that has to be predicted on the next time step as an output. It is favorable to the model to use a sequence-based way to analyze the data, as temporal dependencies and temporal patterns can be learned using the data. Either the value of 't' or the length of the input sequence is a dire parameter that determines the extent of the temporal information that the models can learn.

SVM Structure

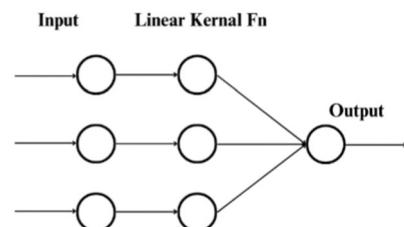


Fig. 2. SVM model structure.

Neural network architecture

Specifically, with the view of operationalizing the proposed forecasting model, a feedforward neural network has been used. The network includes multiple layers: The network can contain multiple levels, such as the system and the subjects filling posts in it.

A typical structure of a Neural Network (NN) design comprises of an input layer, hidden layers and an output layer. The input layer receives the input data and each of the neurons in this layer is in some way related to the feature present in the data. The latent layers are the layers that come in between the input and output layers and these neurons work on the inputs with applying weights, and sigmoid or tanh to put non-linearity in the model. All the results and decisions are given in the output layer and each neuron of the output layer contributes to the special possibility. While in forward propagation the network is in a position to make certain predictions and Sugimura, while using backpropagation, the weights of the network are in a position to be modified depending on the errors. The structure of the NN can be basic and have only the one hidden layer, or can be multilayer which consists from many layers in what is called Deep Learning Network with numerous layers or can have the adjusted layers as in convolutional or Recurrent NNs. Likewise, the structure of NN that Fig. 3 depicted was about organizing relation.

Model training and evaluation

Neural network learning occurs on a sample of the data referred as the training set while the performance is evaluated on another sample known as the testing set. To measure the accuracy of the created model, Performance metrics named Mean Absolute Percentage Error (MAPE) is used. MAPE computes the level of prediction accuracy in proportional levels.

Forecasting and visualization

Having derived the model, it is then applied to make forecast for subsequent periods. In this method, the subsequent values are predicted from the data set and this is done repeatedly. These forecasts are linked with historical data to make visualization of the performance of the model possible. To illustrate the actual data and forecasted values, plots are always created. This in turn allows consideration of the predictive accuracy of the model.

It is a flexible and hierarchical structure that enables the training of the neural network for the patterns which exist in the time series data for robust predictions. Its ability to learn different and more accurate relationships within the data makes it more reliable for forecasting in the time series analysis. This has been made clear by the following representation of the Neural Network model;

$$y_i = f(y_{t-i}) + \epsilon_t \quad (5)$$

where, $y_{t-i} = (y_{t-1}, y_{t-2}, \dots, y_{t-8})'$ comprises a vector that includes past values of the series, and f represents ANN including 2 hidden nodes. The error system $\{\epsilon_t\}$ is presumed to exhibit homoscedasticity. ϵ_{T+1} is a stochastic draw from the error distribution at time $T+1$, then $y_{T+1}^* = f(y_{T+1}) + \epsilon_{T+1}^*$ represents a potential realization from the predicted distribution. for y_{T+1} .

$$y_{T+1}^* = (y_{T+1}^*, y_T, \dots, y_{T-6})'$$

$$y_{T+2} = f(y_{T+1}^*) + \epsilon_{T+1}^* \quad (6)$$

y_{T+2} is the value of the series at time $T+2$. f is a function that describes the relationship between the past values and the future values of the series. y_{T+1}^* is a modified or transformed value of the series at time $T+1$. ϵ_{T+1}^* is an error term at time $T+1$. This method allows a scholar to simulate upcoming sample paths iteratively. Through the repetitive simulation of these paths the researcher acquires an understanding of the distribution for all forthcoming values leveraging the information provided by the fitted ANN.

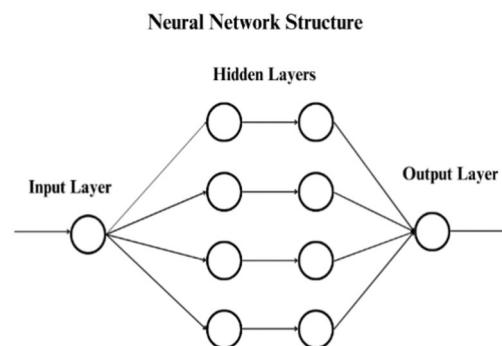


Fig. 3. Neural network nodes.

Descriptive study

The time series analysis includes variables such as fatalities attributed to APM, HAP and AOP as well as DALYs associated with PMP, AOP and HAP. The descriptive overview of these variables for specific Middle East countries is presented in Table 1.

Table 1 offers a detailed overview of descriptive statistics for time-series variables related to pollution in various Middle Eastern countries, including Mortalities and Disability-Adjusted Life Years (DALYs) associated with Particulate Matter Pollution (PMP), Ambient Outdoor Pollution (AOP), and Household Air Pollution (HAP). In the United Arab Emirates, the mean Mortality from PMP is 1418.2, with a standard deviation of 761.60, reflecting moderate variability. The range of Mortality spans from a minimum of 587.9 to a maximum of 3252.0. In the Syrian Arab Republic, the mean Mortality resulting from HAP is 173.9, with a high standard deviation of 1066.18, indicating significant variability. The Mortality range for AOP is between 125.8 and 229.9. In Iraq, the mean Mortality due to PMP is notably high at 20,391, with a substantial standard deviation of 2594.06, underscoring considerable variability. The Mortality for AOP ranges from a minimum of 111.3 to a maximum of 259.4. Kuwait exhibits a mean Mortality from PMP of 905.6, with a lower standard deviation of 319.68, indicating less variability. The range of Mortality in Kuwait extends from a minimum of 4,646 to a maximum of 24,666, showing a narrower range compared to other countries. This analysis reveals distinctive patterns in mortality rates due to different types of pollution across the region. For the Islamic Republic of Iran, mean Mortality resulting from PMP stand at 35,029 that showcases a higher impact and the standard deviation of 3782.674 signifying a significant degree of variability. The minimum of 495.2 and maximum of 1794.2 highlights the broad spectrum of Mortality resulting from AOP.

In Lebanon, the mean values for Mortality resulting from PMP, AOP and HAP stand at 2636, 71.01 and 2692 respectively. The corresponding DALYs means are 72,760, 1231.61 and 73,737. Lebanon shows relatively moderate variability. Saudi Arabia, on the other hand, portrays higher mean values across all categories such as 13,036 for Mortality resulting from PMP and 13,182 for Mortality resulting from HAP. The DALYs mean values are also substantial indicating a significant impact on public health. Saudi Arabia exhibits moderate to high variability especially in Mortality resulting from PMP and DALYs related to PMP and HAP. These findings collectively underscore the diverse impacts of different pollutants on public health, considering both Mortalities and DALYs.

Results

The examination of Mortalities and DALYs attributed to APM, HAP and AOP in Middle East countries that includes Syria, Iraq, Kuwait, United Arab Emirates, Iran, Lebanon, Saudi Arabia and Qatar encompassed the years 1990 to 2019. The study employed ARIMA, ES, SVM and Neural Network methodologies for forecasting purposes. The time series analysis focused on variables such as Mortalities due to APM, HAP and AOP as well as DALYs related to these three pollution categories. The selected countries' data from 1990 to 2019 served as the foundation for employing these predicting techniques, providing a comprehensive understanding of the trends and potential future trajectories in pollution-related health outcomes in the Middle East.

Time series plot

Time series plots were generated for the period spanning 1990 to 2019 to explore data patterns.

The Fig. 4a,e Mortality resulting from PMP and HAP show a rising pattern in selected middle east countries. Figure 4b,f DALYs resulting from HAP and PMP show a rising pattern in selected middle east countries except for Iran where it shows u-shape pattern. Figure 4c,d mortalities and DALYs resulting from AOP show a rising pattern in selected middle east countries with Iran having higher mortalities and DALYs resulting from AOP and subtle increase from 2005–2019.

Model summaries

ARIMA and Exponential Smoothing models were compared using the Bayesian Information Criterion (BIC). The results are presented in Table 2 for all the countries. In the context of time series analysis, this criterion helps in evaluating and selecting the model that best balances complexity and accuracy, ultimately guiding the choice of the most appropriate forecasting model for each country.

The ARIMA(p, d, q) model was selected from the Box-Jenkins Methodology. Model was selected based on smaller BIC score. ES performs better than ARIMA having lower BIC score (see Table 2). Across all countries and metrics, the ES model consistently achieves lower BIC scores compared to the ARIMA model. This indicates that, for the given data, the ES model provides a better fit than the ARIMA model for predicting both DALYs and mortalities. The analysis shows that the ES model is generally superior to the ARIMA model in predicting DALYs and mortalities across the examined Middle Eastern countries. This insight can guide future health data modeling and improve the accuracy of health outcome predictions, ultimately aiding in better health policy formulation and implementation.

Visualization for ARIMA

The forecasting charts for selected Middle East countries is shown below:

The projected values for next 10 years were visualized for ARIMA. In Fig. 5a,e depicting mortality resulting from PMP and HAP, an ascending trend is observed in Iraq, suggesting an increase in the number of mortalities each year for the forecasted period. Conversely, all other figures exhibit relatively constant forecasts, indicating a consistent number for the next 10 years. The Fig. 5c,d, illustrating mortalities and DALYs resulting from AOP, reveal an ascending trend in Kuwait and Iran, signaling a yearly increase in both Mortalities and DALYs resulting from AOP. In Fig. 5b,f representing DALYs resulting from HAP and PMP an upward trend is obvious across all countries that indicates a yearly rise in DALYs.

Descriptive statistics	Mortality resulting from PMP	Mortality resulting from AOP	Mortality resulting from HAP	DALYs resulting from PMP	DALYs resulting from AOP	DALYs resulting from HAP
United Arab Emirates						
N	30	30	30	30	30	30
Minimum	587.9	19.80	602.3	24,480	693.2	24,984
1st quartile	862.6	31.92	885.9	33,543	1124.6	34,362
Median	1130.8	50.11	1167.5	43,967	1763.3	45,255
Mean	1418.2	66.64	1466.7	60,850	2552.0	62,706
3rd quartile	1848.6	88.88	1912.1	83,317	3582.7	85,878
Maximum	3252.0	149.10	3360.7	143,110	5844.7	147,250
SD	761.60	41.63	791.31	36,055.51	1699.103	37,268
Variance	580,036.9	1732.74	626,176.7	1.3e+09	2,886,950	1,388,903,593
Syrian Arab Republic						
N	30	30	30	30	30	30
Minimum	7347	125.8	7477	241,204	2662	244,039
1st quartile	7899	146.1	8020	272,704	3171	275,430
Median	8226	160.1	8336	300,410	3503	303,512
Mean	8731	173.9	8866	293,142	3709	296,033
3rd quartile	10,002	205.7	10,167	313,594	4351	316,877
Maximum	10,491	229.9	10,671	332,398	4729	334,622
SD	1066.18	33.3016	1088.798	26,553.14	643.9592	26,558.19
Variance	1,136,739	1108.997	1,185,482	705,069,242	414,683.4	705,337,238
Iraq						
N	30	30	30	30	30	30
Minimum	15,871	111.3	15,942	695,165	2203	696,552
1st quartile	18,757	144.4	18,854	785,066	3043	787,828
Median	19,920	171.9	20,038	795,065	3669	797,917
Mean	20,391	175.7	20,511	787,507	3676	790,009
3rd quartile	22,749	202.0	22,880	805,383	4231	807,483
Maximum	25,442	259.4	25,626	816,359	5287	819,081
SD	2594.06	40.1023	2622.97	28,603.58	842.836	29,004.17
Variance	6,729,150	1608.195	6,879,957	818,164,584	710,372	841,242,151
Kuwait						
N	30	30	30	30	30	30
Minimum	451.9	4.646	455	18,481	92.14	18,543
1st quartile	639	8.038	644.4	23,348	147.31	23,446
Median	881.1	11.151	888.3	30,816	196.36	30,943
Mean	905.6	12.426	913.7	32,653	214.41	32,793
3rd quartile	1155.6	16.449	1165.9	40,301	272.59	40,483
Maximum	1526.8	24.666	1543.1	54,914	408.84	55,184
SD	319.679	5.80753	323.344	10,876.24	89.4481	10,932.53
Variance	102,194	33.7274	104,551	118,292,567	8000.97	119,520,250
Islamic Republic of Iran						
N	30	30	30	30	30	30
Minimum	29,583	495.2	29,944	1,112,085	10,936	1,124,944
1st quartile	32,049	597.08	32,490	1,142,753	12,674	1,158,268
Median	34,626	44.7	35,246	1,164,769	16,790	1,177,956
Mean	35,029	967.4	35,749	1,165,695	18,722	1,179,616
3rd quartile	37,045	1206.8	37,930	1,174,518	22,798	1,189,787
Maximum	41,839	1794.2	43,203	1,315,278	31,852	1,323,247
SD	3782.674	423.769	4096.866	41,543.53	6849.109	40,474.5
Variance	14,308,621	179,580	16,784,310	1,725,864,909	46,910,294	1,638,185,141
Lebanon						
N	30	30	30	30	30	30
Minimum	2020	46.93	2057	60,937	905.4	61,729
1st quartile	2119	52.36	2160	64,448	988.1	65,272
Median	2436	72.48	2494	67,041	1259.5	67,912
Mean	2636	71.01	2692	72,760	1231.61	73,737

Continued

Descriptive statistics	Mortality resulting from PMP	Mortality resulting from AOP	Mortality resulting from HAP	DALYs resulting from PMP	DALYs resulting from AOP	DALYs resulting from HAP
3rd quartile	3267	89.35	3340	84,711	481.3	85,879
Maximum	3423	93.11	3493	88,591	1554.1	89,749
SD	544.836	17.59203	557.988	10,210.22	240.756	10,371.3
Variance	296,846	309.4797	311,351	104,248,505	57,963.4	107,563,862
Saudi Arabia						
N	30	30	30	30	30	30
Minimum	9105	162.2	9196	354,261	3238	356,080
1st quartile	10,587	194.2	10,715	378,835	4005	381,465
Median	12,906	225	13,050	448,078	4823	451,182
Mean	13,036	225.6	13,182	465,874	4972	469,101
3rd quartile	15,135	249.1	15,288	533,069	5656	536,605
Maximum	17,832	310.1	18,036	654,330	7441	659,239
SD	2734.66	42.0426	2761.734	94,490.13	1212.26	95,272.83
Variance	7,478,361	1767.58	7,627,175	8,928,385,492	1,469,566	9,076,911,465
Qatar						
N	30	30	30	30	30	30
Minimum	174.6	2.145	175.9	7409	53.47	7441
1st quartile	229.4	2.995	231.2	9031	76.01	9078
Median	285.1	3.826	287.4	11,569	98.26	11,627
Mean	313.8	4.437	316.5	13,784	115.55	13,854
3rd quartile	393.4	5.284	396.5	18,568	140.87	18,651
Maximum	539.3	9.23	545	24,894	237.18	25,039
SD	108.113	2.022973	109.306	5683.733	53.238	5715.042
Variance	11,688.4	4.092418	1947.77	32,304,818	2834.29	32,661,700

Table 1. Descriptive statistics for time-series variables in middle east countries.

Visualization for exponential smoothing

The plots were designed for selected Middle East countries for forecasting purposes (2020–2030).

The projected values for the next ten years were visualized for ES. In Fig. 6a,e, depicting mortality resulting from PMP and HAP, an ascending trend is observed in Iraq, suggesting an increase in the number of mortalities each year for the forecasted period. Conversely, all other figures exhibit relatively constant forecasts, indicating a consistent number for the next 10 years. Figure 6c,d, illustrating mortalities and DALYs resulting from AOP, reveal an ascending trend in Kuwait and Iran, signaling a yearly increase in both Mortalities and DALYs resulting from AOP. In Fig. 6b,f showing DALYs resulting from HAP and PMP a rising trend is evident across all Middle East that shows a yearly rise in DALYs.

Visualization for SVM

The plots for eight Middle East countries, examining trends from 2020 to 2030, were generated using SVM projections. In Fig. 7a,b, which represent mortalities and DALYs resulting from PMP, an upward trend is observed in all countries. This indicates an increase in the number of mortalities and DALYs each year. However, in Syria and the Maldives, a downward trend suggests a decrease in both mortalities and DALYs over the same period.

Figure 7e,f, depicting mortalities and DALYs resulting from HAP, show rising trend in all middle east countries, signaling an annual rise in mortalities and DALYs resulting from HAP, except for Kuwait and Iraq where no change is observed. Figure 7c,d, illustrating Mortalities and DALYs resulting from AOP, reveal rising pattern in selected middle east countries except for Iran, which exhibits a different pattern.

Visualizations for NN

The plots designed for eight Middle East countries to analyze trend (2020–2030).

The predicted values for the years 2020–2030 were visualized for ANN models. In Fig. 8a,b illustrating mortalities and DALYs resulting from PMP an upward trend is evident in all countries indicating an annual increase in the number of mortalities and DALYs resulting from PMP, except in Syria and Maldives where a downward trend suggests a decrease in both Mortalities and DALYs. Figure 8e,f, portraying mortalities and DALYs resulting from HAP, exhibit an upward pattern in selected middle east countries, signaling an annual rise in mortalities and DALYs resulting from HAP, except for Kuwait and Iraq where no change is observed. Figure 8c,d, representing mortalities and DALYs resulting from AOP, reveal a rising trend in middle east countries, except in Iran, which displays a different pattern.

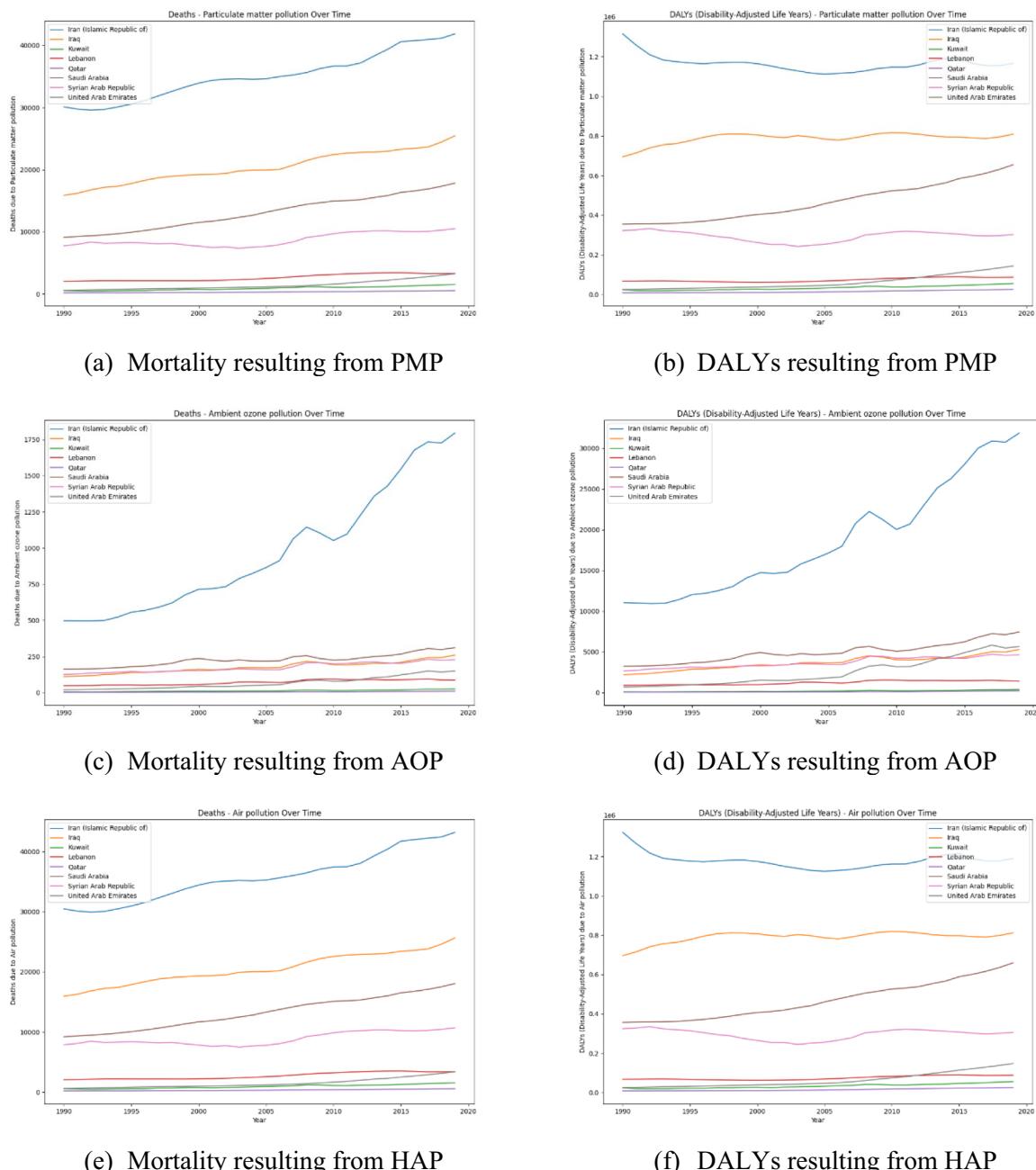


Fig. 4. Time series chart for mortalities and DALYs resulting from pollution.

Comparative study

The study used the ARIMA, Exponential Smoothing (ES), Support Vector Machines (SVM) and Artificial Neural Networks (ANN) methods for data analysis. The degree of accuracy of these models was measured using a statistical criterion known as Mean Absolute Percentage Error (MAPE). MAPE is a very easy to understand approach of determining the accuracy of a forecast because it portrays the percentage difference between the forecast and the actual outcome. This metric is useful to understand how effective each of the method of forecasting is in time series analysis.

Table 3 provides an evaluation of the forecasting performance of different models with measurement in Mean Absolute Percentage Error (MAPE) and with the models applied which includes ARIMA, Exponential Smoothing (ES), Support Vector Machines (SVM), and Artificial Neural Networks (ANN). SVM and ANN are machine and deep learning model respectively has greater accuracy than the traditional statistical models such as ARIMA and ES. Namely, with regard to the set pollution variables in the Middle Eastern countries, SVM was found to outperform all compared classifiers in a consistent manner. In the context of the United Arab Emirates (UAE), it is possible to state that SVM provides lower MAPE values as compared to ARIMA, ES, and ANN across the different flavors of pollution. For instance, in the task of Mortality due to Particulate Matter Pollution (PMP),

BIC score		DALYs			Mortalities		
		PMP	AOP	HAP	PMP	AOP	HAP
		Syrian Arab Republic					
BIC score	Arima	712.93	453.59	712.55	508.36	278.91	509.11
	ES	695.12	427.34	695.19	474.01	252.87	474.64
	Iraq						
	Arima	719.81	475.05	720.06	533.58	290.39	534.26
	ES	671.85	451.26	672.18	506.18	274.52	506.77
	Kuwait						
	Arima	623.12	329.16	624.11	421.38	181.06	422.09
	ES	586.66	311.17	586.97	388.87	161.72	389.59
	United Arab Emirates						
	Arima	683.93	511.15	686.31	471.28	298.59	473.8
	ES	652.17	481.02	654.18	445.03	280.71	447.19
	Islamic Republic of Iran						
	Arima	739.34	594.42	740.46	568.78	438.78	573.77
	ES	698.50	558.36	726.55	532.80	411.94	537.70
	Lebanon						
	Arima	629.02	401.05	629.61	453.78	247.79	457.89
	ES	587.52	371.12	587.96	423.46	226.71	426.12
	Saudi Arabia						
	Arima	738.10	493.80	738.64	541.49	306.26	542.07
	ES	692.75	465.53	693.30	512.70	284.04	513.27
	Qatar						
	Arima	575.87	290.01	574.27	346.73	111.57	347.47
	ES	536.11	271.31	538.41	318.87	102.28	319.68

Table 2. Comparison with BIC.

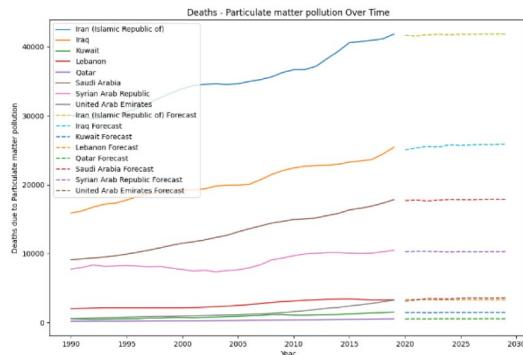
SVM has a phenomenal performance with the help of MAPE 0. 900, which is higher than other models such as ARIMA, which gave an RMSE of 24. 271, ES with 20. 607; and ANN with 22. 44. As with the other pollution variables, SVM gives lower MAPE values for AOP (5. 786) and HAP (0. 520).

While SVM currently exhibits superior performance in this context, it is important to recognize that ANN has the potential to excel, especially with more extensive datasets. The findings highlight SVM's reliability for pollution-related health predictions in the Middle East, but researchers should also consider ANN's scalability and potential for improved performance with larger datasets. This underscores the importance of selecting the most suitable model based on the specific dataset characteristics and forecasting requirements.

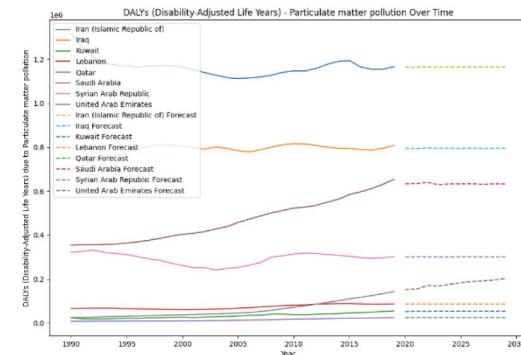
Discussion

The unique aspect of this research lies in its dedicated focus on the Middle East countries and their contributions to air pollution along with the consideration of Air Quality Action Plans (AOP). This research goes a step further by delving into the forecasted health consequences including potential Mortalities and disabilities resulting from air pollution in the Middle East. This research comprises time series analysis and machine learning methods. The usage of machine learning models shows an exit from common practices in air pollution forecasting. The complexity of the research is augmented by the need to gather and integrate data from various sources surrounding air quality, health and environmental factors across several Middle East countries. The study presents a novel solution to navigate this data puzzle. This research extends beyond ordinary prediction, aiming to provide actionable intuitions for policymakers. These insights become precious tools for making informed decisions. The research acknowledges the unique characteristics and challenges of each Middle East country tailoring its forecasts to align with the specific situations in these nations. The research accepts a global perspective suggesting that the methods and findings could offer valuable insights for other regions dealing with similar pollution issues. It contributes to a broader understanding of the health impacts of air pollution in the Middle East offering a fresh perspective on this critical issue and presenting a model that could inform strategies for fighting pollution worldwide.

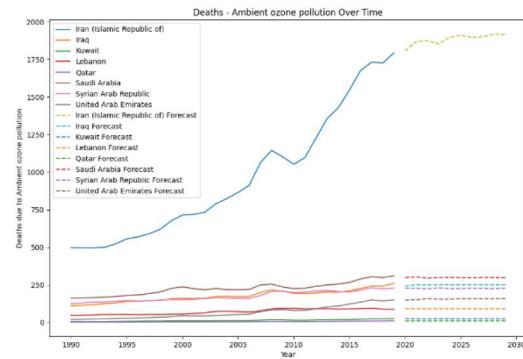
The implications of the results from this study are thereby significant in different fields. This research helps the public health departments to predict pollution and corresponding health effects and makes timely and efficient interventions possible because of the comparative study of the classical statistical models and the more sophisticated machine learning approaches. These understandings can hence be helpful for environmental agencies in improving the means by which air quality is being observed and also improving the management and mitigation of it. The work is useful to help health departments and government agencies minimize inefficiency in the use of resources if they are to apply the most accurate predictive models to areas with increased risks to the health



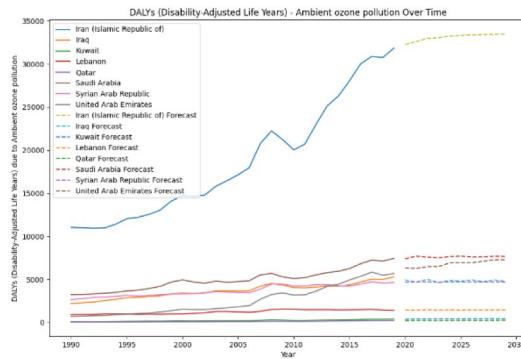
(a) Mortality resulting from PMP



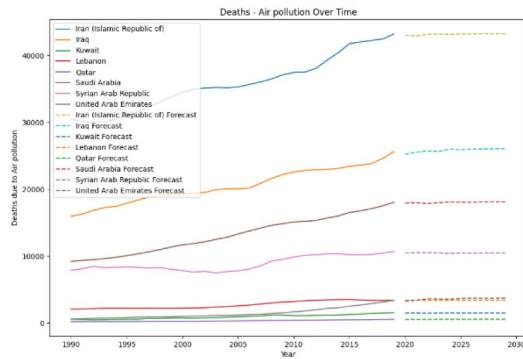
(b) DALYs resulting from PMP



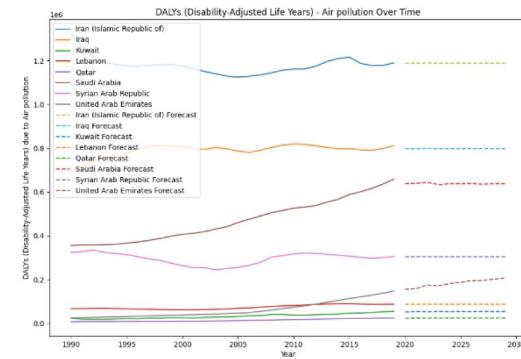
(c) Mortality resulting from AOP



(d) DALYs resulting from AOP



(e) Mortality resulting from HAP



(f) DALYs resulting from HAP

Fig. 5. Plots for forecasting using ARIMA.

of the people. The results may be used by the policymakers to provide regulations and guidelines that is based on facts to control air pollution and ultimately enhance public health. It also assists in creating awareness of the adverse effects of air pollution to the health of people hence increases the level of support by the public on environmental conservation. In conclusion, this study provides the guidelines for the further scientific work; opening the horizon for improvements and innovations in the field of environmental health prediction. The study offers a positive return on investment to the society by improving public health approaches, environmental surveillance, and decisions and policies.

Conclusion

This paper analyses the effects of air quality on health status in Middle Eastern countries employing Traditional and machine- learning forecasting models. Large differences were detected in pollution impacts as revealed by differences in mean values, SD values and range in each considered country. The analysis showed rising trends in mortalities and DALYs due to the three different types of pollution under study from 1990 to 2019. Among the models, SVM consistently outperformed others in accuracy, with a MAPE of 0.900 for predicting PMP mortalities in the UAE, compared to ARIMA's 24.271, ES's 20.607, and ANN's 22.44. While SVM demonstrated the

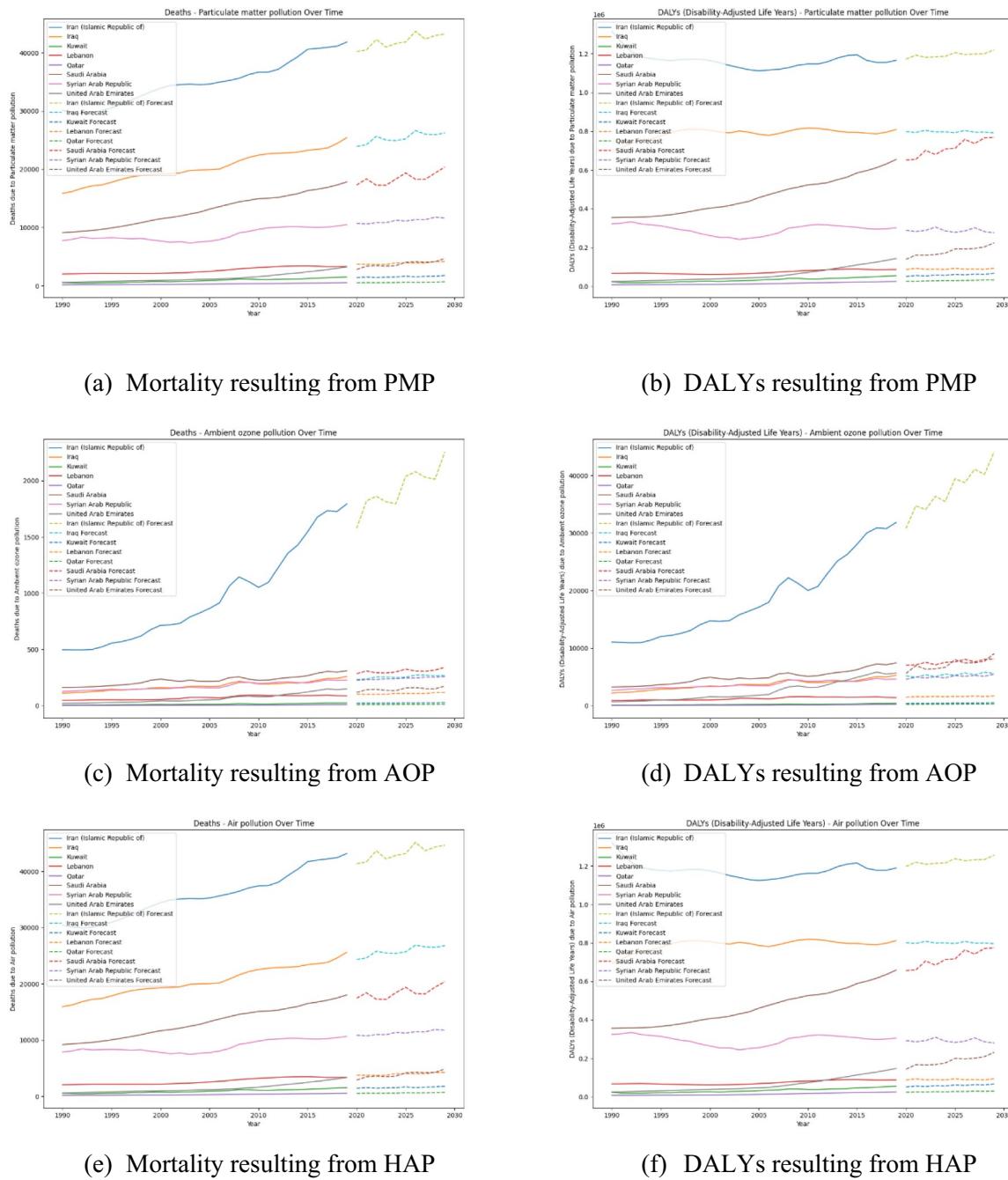


Fig. 6. Plots for forecasting all variables using ES.

best overall performance, ANN showed potential, particularly with larger datasets, as seen in predicting DALYs from PMP in Iraq, where it achieved a MAPE of 0.794 versus SVM's 0.887 and ARIMA's 2.191. The comparison of models using the MAPE metric reveals SVM's reliability and accuracy in predicting health outcomes related to air pollution. However, the study acknowledges the potential of ANN to excel particularly with more extensive datasets. The study emphasizes the uniqueness of its approach comparing traditional time series analysis with advanced neural network methodologies. It extends its implications globally. The study contributes to a broader understanding of this critical issue by offering a framework that could inform strategies worldwide. It presents a novel approach to forecasting and bridging the gap between traditional and advanced methodologies for more accurate and actionable predictions. These findings underscore the importance of selecting appropriate forecasting models based on dataset characteristics and offer valuable insights for policymakers aiming to mitigate the health impacts of air pollution in the Middle East and similar regions globally. The study's forecasts aid in emergency preparedness, while its comparative analysis of forecasting models helps refine predictive methodologies. Overall, it supports a comprehensive approach to mitigating air pollution and its health effects.

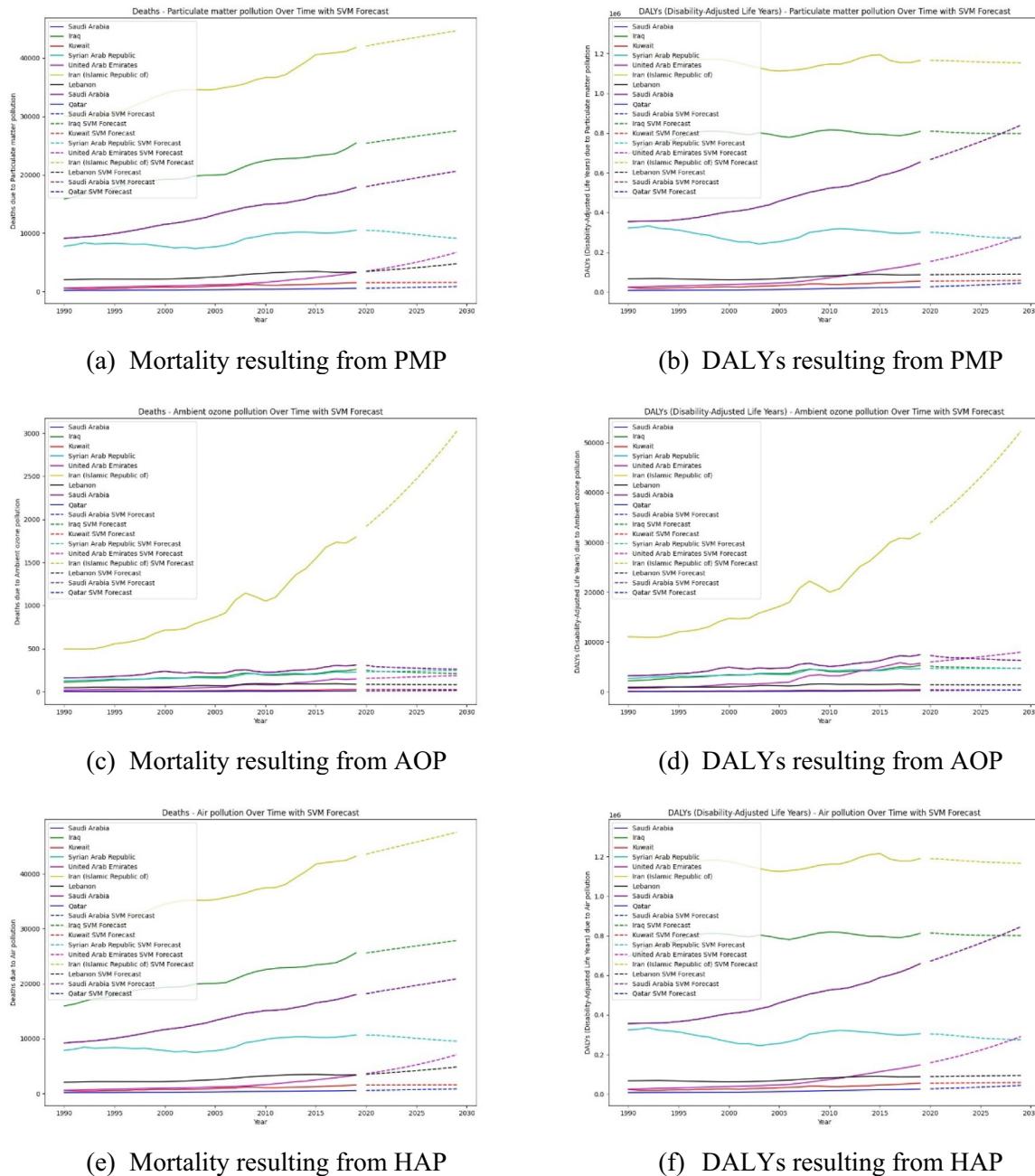


Fig. 7. Plots for forecasting all variables using SVM.

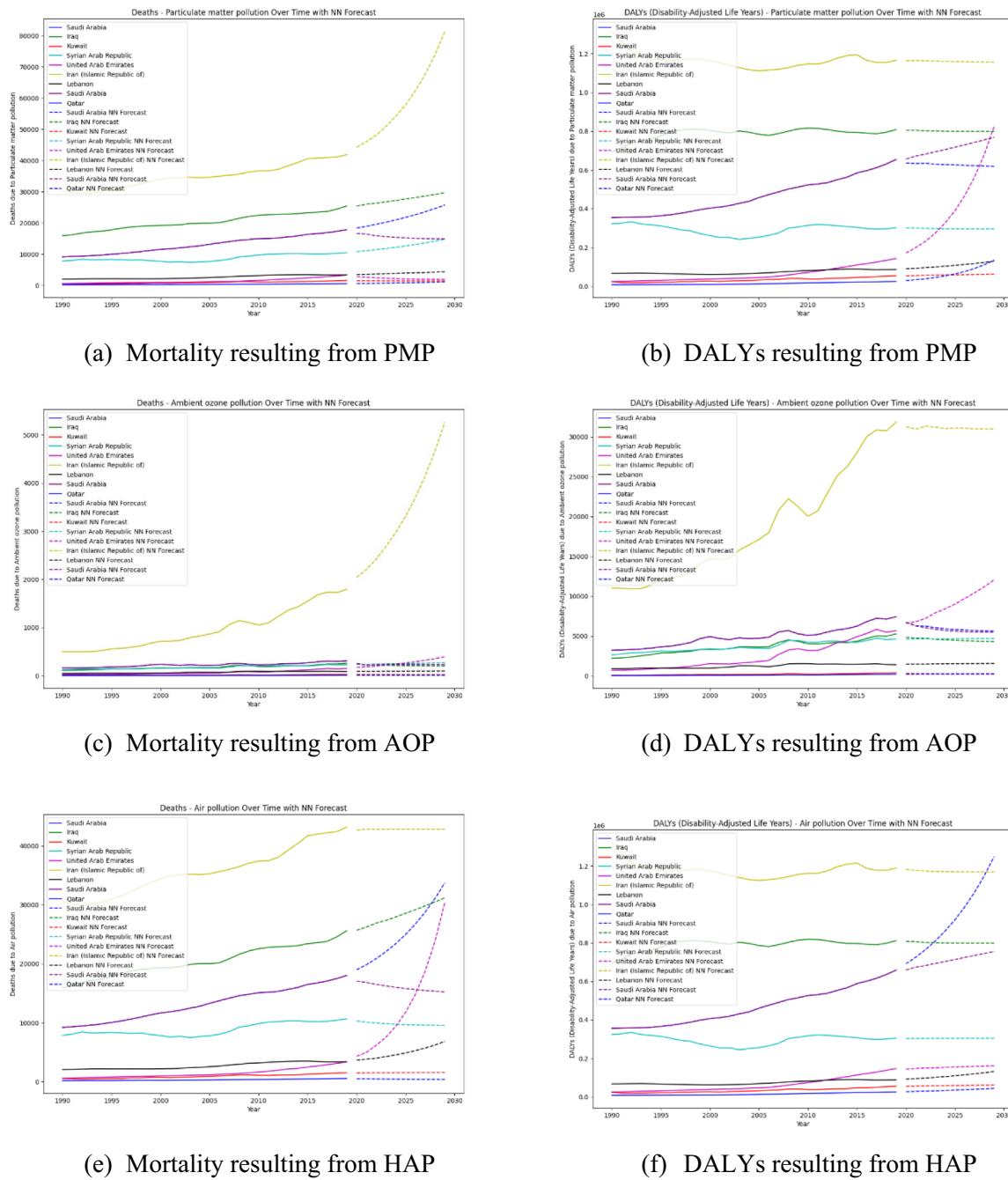


Fig. 8. Plots for forecasting all variables using Artificial NN.

Countries	Measure	Variables	MAPE			
			ARIMA	ES	SVM	NN
Syrian Arab Republic	Mortalities	Particulate matter pollution	5.495	5.264	2.580	0.998
		Ambient ozone pollution	6.274	7.099	3.337	3.010
		Air pollution	5.469	5.226	2.425	3.469
	DALYs	Particulate matter pollution	6.061	5.374	2.460	1.247
		Ambient ozone pollution	6.934	6.260	3.543	3.053
		Air pollution	6.000	5.208	2.364	1.880
Iraq	Mortalities	Particulate matter pollution	4.257	4.035	1.965	2.295
		Ambient ozone pollution	9.819	8.274	8.184	13.567
		Air pollution	4.269	4.059	1.941	3.569
	DALYs	Particulate matter pollution	2.191	2.076	0.794	0.887
		Ambient ozone pollution	10.727	8.239	6.849	10.143
		Air pollution	2.197	2.076	0.803	0.810
Kuwait	Mortalities	Particulate matter pollution	14.063	16.288	5.158	6.778
		Ambient ozone pollution	21.218	22.054	8.000	14.455
		Air pollution	14.107	16.332	5.212	7.107
	DALYs	Particulate matter pollution	12.038	12.109	4.965	6.074
		Ambient ozone pollution	16.170	12.624	10.177	19.439
		Air pollution	12.054	10.017	4.931	5.509
United Arab Emirates	Mortalities	Particulate matter pollution	24.271	20.607	0.900	22.44
		Ambient ozone pollution	29.998	33.737	5.786	6.741
		Air pollution	24.445	19.427	0.520	8.225
	DALYs	Particulate matter pollution	18.156	14.054	0.384	3.965
		Ambient ozone pollution	27.416	33.233	5.290	7.300
		Air pollution	18.355	14.231	0.677	1.903
Islamic Republic of Iran	Mortalities	Particulate matter pollution	4.107	4.098	0.570	4.783
		Ambient ozone pollution	17.883	14.976	3.081	7.274
		Air pollution	4.394	4.339	0.637	1.147
	DALYs	Particulate matter pollution	2.095	2.184	0.600	0.721
		Ambient ozone pollution	12.269	10.570	2.934	1.676
		Air pollution	2.147	2.144	0.645	0.614
Lebanon	Mortalities	Particulate matter pollution	6.652	8.331	4.457	4.495
		Ambient ozone pollution	9.705	10.399	2.873	4.583
		Air pollution	5.954	8.188	4.567	9.580
	DALYs	Particulate matter pollution	4.292	4.474	1.564	7.570
		Ambient ozone pollution	6.883	7.074	3.027	4.887
		Air pollution	4.256	4.462	1.372	3.493
Saudi Arabia	Mortalities	Particulate matter pollution	7.462	7.519	0.923	1.927
		Ambient ozone pollution	7.987	8.246	6.188	4.849
		Air pollution	7.456	7.521	0.951	1.535
	DALYs	Particulate matter pollution	5.393	4.868	0.714	1.936
		Ambient ozone pollution	9.128	9.218	6.388	13.527
		Air pollution	5.412	4.886	0.736	2.533
Qatar	Mortalities	Particulate matter pollution	11.023	9.845	2.001	3.165
		Ambient ozone pollution	20.558	19.040	4.103	14.354
		Air pollution	11.096	9.892	1.959	4.615
	DALYs	Particulate matter pollution	11.590	9.889	3.886	6.376
		Ambient ozone pollution	14.463	12.144	3.016	3.258
		Air pollution	10.678	10.189	3.781	3.893

Table 3. Comparison of models from MAPE.

Data availability

All data analyzed in the course of this research were sourced from the “Institute of Health Metrics and Evaluation GBD” website, covering the years 1990 to 2019. The data can be accessed through the following link: <https://ghdx.healthdata.org/gbd-2019>.

Received: 15 March 2024; Accepted: 30 August 2024
 Published online: 14 September 2024

References

- European Environmental Agency. (2022, April 15). <http://www.eea.europa.eu/>.
- GBD Results. (n.d.). Institute for Health Metrics and Evaluation. <https://vizhub.healthdata.org/gbd-results/>.
- Texas Disposal Systems. (2022, April 15). <https://www.texasdisposal.com/blog/land-pollution/>.
- Harvard T.H. Chan, School of Public Health. (2022, March 22). <https://www.hsp.harvard.edu/ehep/82-2/#:~:text=Water%20pollution%20is%20the%20contamination,make%20their%20way%20to%20water>.
- National Institute of Environmental Health Science (NIH). (2021). <https://www.niehs.nih.gov/>.
- National Institute of Environmental Health Science (NIH). (2022). <https://www.niehs.nih.gov/>.
- World Health Organization. (2022, April 21). <https://www.who.int/data/gho/data/themes/air-pollution/household-air-pollution2022>.
- Pan American Health Organization. (2022). <https://www.paho.org/en/topics/air-quality-and-health/ambient-and-household-air-pollution-and-health-frequently-asked>.
- Perera, F., Ashrafi, A., Kinney, P. & Mills, D. Towards a fuller assessment of benefits to children's health of reducing air pollution and mitigating climate change due to fossil fuel combustion. *Environmental Research* **172**, 55–72. <https://doi.org/10.1016/j.envres.2018.12.016> (2019).
- Ejohwomu, O. A. *et al.* Modelling and forecasting temporal PM2.5 concentration using ensemble machine learning methods. *Buildings* **12**(1), 46. <https://doi.org/10.3390/buildings12010046> (2022).
- Lara-Benitez, P., Carranza-Garcia, M. & Riquelme, J. C. An experimental review on deep learning architectures. *International Journal of Neural Systems* **31**(3), 1–25. <https://doi.org/10.1142/S0129065721300011> (2021).
- Bhatti, U. A. *et al.* Time series analysis and forecasting of air pollution particulate matter (PM2.5): An SARIMA and factor analysis approach. *IEEE Access* **9**, 41019–41031. <https://doi.org/10.1109/ACCESS.2021.3060744> (2021).
- Guan, Y. *et al.* Molecular epidemiology of the novel coronavirus that causes severe acute respiratory syndrome. *Lancet (London, England)* **363**(9403), 99–104. [https://doi.org/10.1016/s0140-6736\(03\)15259-2](https://doi.org/10.1016/s0140-6736(03)15259-2) (2004).
- Cao, Q., Rui, G. & Liang, Y. Study on PM2.5 pollution and the mortality due to lung cancer in China based on the geographic weighted regression model. *BMC Public Health* **18**(925), 1–10. <https://doi.org/10.1186/s12889-018-5844-4> (2018).
- Niazkar, H. R. & Niazkar, M. Application of artificial neural networks to predict the COVID-19 outbreak. *Global Health Research and Policy* <https://doi.org/10.1186/s41256-020-00175-y> (2020).
- Jiang, W., Wu, X., Gong, Y., Yu, W. & Zhong, X. Holt-Winters smoothing enhanced by fruit fly optimization algorithm to forecast monthly electricity consumption. *Energy* **193**, 116779. <https://doi.org/10.1016/j.energy.2019.116779> (2020).
- Zhang, X. *et al.* Comparative study of four time series methods in forecasting typhoid fever incidence in China. *PLoS ONE* **8**(5), e63116. <https://doi.org/10.1371/journal.pone.0063116> (2013).
- Zhang, Y. & Yew, W. W. Mechanisms of drug resistance in *Mycobacterium tuberculosis*: Update 2015. *The International Journal of Tuberculosis and Lung Disease: The Official Journal of the International Union against Tuberculosis and Lung Disease* **19**(11), 1276–1289. <https://doi.org/10.5588/ijtdl.15.0389> (2015).
- Al-Najjar, H. & Al-Rousan, N. A classifier prediction model to predict the status of Coronavirus COVID-19 patients in South Korea. *European Review* (2020). <https://www.europeanreview.org/article/20709>
- Guo, Q. & He, Z. Prediction of the confirmed cases and deaths of global COVID-19 using artificial intelligence. *Environmental Science and Pollution Research* **28**, 11672–11682 (2021).
- Kozuch, A., Cywicka, D. & Adamowicz, K. A comparison of artificial neural network and time series models for timber price forecasting. *Forests* **14**(2), 177–177. <https://doi.org/10.3390/f14020177> (2023).
- Salih, S. O. *et al.* Forecasting causes of death in northern Iraq using neural network. *Journal of Statistical Theory and Applications* **21**(2), 58–77. <https://doi.org/10.1007/s44199-022-00042-4> (2022).
- Al Yammahi, A. & Aung, Z. Forecasting the concentration of NO₂ using statistical and machine learning methods: A case study in the UAE. *Heliyon* <https://doi.org/10.1016/j.heliyon.2022.e12584> (2022).

Acknowledgements

The authors extend their appreciation to Taif University, Saudi Arabia, for supporting this work through Project Number (TU-DSPP-2024-94).

Author contributions

A.B. and M.N.A. conceived and designed the study. A.B. and N.M. conducted the data collection and preprocessing. M.N.A. and U.Y. performed the statistical analysis and modeling. O.A. worked on the evaluation of Time Series/ARIMA models and contributed to the analysis. O.A. also improved the language and presentation of the manuscript. N.M. and M.N prepared the figures and tables. E.E.M. reviewed the manuscript and made technical corrections. E.E.M. also contributed to the analysis of figures. M.N. is the corresponding author and managed the manuscript submission process. All authors reviewed the manuscript and approved the final version.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024