Zhang et al. BMC Public Health (2024) 24:3269
https://doi.org/10.1186/s12889-024-20713-4

BMC Public Health

RESEARCH

Open Access

The use of machine and deep learning to model the relationship between discomfort temperature and labor productivity loss among petrochemical workers

Yilin Zhang[1†], Yifeng Chen[1†], Qingling Su[2], Xiaoyin Huang[2], Qingyu Li[1], Yan Yang[1], Zitong Zhang[1], Jiake Chen[1], Zhihong Xiao[1], Rong Xu[3], Qing Zu[3], Shanshan Du[2], Wei Zheng[3*], Weimin Ye[2*] and Jianjun Xiang[1,4*]

Abstract

Background Workplace may not only increase the risk of heat-related illnesses and injuries but also compromise work efficiency, particularly in a warming climate. This study aimed to utilize machine learning (ML) and deep learning (DL) algorithms to quantify the impact of temperature discomfort on productivity loss among petrochemical workers and to identify key influencing factors.

Methods A cross-sectional face-to-face questionnaire survey was conducted among petrochemical workers between May and September 2023 in Fujian Province, China. Initial feature selection was performed using Lasso regression. The dataset was divided into training (70%), validation (20%), and testing (10%) sets. Six predictive models were evaluated: support vector machine (SVM), random forest (RF), extreme gradient boosting (XGBoost), Gaussian Naive Bayes (GNB), multilayer perceptron (MLP), and logistic regression (LR). The most effective model was further analyzed with SHapley Additive exPlanations (SHAP).

Results Among the 2393 workers surveyed, 58.4% (1,747) reported productivity loss when working in high temperatures. Lasso regression identified twenty-seven predictive factors such as educational level and smoking. All six models displayed strong prediction accuracy (SVM = 0.775, RF = 0.760, XGBoost = 0.727, GNB = 0.863, MLP = 0.738, LR = 0.680). GNB model showed the best performance, with a cutoff of 0.869, accuracy of 0.863, precision of 0.897, sensitivity of 0.918, specificity of 0.715, and an F1-score of 0.642, indicating its efficacy as a predictive tool. SHAP analysis showed that occupational health training (SHAP value: -3.56), protective measures (-2.61), and less physically

†Yilin Zhang and Yifeng Chen contributed equally to this work.

*Correspondence:
Wei Zheng
zheng77wei@hotmail.com
Weimin Ye
ywm@fjmu.edu.cn
Jianjun Xiang
jianjun.xiang@fjmu.edu.cn
Full list of author information is available at the end of the article

demanding jobs (-1.75) were negatively associated with heat-attributed productivity loss, whereas lack of air conditioning (1.92), noise (2.64), vibration (1.15), and dust (0.95) increased the risk of heat-induced productivity loss.

Conclusions Temperature discomfort significantly undermined labor productivity in the petrochemical sector, and this impact may worsen in a warming climate if adaptation and prevention measures are insufficient. To effectively reduce heat-related productivity loss, there is a need to strengthen occupational health training and implement strict controls for occupational hazards, minimizing the potential combined effects of heat with other exposures.

Keywords Temperature discomfort, High temperature, Labor productivity loss, Machine learning, SHAP

Background

The petrochemical industry plays a crucial role in the global economy [1], while being associated with substantial health risks [2]. Petrochemical workers are exposed to a variety of occupational hazards during the production and processing of petrochemical products, such as heat stress, noise, petrol, and hydrogen sulfide [3, 4]. Among these hazards, temperature discomfort has become a pressing occupational health and safety issue, especially in the context of climate change [5]. Studies have indicated that workplace heat exposure may have deleterious effects on workers' health and increase the risk of occupational injuries/accidents [3, 4, 6]. Long-term heat exposure may lead to some chronic health outcomes, such as cardiovascular dysfunction [7], hypertension [8], and abnormal immune parameters [9]. In addition to occupational health and safety [10], empirical evidence from Australia suggested that high temperatures may burden the global economy [11], affecting all levels of our society. Dasgupta and colleagues found that a 3.0 °C increase in global temperatures could reduce the total workforce by 18% [12]. However, most studies investigating the risks associated with temperature discomfort have been conducted in high-income countries/region [10], with limited studies from heat-vulnerable middle-low-income countries. Moreover, scant attention has been paid to the impact of temperature discomfort on labor productivity loss among petrochemical workers, a population susceptible to heat-related illnesses and injuries [13].

In reflecting the real occupational settings where workers are typically exposed to a combination of multiple hazards, conventional analytical approaches, such as regression analysis and time-series forecasting, often fall short of comprehensively and precisely capturing the multifaceted effects of high-temperature environments on occupational outcomes [14]. Their limitations stem primarily from their intrinsic constraints: a limited ability to model non-linear relationships, challenges of capturing interactions between multiple variables, and difficulties in adequately handling high-dimensional data [14]. Specifically, when addressing complex and interactive effects of multiple exposures, conventional methodologies may overlook subtle patterns and interdependencies that machine learning and deep learning techniques are better suited to reveal [15]. Moreover, machine learning and deep learning are robust computational tools capable of analyzing large datasets and exploring intricate patterns and relationships [16]. Their effectiveness spans diverse domains, from disease prediction and patient outcomes in healthcare [17], to ecosystem conservation and pollution assessment in environmental science [18, 19], and understanding behavior and cognitive functions in psychology [20].

This investigation employed multiple machine learning and deep learning models — SVM, RF, XGBoost, GNB, MLP, and LR — to address specific challenges inherent to large-scale data analysis. SVM excels in processing high-dimensional data but can be computationally intensive on massive datasets [21]. RF is widely used at handling numerous features and missing values, yet it is prone to overfitting in noisy data [22]. XGBoost is valued for its speed and performance yet lacks interpretability and is sensitive to hyperparameter settings [23]. GNB offers simplicity and scalability, however, it assumes independence between features, which is often unrealistic [23]. MLP is effective at identifying complex patterns but may overfit due to its complexity [24]. Conversely, LR is straightforward but limited to linear relationships [25]. Considering these characteristics, this study adopted a multimodal approach, leveraging the strengths of each model to offset others' weaknesses. This method aims to integrate multiple models' outcomes, resulting in more accurate and robust predictions, thereby enhancing the overall efficacy and reliability of the analytical process.

Currently, few studies have investigated the impact of temperature discomfort on the labor productivity loss of petrochemical workers, particularly through machine learning and deep learning methods. Based on forecasting models, this study aimed to assess the impact of temperature discomfort on the productivity of petrochemical workers and identify the influencing factors. Findings of this study may provide valuable evidence for decision-making and heat safety management in the petrochemical industry.

Methods

Study design and participant recruitment

A cross-sectional questionnaire survey was conducted among petrochemical workers between May and September 2023 at the Quangang Petrochemical Industrial Park (QPIP), Quanzhou, Fujian Province, China. Located on the west bank of the Taiwan Straits and the south bank of Meizhou Bay, QPIP was launched in 2005. It covers an area of about 30 square kilometers and has a population of around 360,000. Currently, 45 petrochemical related enterprises have settled in the QPIP, which has been recognized as one of the top 20 chemical industrial parks in China and a national circular transformation model park. The industrial chain of QPIP mainly includes ethylene, propylene, carbon tetrachloride, benzene and paraxylene, with a total output value of about $15 billion [26].

According to the Law of the People's Republic of China on the Prevention and Control of Occupational Diseases, employers are required to organize regular occupational health examinations for workers exposed to occupational hazards [27]. During the study period, workers from the two leading petrochemical plants of QPIP were recruited to participate in the face-to-face questionnaire survey while undergoing their occupational health examinations in the Medical Examination Center of Minnan Branch of the First Affiliated Hospital of Fujian Medical University. Inclusion criteria were: (a) Currently registered petrochemical workers at QPIP; (b) Workers undergoing occupational health examinations during the study period; (c) Individuals who consent to participate and can provide informed consent. Exclusion criteria were: (a) Workers younger than 18 or older than 65 years; (b) Individuals with cognitive impairments or mental health conditions that preclude understanding of the questionnaire or independent completion; (c) Workers who have not worked in QPIP in the three months prior to the study commencement.

Trained interviewers administered the electronic questionnaires via iPads. All conversations between participants and interviewers were recorded, and the entered data were double-checked against the recordings on the same day to prevent data entry errors or misclassification. Participation in this study was completely voluntary, with no incentives provided. The study was approved by the Fujian Medical University Ethics Committee (Approval No. 2022−111).

Questionnaire design

Following an extensive literature review and consultations with relevant experts, including senior occupational hygienists in petrochemical plants and university professors in occupational health, we developed a survey instrument. The questionnaire was initially piloted with a representative sample of 50 petrochemical workers to ensure clarity and relevance. The questionnaire comprised three parts. The first part mainly requested the following demographic information and individual lifestyle habits: gender, age, body mass index (BMI), marital status, education level, annual income, medication use, sleep quality, smoking, and alcohol consumption. The second part included questions on employment details such as task group, job category, working hours per week, use of personal protective equipment (PPE) such as reflective vests, safety boots, helmet, gloves and overall. Occupational hazards exposure information was collected through yes/no questions. The third part included six questions to measure the impact of high temperatures on work efficiency, using a 5-point Likert scale (strongly disagree, disagree, neutral, agree, strongly agree), with scores assigned from 1 to 5, respectively. The total score could range from 6 to 30. In this study, workers with scores between 6 and 18 were categorized as 'No labor productivity loss', while those with scores between 19 and 30 were categorized as experiencing 'Labor productivity loss'. For specific questions, please refer to the supplementary file. The Cronbach's alpha coefficient for the scale was 0.81, indicating good reliability [28].

Statistical analysis

Descriptive analyses were conducted to characterize normally distributed continuous variables, presented as mean ± standard deviation (Mean ± SD), while non-normally distributed data were reported using median ± inter-quartile range (Med ± IQR). To compare differences in quantitative variables between groups, an independent samples t-test was employed. Categorical variables were summarized using frequencies and percentages. The Chi-square (χ^2) test or Fisher's exact test was used for statistical analysis of categorical variables.

Feature selection for predictive modeling utilized Lasso regression. This study applied five-fold cross-validation to divide the data into three distinct sets: training (70%), validation (20%), and testing (10%) [29]. During the training phase, six predictive algorithms were explored, including SVM, RF, XGBoost, GNB, MLP, and LR. SVM: C (regularization factor) was set to 1.0, kernel function was the sigmoid type, and tol (tolerance for convergence) was 0.001. RF: criterion = "gini", max depth = 500, min impurity decrease = 0.0, and *n* estimators = 436. XGBoost: objective function optimized for binary: logistic, learning rate set to 0.1, maximum tree depth capped at 6, minimum sum of instance weight (min child weight) fixed at 1, and L2 regularization term (reg lambda) at 0.5. GNB: a model was constructed utilizing the naiveBayes function from the e1071 package in R, with the priors (a priori probability) set to default as None and var smoothing established at 1e-07. MLP: activation function (nonlinear function) = logistic, hidden layer sizes = (10, 10), and max iterations = 10. LR: maximum iterations (max iter) = 100, regularization type (penalty) ='1', convergence metric (tol) = 0.0001, and regularization factor (C) = 0.123.

During the validation phase, hyperparameter optimization was performed to fine-tune each model. The test set facilitated the assessment of model efficacy using various metrics, such as area under the curve (AUC), accuracy, precision, sensitivity, specificity, F1-score, and cutoff value. These metrics guided the selection of the optimal model. Interpretability of the selected model was enhanced through SHapley Additive exPlanations (SHAP) value analysis. Statistical analyses were conducted utilizing SPSS software (version 25.0) and the R (version 4.1.3), with results presented with a 95% confidence interval (CI).

Extraction of key features

Prior to applying Lasso regression for variable selection, it is crucial to consider its key assumptions. Lasso regression, a method known for its ability to perform both variable selection and regularization to prevent overfitting, assumes that the relationships between the predictor variables and the outcome are linear and additive [30]. Another critical assumption is the independence of errors, meaning that the residuals from the models are uncorrelated with each other [31]. Given the potential for multicollinearity among predictor variables, a common challenge in datasets with numerous features, Lasso regression is particularly adept at handling such conditions due to its regularization component [32].

In this study, we employed Lasso regression to streamline our predictive model. By introducing a penalty term (L1 norm) into the cost function, Lasso regression constrains the absolute values of the coefficients, effectively shrinking some coefficients to zero [32]. This regularization process not only reduces the complexity of the model by performing feature selection but also mitigates the risk of multicollinearity, thereby enhancing model interpretability and reliability [33]. Specifically, we utilized the Lasso CV algorithm, which integrates five-fold cross-validation. This approach systematically evaluates the model's performance with different data subsets and excludes variables with coefficients reduced to zero [34]. Such a methodical exclusion based on cross-validation enhances the model's robustness and generalizability, aligning with our aim to efficiently identify key factors influencing productivity loss among petrochemical workers due to temperature discomfort.

Construction of machine and deep learning models

To address the multifaceted nature of temperature discomfort and the associated productivity loss, this study employed a variety of ML models. Given the unique strengths and biases inherent in different algorithms based on their architecture and underlying assumptions, utilizing a diverse array of analytical tools facilitates the selection of the optimal model to ensure robust predictions. In this research, the following six models were adopted to explore the predictive efficiency of temperature discomfort and productivity loss: (1) Support vector machine (SVM): This model exhibits strong generalization capabilities in handling high-dimensional data and is particularly adept at solving nonlinear problems [21]; (2) Random forest (RF): As part of ensemble learning techniques, RF can process a large number of features, providing highly accurate classifications and possessing good noise resistance [22]; (3) Extreme gradient boosting (XGBoost): An enhanced version of decision trees, XGBoost optimizes performance through constructing multiple models and is renowned for handling large datasets and complex pattern recognition [23]; (4) Gaussian Naive Bayes (GNB): Applicable to environments with smaller data volumes or known distributions, this method is easy to implement and execute swiftly [35]; (5) Multilayer perceptron (MLP): A type of deep neural network suitable for problems with complex feature correlations and multiple layers, it can learn deeper representations of the data [24]; and (6) Logistic regression (LR): Although a more traditional technique, it demonstrates stable and interpretable performance in binary classification problems [25].

Assessment of variable significance

ML and DL models are often considered to be opaque due to the complexity involved in understanding how these algorithms generate precise predictions for specific groups. To address this challenge, our study employed SHAP (SHapley Additive exPlanations) values, a comprehensive framework initially derived from the Shapley values in cooperative game theory [36].

SHAP values provide a powerful framework derived from Shapley values in cooperative game theory [37]. These values offer a robust method by articulating how each feature in the model contributes to the prediction outcome both individually and in conjunction with other features [37]. This capacity to distribute the 'payout' (prediction output) among the features proportionately according to their contribution enables us to dissect the decision-making process of complex ML and DL models. Consequently, this detailed dissection helps identify primary determinants adversely affecting work efficiency under high-temperature conditions in the petrochemical industry.

Practically, SHAP values assess the marginal contribution of each feature towards the predictive accuracy at any given instance [38]. For our optimal model, this translates into identifying which variables most significantly impact productivity loss due to temperature discomfort. By examining interactions between features in the context of their contribution to the output, SHAP also guides understanding synergistic effects that may not be readily apparent in traditional statistical or even some machine learning evaluations.

Results

Comparison of demographic characteristics between affected and unaffected workers

In this study, 3,000 workers were recruited to participate in the survey, with 2,586 accepting, resulting in a participation rate of 86.20%. Of the questionnaires completed, 2,393 were valid, yielding an effective response rate of 92.54%. As shown in Table 1, most of the respondents (1,747, 58.37%) reported productivity loss when working in high-temperature conditions. Moreover, male workers were more likely to report productivity loss due to heat exposure compared to their female counterparts.

The vast majority of the participants were of Han nationality (99.12%), with Han workers showing a higher proportion of self-reported productivity loss due to heat exposure than their counterparts from ethnic minorities (9 Manchu, 11 Hui, and 1 She) ($P=0.008$). The average age of the participants was 38.07 ± 11.15 years, with no significant difference in age between those affected and unaffected by heat-related productivity loss ($P=0.192$). However, variations in BMI were observed. The overall average BMI was 24.37 ± 3.82 kg/m^2. Among workers with a normal BMI, those reporting productivity loss had a higher BMI than those without productivity loss (21.81 ± 1.41 vs. 21.51 ± 1.51, $P=0.001$). Similarly, in the obese category, affected workers also had a higher BMI than unaffected ones (30.95 ± 4.87 vs. 30.12 ± 1.91, $P=0.042$). However, no significant differences in BMI were found between the normal and overweight groups ($P>0.005$).

Education levels showed a marked difference between the two groups. Unaffected workers had higher education levels, with more individuals holding secondary education or below (21.67% vs. 26.94%) and postgraduate degrees (1.70% vs. 2.00%). This difference was statistically significant, indicating that educational attainment might influence susceptibility to productivity loss due to heat exposure. A higher percentage of unaffected workers were married (73.22%), compared to affected workers (67.37%) ($P=0.028$). Additionally, affected workers had a higher prevalence of chronic diseases compared to unaffected workers (23.93% vs. 16.10%, $P<0.001$), with hypertension being the most common (87.21%), followed by diabetes (12.79%). Workers who drank at least once per month had a higher percentage of productivity loss compared to those drank less frequently. No significant difference in smoking habits was observed between the two groups ($P=0.872$). However, medication use was significantly higher among affected workers (10.36%) compared to unaffected workers (7.59%).

Table 2 highlights significant correlations between occupational factors and heat-induced labor productivity loss. In terms of nature of work, workers in aromatic and olefin production (81.33%) had the highest percentage of labor productivity loss due to heat exposure, followed by petroleum refining (76.64%), storage and transportation (74.34%), and service operation (68.32%). Workers affected by heat-induced efficiency loss had a longer average employment duration (15.28 ± 12.15 years) compared to their unaffected counterparts (13.91 ± 11.27 years, $P=0.041$). A similar pattern was observed in weekly working hours, with affected workers working longer hours (48.92 ± 6.55 h) than those unaffected (44.45 ± 7.62 h, $P=0.048$). Either indoor frontline workers (74.42%) or outdoor frontline workers (73.25%) had a significantly higher percentage of heat-induced labor productivity loss than managerial or supervisory staff (69.52%). Workplace greenery did not significantly impact productivity loss due to heat ($P=0.25$). The presence of heat control measures like electric fans and air-conditioning units were significantly associated with productivity loss ($P=0.017$). Exposure to occupational hazards like vibration, noise, dust, fumes, and oil mist were significant determinants of work efficiency, with exposed workers consistently showing higher labor productivity loss across these factors ($P<0.001$). The level of physical strain was positively associated with labor productivity loss. Workers not required to operate within confined spaces ($n=1923$, 80.36%, $P<0.001$) and those who received occupational health training ($n=2213$, 92.48%, $P<0.001$) reported less decline in work efficiency. Workers using PPE (75.07%) experienced a higher percentage of productivity loss than those not using (44.79%).

Selection of critical variables

To minimize multicollinearity among the various predictors, Lasso regression was employed for dimensionality reduction across all surveyed variables, facilitating the extraction of key features. From the initial set of variables, Lasso CV identified 27 significant factors including educational level, marital status, health status, tobacco use, the nature of physical activity at work, frequency of sleep difficulties, sleep quality, and fatigue levels. The optimal lambda value determined through cross-validation was 0.004. Notably, variables such as education level, use of PPE, access to air conditioning, and occupational health training were given substantial weights, as illustrated in Fig. 1.

Comparison of ML and DL models

Table 3 shows the AUC (95%CI), accuracy, sensitivity, specificity, F1-score, cutoff, and other indices for all

Table 1 Demographic characteristics *n* (%) of surveyed petrochemical workers

Variable	Category	Total (*n* = 2393)	Labor productivity loss		χ^2/t	*P*
			No (*n* = 646)	Yes (*n* = 1747)		
Gender	Female	460 (19.22)	181 (28.02)	279 (15.97)	44.089	< 0.001
	Male	1933 (80.78)	465 (71.98)	1468 (84.03)		
Nationality	Han	2372 (99.12)	635 (98.30)	1737 (99.43)	6.928	0.008
	Other	21 (0.88)	11 (1.70)	10 (0.57)		
Age	(Years, Mean ± SD)	38.07 ± 11.15	39.09 ± 10.97	38.57 ± 11.30	1.303	0.192
BMI (kg/m^2, Mean ± SD)	Underweight (BMI < 18.5)	17.45 ± 0.92	17.42 ± 0.88	17.46 ± 0.95	-0.206	0.837
	Normal (18.5 ≤ BMI < 24.0)	21.73 ± 1.45	21.51 ± 1.51	21.81 ± 1.41	-3.217	0.001
	Overweight (24.0 ≤ BMI < 28.0)	25.67 ± 1.11	25.73 ± 1.10	25.65 ± 1.12	0.903	0.367
	Obese (BMI ≥ 28.0)	30.75 ± 4.34	30.12 ± 1.91	30.95 ± 4.87	-2.046	0.042
Educational attainment	Secondary or below	314 (13.12)	140 (21.67)	174 (26.94)	72.105	< 0.001
	High school	672 (28.08)	187 (28.95)	485 (27.76)		
	Technical/vocational college	740 (30.92)	160 (24.77)	580 (33.20)		
	Undergraduate	621 (25.95)	148 (22.91)	473 (27.08)		
	Postgraduate or above	46 (1.92)	11 (1.70)	35 (2.00)		
Marital status	Single	687 (28.71)	156 (24.15)	531 (30.40)		0.028*
	Married	1650 (68.95)	473 (73.22)	1177 (67.37)		
	Divorced	50 (2.09)	15 (2.32)	35 (2.00)		
	Widowed	6 (0.25)	2 (0.31)	4 (0.23)		
Health status	Without chronic diseases	1871 (78.19)	542 (83.90)	1329 (76.07)	16.943	< 0.001
	With chronic diseases	522 (21.81)	104 (16.10)	418 (23.93)		
Medication use	No	2163 (90.39)	597 (92.42)	1566 (89.64)	4.182	0.041
	Yes	230 (9.61)	49 (7.59)	181 (10.36)		
Smoking status	Never	1404(58.67)	387(59.91)	1017(58.21)	0.641	0.887
	Occasionally	143(5.98)	38(5.88)	105(6.01)		
	Daily	680(28.42)	179(27.71)	501(28.68)		
	Quit	166(6.94)	42(6.50)	124(7.10)		
Alcohol consumption	Never	115 (4.81)	29 (4.49)	86 (4.92)	14.251	0.027
	1–2 times per year	453 (18.93)	121 (18.73)	332 (19.00)		
	Less than once per month	1180 (49.31)	346 (53.56)	834 (47.74)		
	Monthly	313 (13.08)	76 (11.77)	237 (13.57)		
	1–2 times per week	222 (9.28)	43 (6.66)	179 (10.25)		
	3–5 times per week	77 (3.22)	18 (2.79)	59 (3.38)		
	Daily	33 (1.38)	13 (2.01)	20 (1.15)		
Annual income (CNY)	< 10,000	35 (1.46)	7 (1.08)	28 (1.60)	24.293	< 0.001
	10,000–29,900	42 (1.76)	11 (1.70)	31 (1.77)		
	30,000–59,900	118 (4.93)	44 (6.81)	74 (4.24)		
	60,000–99,900	335 (14.00)	117 (18.11)	218 (12.48)		
	100,000-149,900	707 (29.55)	193 (29.88)	514 (29.42)		
	150,000-299,900	861 (35.98)	208 (32.20)	653 (37.38)		
	≥ 3,000,000	295 (12.33)	66 (10.22)	229 (13.11)		
Life satisfaction	Very satisfied	475 (19.85)	152 (23.53)	323 (18.49)		0.016*
	Mostly satisfied	1876 (78.40)	487 (75.39)	1389 (79.51)		
	Unsatisfied	36 (1.50)	5 (0.77)	31 (1.77)		
	Very unsatisfied	6 (0.25)	2 (0.31)	4 (0.23)		
Insufficient energy to perform tasks (Weekly)	Never	1986 (82.99)	581 (89.94)	1405 (80.42)	30.253	< 0.001
	< 1 time	105 (4.39)	17 (2.63)	88 (5.04)		
	1–2 times	209 (8.73)	33 (5.11)	176 (10.07)		
	≥ 3 times	93 (3.89)	15 (2.32)	78 (4.47)		
Frequency of feeling sleepy (Weekly)	Never	1960 (81.91)	568 (87.93)	1392 (79.68)	22.676	< 0.001
	< 1 time	92 (3.85)	18 (2.79)	74 (4.24)		
	1–2 times	220 (9.19)	35 (5.42)	185 (10.59)		
	≥ 3 times	121 (5.06)	25 (3.87)	96 (5.50)		

Table 1 (continued)

Variable	Category	Total ($n=2393$)	Labor productivity loss		χ^2/t	*P*
			No ($n=646$)	Yes ($n=1747$)		
Sleep quality	Good	1999(83.54)	567 (87.77)	1432 (81.97)	11.542	< 0.001
	Poor	394(16.465)	79 (12.23)	315 (18.03)		
Sleep onset difficulty (Weekly)	Rarely	2001 (83.62)	570 (88.24)	1431 (81.91)	16.917	< 0.001
	< 1 time	72 (3.01)	20 (3.10)	52 (2.98)		
	1–2 times	320 (13.37)	56 (8.67)	264 (15.11)		
Noontime break habit	No break	529 (22.11)	164 (25.39)	365 (20.89)	5.775	0.056
	Seasonal (Typically in summer)	224 (9.36)	61 (9.44)	163 (9.33)		
	All-year round	1640 (68.53)	421 (65.17)	1219 (69.78)		
Physical activity level	Sedentary	1445 (60.38)	444 (68.73)	1001 (57.30)		< 0.001*
	Slightly active	15 (0.63)	7 (1.08)	8 (0.46)		
	Moderately active	918 (38.36)	194 (30.03)	724 (41.44)		
	Vigorously active	15 (0.63)	1 (0.16)	14 (0.80)		

Note: '*' indicates the *p*-values calculated using Fisher's exact test

models. The GNB model demonstrated the highest AUC of 0.722 (95%CI: 0.680–0.763) and achieved the best accuracy at 86.3% and sensitivity at 91.8%, although it had a relatively lower specificity of 71.5%, as shown in Table 3; Fig. 2. Similarly, the SVM model exhibited consistent predictive ability, with an AUC of 0.693 (95%CI: 0.647–0.738), and corresponding accuracy, sensitivity, specificity, and F1-score of 77.5%, 57.1%, 91.5%, and 67.3%, respectively. The AUC values for RF, XGBoost, MLP, and LR models were 0.646 (95%CI: 0.601–0.690), 0.643 (95%CI: 0.596–0.689), 0.692 (95%CI: 0.648–0.736), and 0.720 (95%CI: 0.677–0.763) respectively. Although XGBoost had the highest sensitivity at 97.9%, it exhibited a significantly lower specificity rate of 55.4%, indicating a potential bias.

In the validation set, GNB model continued to perform well, with an AUC of 0.681 (95%CI: 0.670–0.794) and an accuracy of 82.5%. Other models such as SVM, RF, XGBoost, MLP, and LR had AUC values of 0.662 (95%CI: 0.594–0.730), 0.663 (95% CI: 0.598–0.729), 0.660 (95% CI: 0.594–0.727), 0.678 (95%CI: 0.613–0.743), and 0.687 (95% CI: 0.623–0.751), respectively. The MLP model performed moderately on the validation set, with an accuracy of 72.1% and an F1-score of 80.1%. Overall, the GNB model outperformed the others and was selected as the top choice for classification modeling, followed by the LR and SVM models. Consequently, the GNB model was chosen to build a SHAP model to enhance interpretability.

Visualization of feature importance

To assess the impact of various feature variables on our model's output, we applied SHAP values to interpret the GNB model used in this study. Features were ranked based on their average SHAP values in descending order of influence. As shown in Fig. 3, occupational health training was identified as the most significant factor affecting the productivity of petrochemical workers under high temperatures (SHAP = -3.56), followed by exposure to noise (SHAP=2.64) and the availability of protective measures (SHAP = -2.61). Protective factors that significantly preserved work efficiency under these conditions included occupational health training (SHAP = -3.56), implementation of protective measures (SHAP = -2.61), engaging in less physically demanding roles (SHAP = -1.75), being female (SHAP = -1.37), taking a midday break (SHAP = -0.53), and belonging to higher income brackets (SHAP = -0.57). Conversely, factors that negatively impacted work efficiency included the lack of air conditioning (SHAP=1.92), persistent noise (SHAP=2.64), exposure to vibration (SHAP=1.15), and exposure to dust and smoke (SHAP=0.95). Additional detrimental influences included work-environment induced fatigue (SHAP=0.49), poor sleep quality (SHAP=0.98), involvement in physically exhausting tasks (SHAP=0.81), longer weekly work hours (SHAP=1.10), higher BMI (SHAP=0.59), and working in confined spaces (SHAP=0.80). These factors collectively contributed to significant reductions in labor productivity.

Discussion

Petrochemical industry plays a major role in global economy and growth [39]. The climate crisis highlights the urgent need to assess the impact of rising temperatures on labor productivity loss. Results of this study suggested that heat exposure significantly reduced the efficiency of petrochemical workers, who are vital to meeting the growing energy demands. Using sophisticated machine learning and deep learning techniques, this study investigated the self-reported productivity loss due to heat exposure among petrochemical workers. Moreover, through SHAP value analysis of the best-performing model (GNB), this study identified the key factors that compromise productivity when working

Table 2 Occupational information *n* (%) of surveyed petrochemical workers

Variable	Category	Total ($n=2393$)	Labor productivity loss		χ^2/t	*P*
			No ($n=646$)	Yes ($n=1747$)		
Nature of work	Petroleum refining	381 (15.92)	89 (13.78)	292 (16.71)	7179	< 0.001
	Aromatics and olefin production	407 (17.01)	76 (11.65)	331 (18.90)		
	Storage and transportation	456 (19.06)	117 (18.11)	339 (19.41)		
	Service operation	1149 (48.02)	364 (56.35)	785 (44.93)		
Length of service	(Years, Mean ± SD)	14.61 ± 3.77	13.91 ± 11.27	15.28 ± 12.15	-2.043	0.041
Weekly working hours	(Hours, Mean ± SD)	46.74 ± 16.59	44.45 ± 7.62	48.92 ± 6.55	-2.659	0.048
Occupational classification	Frontline workers (External Ops)	1589 (66.40)	425 (65.79)	1164 (66.63)	43.303	< 0.001
	Frontline workers (Internal Ops)	305 (12.75)	78 (12.07)	227 (12.99)		
	Management/supervisory	351 (14.67)	107 (16.56)	244 (13.97)		
	Administrative staff	148 (6.19)	36 (5.57)	112 (6.41)		
Greenness at workplace	No	246 (10.28)	74 (11.46)	172 (9.85)	1.325	0.250
	Yes	2147 (89.72)	572 (88.55)	1575 (90.16)		
Heat control measures	Electric fan & air conditioning	313 (13.08)	106 (16.41)	207 (11.85)	10.132	0.017
	Electric fan only	84 (3.51)	26 (4.03)	58 (3.32)		
	Air conditioning only	1513 (63.23)	395 (61.15)	1118 (63.99)		
	None	483 (20.18)	119 (18.42)	364 (20.84)		
Vibration exposure	No	1397 (58.38)	471 (72.91)	926 (53.01)	76.902	< 0.001
	Yes	996 (41.62)	175 (27.09)	821 (46.99)		
Noise exposure	No	865 (36.15)	335 (51.86)	530 (30.34)	94.625	< 0.001
	Yes	1528 (63.85)	311 (48.14)	1217 (69.66)		
Dust and fumes exposure	No	1645 (68.74)	491 (76.01)	1154 (66.06)	21.730	< 0.001
	Yes	748 (31.26)	155 (23.99)	593 (33.94)		
Oil mist exposure	No	2166 (90.51)	611 (94.58)	1555 (89.01)	17.055	< 0.001
	Yes	227 (9.49)	35 (5.42)	192 (10.99)		
Physical strain level	Very severe	7 (0.29)	0 (0.000)	7 (0.40)		< 0.001*
	Severe	143 (5.98)	17 (2.63)	126 (7.21)		
	Moderate	1238 (51.73)	259 (40.09)	979 (56.04)		
	Mild	889 (37.15)	317 (49.07)	572 (32.74)		
	None	116 (4.85)	53 (8.20)	63 (3.61)		
Operate in confined space	Not required	1923 (80.36)	560 (86.69)	1363 (78.02)	22.756	< 0.001
	Occasionally (1–2 times per month)	424 (17.72)	76 (11.77)	348 (19.92)		
	Often(at least 3 times per week)	46 (1.92)	10 (1.55)	36 (2.06)		
Occupational health training	No	180 (7.52)	86 (13.31)	94 (5.38)	42.656	< 0.001
	Yes	2213 (92.48)	560 (86.69)	1653 (94.62)		
Use of PPE	No	163 (6.81)	90 (13.93)	73 (4.18)	70.677	< 0.001
	Yes	2230 (93.19)	556 (86.07)	1674 (95.82)		

Note: *' indicates the *p*-values calculated using Fisher's exact test

in hot environments: heat-related health training, the implementation of protective measures, and the number of work hours per week. Findings of this study may provide valuable insights for the development of targeted strategies to bolster workers heat resilience and maintain sustained productivity amidst the challenges posed by a changing climate.

Model construction and evaluation

This research applied an array of machine learning, deep learning, and conventional statistical methods to construct various models for the investigation of factors affecting the work efficiency of petrochemical workers exposed to occupational high temperatures. Our findings suggested that each of the six models exhibited commendable authenticity, reliability, and predictive effectiveness, with the GNB model showing the highest efficacy, followed by LR, SVM, and MLP, while the RF model showed the lowest efficacy in this study. The superior performance of the GNB model in this study aligns with the findings of Taghizadeh et al. [40]. Similarly, Lips et al. found the LR model outperformed the MLP algorithm [41], which is consistent with our results. Although the MLP model has strong non-linear mapping, self-learning, and adaptive capabilities, Boudreault et al. noted that it struggled during summertime modeling [18], a finding echoed by our research. Moreover, Boudreault et al. observed improved calibration with larger

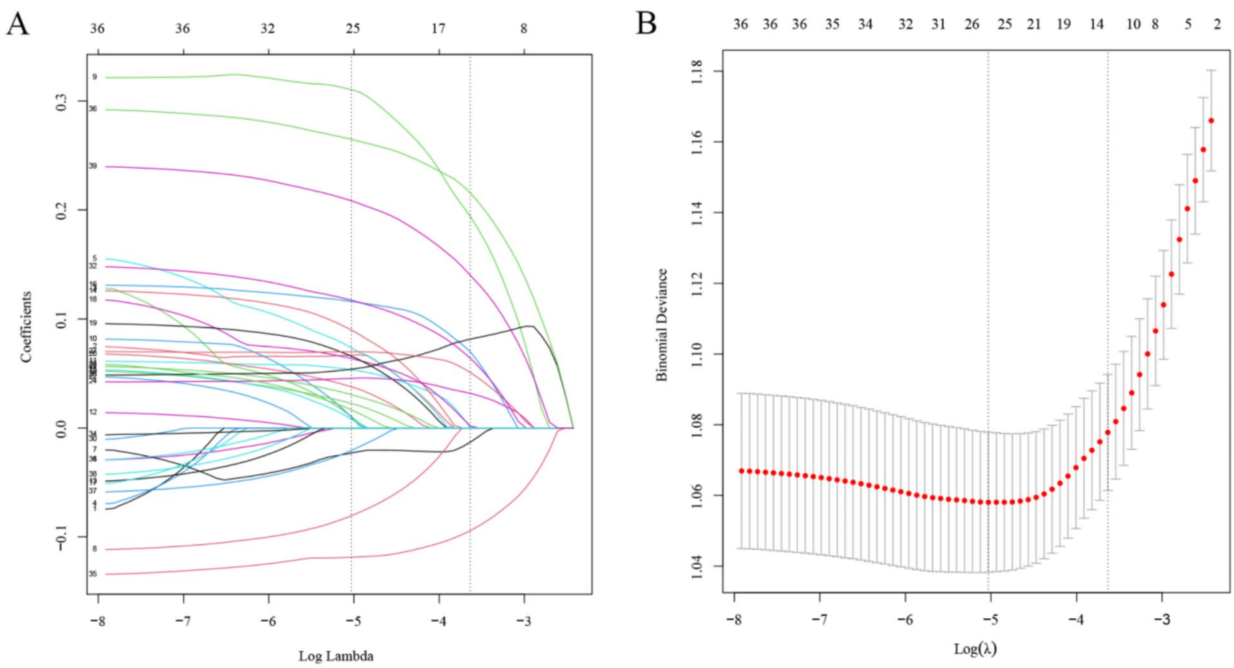

Fig. 1 Lasso regression feature selection for temperature discomfort and labor productivity. (**A**) Spatial distribution of Lasso coefficients for 27 predictors was depicted by lines of varying colors, each representing a distinct predictor. (**B**) Lasso coefficient paths were plotted as functions of the log-transformed regularization parameter (lambda). Optimal model parameters associated with specific lambda values were highlighted with red markers.

Table 3 Performance comparison of different models in the training and validation sets

Model	AUC (95%CI)	Accuracy	Sensitivity	Specificity	F1-Score	Cutoff
Training Set						
SVM	0.693 (0.647–0.738)	0.775	0.571	0.915	0.673	0.770
RF	0.646 (0.601–0.690)	0.760	0.774	0.750	0.750	0.650
XGBoost	0.643 (0.596–0.689)	0.727	0.979	0.554	0.745	0.678
GNB	0.722 (0.680–0.763)	0.863	0.918	0.715	0.642	0.869
MLP	0.692 (0.648–0.736)	0.738	0.769	0.653	0.811	0.699
LR	0.720 (0.677–0.763)	0.680	0.692	0.672	0.638	0.724
Validation set						
SVM	0.662 (0.594–0.730)	0.605	0.599	0.586	0.655	0.768
RF	0.663 (0.598–0.729)	0.674	0.579	0.574	0.654	0.620
XGBoost	0.660 (0.594–0.727)	0.616	0.711	0.470	0.743	0.894
GNB	0.681 (0.670–0.794)	0.825	0.812	0.713	0.624	0.851
MLP	0.678 (0.613–0.743)	0.721	0.762	0.624	0.801	0.678
LR	0.687 (0.623–0.751)	0.680	0.592	0.692	0.672	0.638

datasets, suggesting the need for further validation to assess performance. Marien et al.'s investigation into the annual modeling of myocardial infarction using various machine learning and deep learning approaches found that the MLP model's efficacy surpassed that of RF [42], paralleling our study's outcome. The RF model, which seeks to maximize predictive accuracy by reducing variance rather than bias, has been recognized for its strong predictive capacity, as identified by Stretch et al. RF [43], corroborating the results of our study.

Analysis of influencing factors

Heat-related training

This study found that workers who received adequate occupational health training were better equipped to manage heat stress when working in high-temperature environments, thereby reducing heat-attributed productivity loss. Compliance with heat prevention measures is crucial for mitigating the adverse effects of temperature discomfort on productivity [44]. Workers exposed to high temperatures may experience various health issues, including heat exhaustion [45], stroke [46], and nephrolithiasis [47]. Proactive protective strategies—such as

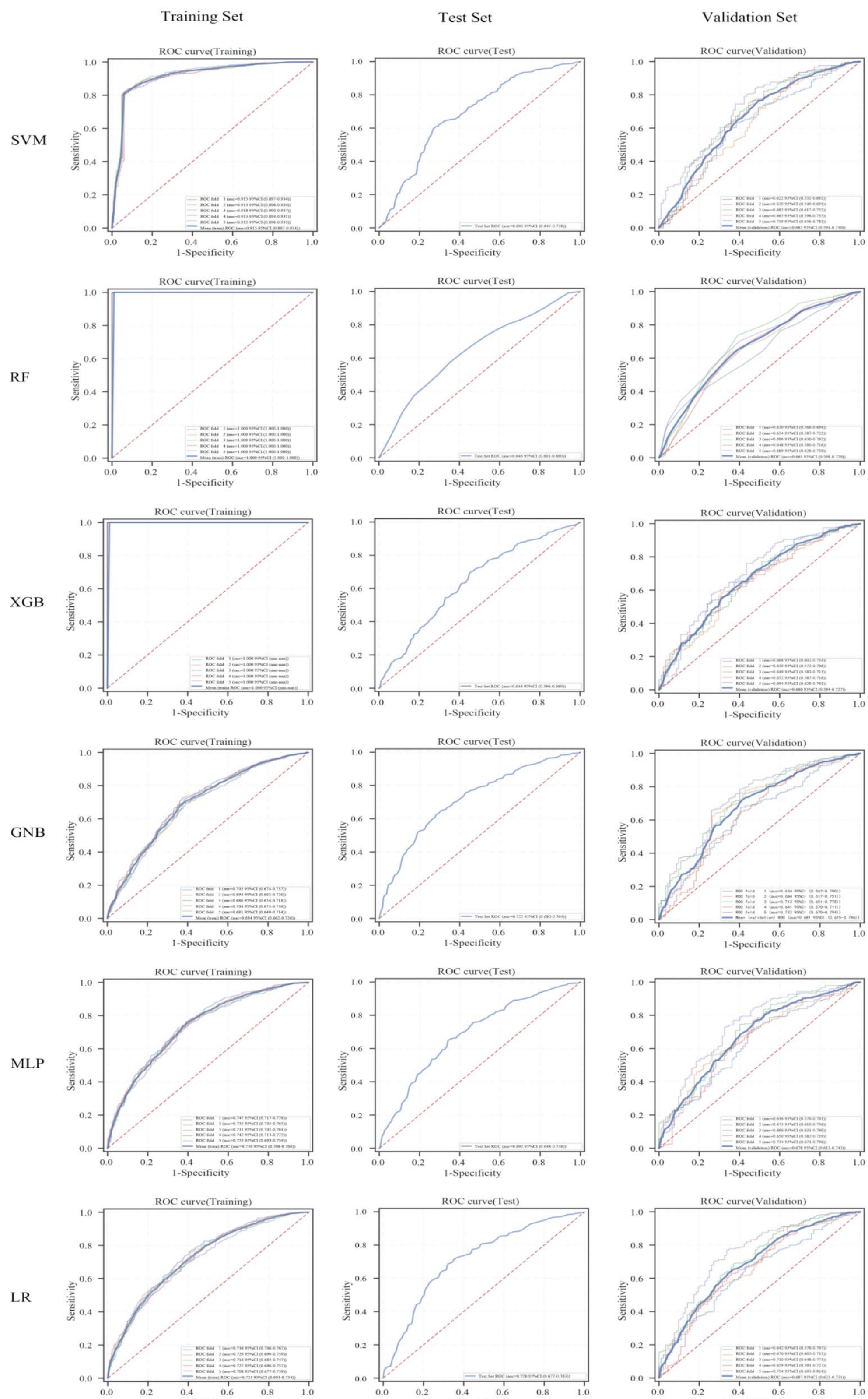

Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 ROC curves of SVM, RF, XGBoost, GNB, MLP, and LR models across training, testing, and validation sets. This figure shows the assessment of six prominent machine learning models using their ROC curves and AUC values across three distinct dataset partitions: training, testing, and validation. The evaluated models include Support Vector Machine (SVM), Random Forest (RF), Extreme Gradient Boosting (XGB), Gaussian Naive Bayes (GNB), Multilayer Perceptron (MLP), and Logistic Regression (LR). The ROC curves represent a graphical comparison of the true positive rate (sensitivity) against the false positive rate (1-specificity) at various threshold levels, offering an aggregate measure of model performance throughout all classification thresholds. A higher AUC value signifies a model's superior capability to differentiate between classes accurately. This comparative visualization facilitates a critical analysis of each model's predictive efficiency and generalizability across different data segments, underscoring the importance of model selection in the application of machine learning for classification tasks

wearing breathable, moisture-wicking and heat-dissipating work attire; taking regular breaks for hydration; and maintaining moderate work intensity—can effectively prevent and mitigate these health issues due to temperature discomfort.

Moreover, reinforcing occupational health training and protective measures is vital for long-term vocational development. A randomized controlled trial by Harsini et al. revealed that approximately 90% of workplace accidents were due to unsafe behaviors [48], underscoring the urgent need for immediate implementation of robust occupational health education initiatives. Workers consistently exposed to high-temperature environments are prone to cumulative health effects from ongoing heat exposure. However, through occupational health education, they can learn how to reduce and manage the impact of heat exposure, thus reducing or preventing the exacerbation of potential health issues. Therefore, a strong commitment to occupational safety and health management in the workplace is both essential and urgent.

Duration of weekly working hours

We found the duration of weekly working hours is a significant determinant influencing labor productivity loss. Extended working hours and work intensity can exacerbate the physical strain on petrochemical workers, thereby negatively affecting their job performance and productivity. Evidence has shown an inverse relationship between the duration of vacation time and the incidence of occupational injuries in high-temperature jobs [49]. Similarly, Li's study demonstrated an inverse correlation between working hours and labor productivity loss under high temperatures [50], aligning with the findings of our research. Potential factors contributing to this relationship include: heightened physical and mental exhaustion [6], reduced attention, increased psychological burden [51], and heightened psychological stress, all of which can compromise work efficiency in hot conditions.

Moreover, temperature discomfort may diminish the human body's physiological adaptability, affecting functions like thermoregulation and cardiovascular systems [52], further impacting the workers' operational state and efficiency. Therefore, in management practices, it is essential to arrange reasonable working hours, avoid excessively long continuous work periods, and provide adequate rest and adjustment measures to enhance the efficiency and health of workers in high-temperature environments.

Co-exposure to other occupational hazards

Petrochemical workers are generally exposed to multiple types of occupational hazards simultaneously in the workplace. Our findings indicated that the efficiency of petrochemical workers exposed to high temperatures was also influenced by occupational hazards such as noise, vibration, dust, and fumes encountered during production operations. Exposure to these occupational hazards not only has detrimental effects on workers' health [53], such as damage to the eardrum from noise [54], impairment of the nervous and circulatory systems from vibrations [55], respiratory diseases, but also lead to psychological fatigue, anxiety, and stress responses, which in turn, can impair workers' concentration and state of mind [56]. Prolonged exposure to dust and fumes may also cause emotional fluctuations and discomfort among workers [56], further diminishing work efficiency.

The combined effects of occupational hazards such as noise, dust, fumes, and vibration, along with heat, may create a complex risk environment that can significantly impact petrochemical workers' health and efficiency. Workers exposed to multiple hazards are more vulnerable because their bodies are under greater stress, making it harder to regulate temperature and maintain hydration [57]. In noisy and hot environments, workers may experience higher levels of discomfort, leading to quicker fatigue and reduced concentration. Inability to communicate effectively in a noisy environment can also prevent timely interventions for heat-related illnesses. Dust particles are more likely to be inhaled in hot conditions due to increased breathing rates, which may aggravate respiratory conditions [58]. Heat can increase the volatility of certain chemicals, resulting in increased inhalation and absorption of toxic substances [59]. The physical strain from operating vibrating machinery is amplified in hot conditions, leading to faster onset of fatigue, reduced productivity, and a higher likelihood of mistakes and accidents. Therefore, effective management strategies and technical measures, including improving the work environment, providing essential personal protective equipment, conducting health monitoring, and offering relevant training, are necessary to safeguard the health

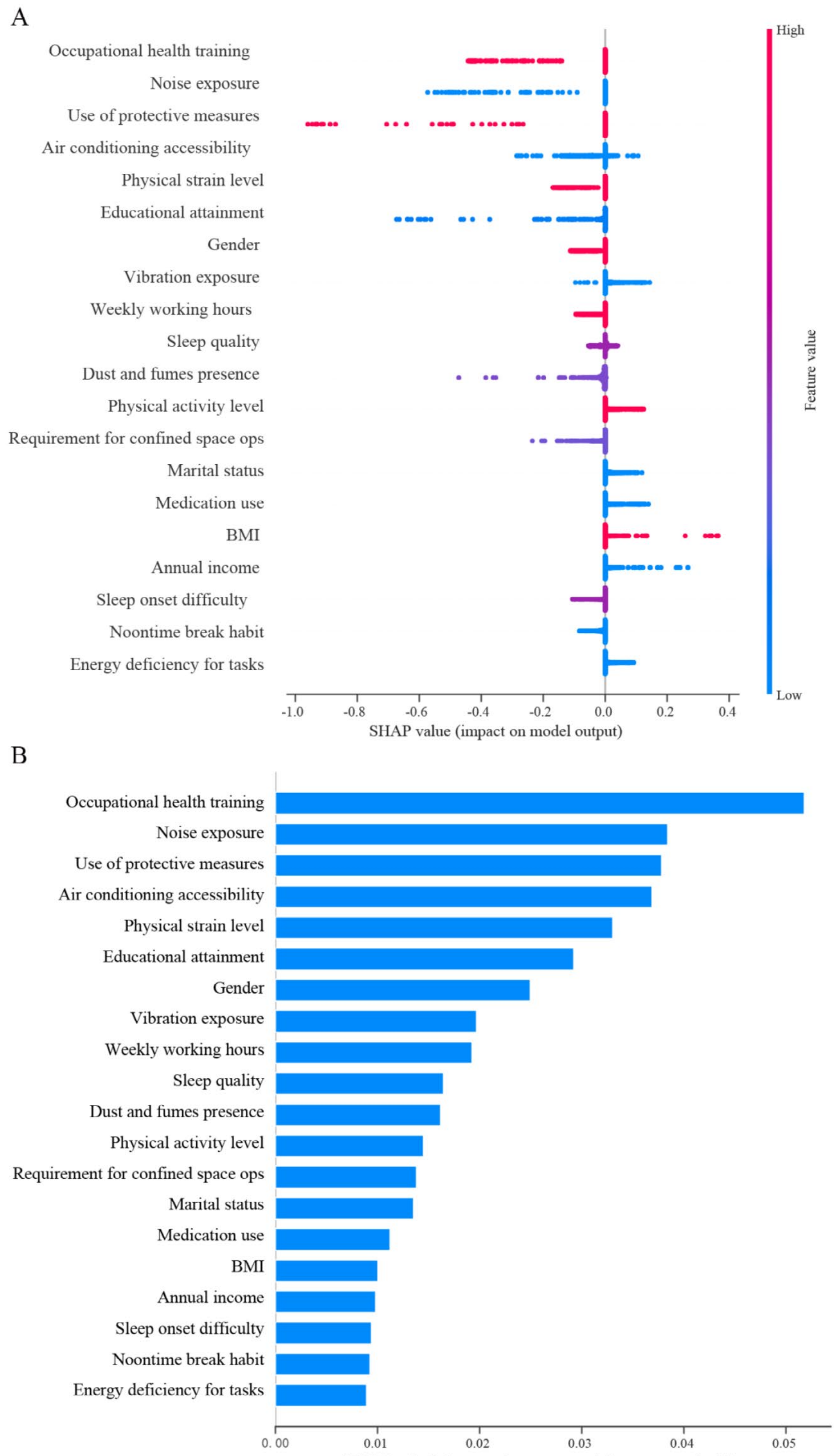

Fig. 3 SHAP summary plot for GNB model of petrochemical workers' productivity loss. (**A**) SHAP value impact on petrochemical workers' productivity: This plot illustrates the relationship between feature-wise SHAP values and the workers' productivity. Features with greater SHAP values exert a more significant influence on productivity enhancement, depicted by a transition from blue (lower impact) to red (higher impact) across the axis. (**B**) Feature importance landscape: Exhibits the ranked importance of features based on the mean absolute SHAP values, providing a quantitative analysis of each feature's contribution to predictive accuracy

and work efficiency of petrochemical workers in high-temperature environments.

Personal characteristics

Furthermore, we found personal characteristics including education level, gender, sleep quality, marital status, disease presence, BMI, and annual household income have significantly impacted workers' productivity. These findings underscored the potential influences of individual traits on job performance. Regarding gender differences, this study suggested that male workers were more susceptible to the impacts of high ambient temperatures, which is supported by McInnes in earlier research [60]. Firstly, men often engage in more physically demanding tasks than women [61], requiring increased physical exertion and endurance. In high-temperature environments, heat stress exacerbates the body's burden, making them more prone to fatigue and loss of energy. Secondly, males are physiologically more sensitive to high temperatures. Compared to females, males were proved to have poorer temperature regulation capabilities [62, 63], making them more vulnerable to heat-related illnesses. This may lead to discomfort and fatigue in high-temperature environments, thereby affecting work efficiency. Additionally, males exhibit higher sweat production under heat stress conditions [64]. While sweating helps dissipate heat, excessive sweating can result in fluid loss and dehydration, ultimately impacting bodily functions and concentration [65]. These factors may contribute to decreased work efficiency among male petrochemical workers in high-temperature environments. Nevertheless, it is important to note that these observations represent general trends, and individual variability persists. The implementation of adequate protective protocols, judicious work-rest cycles, and comprehensive health education is essential to mitigate the impact of high-temperature conditions on male petrochemical workers.

Regarding BMI, this study observed that workers who experienced productivity loss exhibited higher BMI levels compared to those without productivity loss, both in the normal BMI group and the obese group. Li et al. indicated that BMI could reduce labor productivity in hot environments [50], which is consistent with the results of this study. This finding aligns with research by Giersch et al., which documented a positive correlation between exertional heat stress (EHS) and BMI, indicating a 3% increase in the risk of EHS for every unit increase in BMI [66]. The reasons behind these observations could be multifactorial. One possible explanation is the physiological burden imposed by higher BMI, which can exacerbate the cardiovascular [67] and metabolic strain [68] during physical exertion, thereby reducing work efficiency. Additionally, individuals with higher BMI may have a reduced capacity for heat dissipation [69], increasing their susceptibility to heat-related stress and its subsequent impact on productivity. Biomechanical inefficiencies associated with obesity might also contribute to quicker fatigue onset [70], impacting overall work performance. Further research is essential to elucidate the detailed mechanisms and explore potential strategic interventions to mitigate these effects.

Interestingly, the results of univariate analysis in this study indicated that alcohol consumption was associated with productivity loss among workers. However, workers who drank almost daily exhibited a lower incidence of productivity loss compared to those who never drank. Theoretically, alcohol can exacerbate heat strain and subsequently reduce work efficiency [57]. However, multiple factors in the workplace are associated with productivity loss, such as job requirements and heat control measures. Most participants who drank daily were frontline workers. Probably, they did not report the truth as alcohol drinking is prohibited in the workplace, especially for petrochemical workers. In addition, only 33 workers reported "drink daily" in this study. The relatively small sample size may also contribute to this phenomenon. If divide the alcohol consumption into two options (less than once per month, at least once per month), we found workers who drank at least once per month had a higher percentage of productivity loss compared to those drank less frequently. Since frontline workers tend to have higher physical workloads, they are more vulnerable to heat strain and related productivity loss [71]. Therefore, further investigation into the role of occupational settings and responsibilities may provide deeper insights into strategies that can mitigate productivity loss because of alcohol consumption.

With respect to the influence of sleep quality on work efficiency, research by Tawatsupa et al. has demonstrated that sleep quality in hot environments plays a crucial role in overall productivity [72], which aligns with the findings of the current study. Adequate and restorative sleep is essential for the recovery and regulation of all body systems [73]. In high-temperature conditions, petrochemical plant workers may experience increased physical fatigue and dehydration. Good quality sleep is imperative for restoring physical energy, maintaining physiological stability, and preserving emotional balance and cognitive functioning. Conversely, poor sleep can exacerbate physical exhaustion, reduce stamina, and lead to heightened anxiety, increased stress, and greater emotional variability [74], all contributing to suboptimal mental states during work hours and difficulty in sustaining work efficiency.

Limitations

This study has certain limitations that should be acknowledged. First, the study was conducted in Quangang, a region characterized by a subtropical climate, which may limit the generalizability of the findings to areas with different climatic conditions. Second, thermal comfort is essentially determined by external climatic conditions, work clothes, use of PPE, and individual's adaptability. In this study, heat exposure and other occupational hazards exposure information were collected through a self-reported questionnaire survey. This is a potential risk of recall bias in terms of exposure level. Moreover, we did not ask workers the specific types of PPE and work clothes they wore in the workplace. Different types of PPE and work clothes may have different effects on thermal comfort and heat-related labor productivity loss, which further research is warranted. Third, certain medications (e.g. diuretics, beta-blockers, anticholinergics, and antipsychotics) can modulate thermoregulatory function, resulting in excess physiological strain and predisposing workers to adverse health outcomes [75]. The impact of different medications on heat-related productivity loss may vary, depending on multiple factors such as the type of medication, dosage and duration of use. In this survey, however, we did not collect this specific information.

Conclusion

This study highlights a significant decline in work efficiency among petrochemical workers due to heat exposure in the workplace. To mitigate the likely increasing productivity loss attributed to heat in the context of climate change, it is imperative to consider several proactive measures. These may include improving occupational health training, implementing rigorous noise pollution control measures, establishing climate-controlled work environments, adapting labor-intensive tasks, and strengthening educational programs within the petrochemical industry. Prioritizing these initiatives is warranted to better safeguard the well-being and productivity of workers in the petrochemical sector under challenging environmental conditions.

Abbreviations

SVM	Support Vector Machine
RF	Random Forest
XGBoost	eXtreme Gradient Boosting
GNB	Gaussian Naive Bayes
MLP	Multilayer Perceptron
LR	Logistic Regression
SHAP	SHapley Additive exPlanations
ML	Machine Learning
DL	Deep Learning

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12889-024-20713-4.

Supplementary Material 1

Acknowledgements

The authors would like to thank the Minnan Branch of the First Affiliated Hospital of Fujian Medical University for their support during the survey. All survey participants and the interviewers (Wenzhu Chen, Suqun Chen, Ying Xu, Xiaoyan Zeng, Yian Guo, Sufen Ke, Suyu Chen, and Meimei Lin) are greatly appreciated for their contributions.

Author contributions

YZ conducted data analysis and drafted the manuscript. YC and QL conducted the majority of the field work, including data collection and initial analyses. SD contributed to the interpretation of results and manuscript revisions. YY and ZZ assisted in the field investigations and provided critical revisions to the draft. JC, ZX, and JW were involved in the data collection as well as offering substantive suggestions for the manuscript. WZ participated extensively in designing research methodology and supervising the field investigations. LW supported the study design and execution, and provided guidance on statistical analysis. WY contributed significantly to the manuscript concept, reviewing literature, and editing the manuscript content for important intellectual content. JX, as the corresponding author, played a pivotal role in the conception of the article, oversaw the entire project, contributed to manuscript writing and revisions, and undertook additional data collection. All authors discussed the results, commented on the manuscript at all stages, and read and approved the final manuscript.

Funding

This study was supported by the Minjiang Scholar Start-up Research Fund of Fujian Province (Grant No. 2019-9202001001) and 2021 Natural Science Foundation of Fujian Province of China (2021J01722).

Data availability

Data is provided within the manuscript or supplementary information files. The datasets are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

The study protocol was approved by the Ethics Committee of Fujian Medical University, with approval number (2022–111). All participants provided written informed consents before participating in the study. The research was conducted in accordance with the ethical guidelines and principles outlined in the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Preventive Medicine, School of Public Health, Fujian Medical University; and Key Laboratory of Environment and Health, Fujian Province University, 1 North Xue-Fu Rd, Minhou, Fuzhou 350122, Fujian Province, China
[2]Department of Epidemiology and Health Statistics, School of Public Health, Fujian Medical University, Fuzhou 350122, Fujian Province, China
[3]Minnan Branch of the First Affiliated Hospital of Fujian Medical University, Quangang, Quanzhou 362100, Fujian Province, China
[4]School of Public Health, The University of Adelaide, North Terrace Campus, Adelaide, South Australia 5005, Australia

Received: 27 May 2024 / Accepted: 12 November 2024
Published online: 25 November 2024

References

1. Tong R, Yang Y, Shao G, Zhang Y, Dou S, Jiang W. Emission sources and probabilistic health risk of volatile organic compounds emitted from production areas in a petrochemical refinery in Hainan, China. Hum Ecol Risk Assessment: Int J. 2020;26(5):1407–27. https://doi.org/10.1080/10807039.2019.1579049

2. Tong LZ, Pu ZM, Chen K, Yi JJ. Sustainable maintenance supplier performance evaluation based on an extend fuzzy PROMETHEE II approach in petrochemical industry. Journal of Cleaner Production. 2020; 273. httpshttps://doi.org/10.1016/j.jclepro.2020.122771
3. Sevan G, Pouya M, Ehsan A, Hoda N, Abbas Abbas R, Hamid M. Biological treatment of toxic refinery spent sulfidic caustic at low dilution by sulfur-oxidizing fungi. Journal of environmental chemical engineering. 2018. httpshttps://doi.org/10.1016/j.jece.2018.04.026
4. Zhang Y, Liu Y, Li ZX, Liu X, Chen QF, Qin JY, Liao QL, Du R, Deng QF, Xiao YM, et al. Effects of coexposure to noise and mixture of toluene, ethylbenzene, xylene, and styrene (TEXS) on hearing loss in petrochemical workers of southern China. Environ Sci Pollut Res. 2023;30(11):31597–607. https://doi.org/10.1007/s11356-022-24414-6
5. Rathod SB, Sorte SR, Patel S. The Effect of High Temperature on Cardiovascular autonomic function tests in steel plant furnace worker. Indian J Occup Environ Med. 2021;25(2):67–71. https://doi.org/10.4103/ijoem.IJOEM_193_20
6. Martínez-Solanas È, López-Ruiz M, Wellenius GA, Gasparrini A, Sunyer J, Benavides FG, Basagaña X. Evaluation of the impact of ambient temperatures on Occupational Injuries in Spain. Environ Health Perspect. 2018;126(6):067002. https://doi.org/10.1289/ehp2590
7. Wang L, Yu M, Zhang S, Li X, Yuan J. Associations of Occupational Heat stress and noise exposure with carotid atherosclerosis among Chinese steelworkers: a cross-sectional survey. Int J Environ Res Public Health. 2021;19(1). https://doi.org/10.3390/ijerph19010024
8. Zhou F, Shrestha A, Mai S, Tao Z, Li J, Wang Z, Meng X. Relationship between occupational noise exposure and hypertension: a cross-sectional study in steel factories. Am J Ind Med. 2019;62(11):961–8. https://doi.org/10.1002/ajim.23034
9. Jafari MJ, Pirposhteh EA, Dehghan SF, Khodakarim S, Jafari M. Relationship between heat stress exposure and some immunological parameters among foundry workers. Int J Biometeorol. 2020;64(5):853–61. https://doi.org/10.1007/s00484-020-01874-4
10. Borg MA, Xiang J, Anikeeva O, Ostendorf B, Varghese B, Dear K, Pisaniello D, Hansen A, Zander K, Sim MR, et al. Current and projected heatwave-attributable occupational injuries, illnesses, and associated economic burden in Australia. Environ Res. 2023;236(Pt 2):116852. https://doi.org/10.1016/j.envres.2023.116852
11. Borg MA, Xiang J, Anikeeva O, Pisaniello D, Hansen A, Zander K, Dear K, Sim MR, Bi P. Occupational heat stress and economic burden: a review of global evidence. Environ Res. 2021;195:110781. https://doi.org/10.1016/j.envres.2021.110781
12. Dasgupta S, van Maanen N, Gosling SN, Piontek F, Otto C, Schleussner CF. Effects of climate change on combined labour productivity and supply: an empirical, multi-model study. Lancet Planet Health. 2021;5(7):E455–65.
13. Shen RZ, Ye ZC, Gao J, Hou YP, Ye HC. Climate change risk perception in global: correlation with petroleum and liver disease: a meta-analysis. Ecotoxicol Environ Saf. 2018;166:453–61. https://doi.org/10.1016/j.ecoenv.2018.09.080
14. Shin-Li L. Integrating heuristic time series with modified grey forecasting for renewable energy in Taiwan. Renewable Energy. 2019. httpshttps://doi.org/10.1016/j.renene.2018.08.092
15. Wang Z, Han YM, Li CF, Geng ZQ, Fan JZ. Input-output networks considering graphlet-based analysis for production optimization: Application in ethylene plants. Journal of Cleaner Production. 2021; 278. httpshttps://doi.org/10.1016/j.jclepro.2020.123955
16. Yang L, Liu X, Zhu W, Zhao L, Beroza GC. Toward improved urban earthquake monitoring through deep-learning-based noise suppression. Sci Adv. 2022; 8(15):eabl3564. httpshttps://doi.org/10.1126/sciadv.abl3564
17. Lee M, Yeo NY, Ahn HJ, Lim JS, Kim Y, Lee SH, Oh MS, Lee BC, Yu KH, Kim C. Prediction of post-stroke cognitive impairment after acute ischemic stroke using machine learning. Alzheimers Res Ther. 2023;15(1):147. https://doi.org/10.1186/s13195-023-01289-4
18. Boudreault J, Campagna C, Chebana F. Machine and deep learning for modelling heat-health relationships. Sci Total Environ. 2023;892:164660. https://doi.org/10.1016/j.scitotenv.2023.164660
19. Li X, Zhao Y, Zhang D, Kuang L, Huang H, Chen W, Fu X, Wu Y, Li T, Zhang J, et al. Development of an interpretable machine learning model associated with heavy metals' exposure to identify coronary heart disease among US adults via SHAP: findings of the US NHANES from 2003 to 2018. Chemosphere. 2023;311(Pt 1):137039. https://doi.org/10.1016/j.chemosphere.2022.137039
20. Ogata S, Takegami M, Ozaki T, Nakashima T, Onozuka D, Murata S, Nakaoku Y, Suzuki K, Hagihara A, Noguchi T, et al. Heatstroke predictions by machine learning, weather information, and an all-population registry for 12-hour heatstroke alerts. Nat Commun. 2021;12(1):4575. https://doi.org/10.1038/s41467-021-24823-0
21. Sun DL, Gu QY, Wen HJ, Xu JH, Zhang YL, Shi SX, Xue MM, Zhou XZ. Assessment of landslide susceptibility along mountain highways based on different machine learning algorithms and mapping units by hybrid factors screening and sample optimization. Gondwana Res. 2023;123:89–106. https://doi.org/10.1016/j.gr.2022.07.013
22. Hu JC, Szymczak S. A review on longitudinal data analysis with random forest. Brief Bioinform. 2023;24(2). https://doi.org/10.1093/bib/bbad002
23. Jannusch K, Dietzel F, Bruckmann NM, Morawitz J, Boschheidgen M, Minko P, Bittner AK, Mohrmann S, Quick HH, Herrmann K, et al. Eur J Nucl Med Mol Imaging. 2023. https://doi.org/10.1007/s00259-023-06513-9. Prediction of therapy response of breast cancer patients with machine learning based on clinical data and imaging data derived frombreast < SUP > 18 F FDG-PET/MRI
24. Ali R, Hussain J, Lee SW. Multilayer perceptron-based self-care early prediction of children with disabilities. Digit Health. 2023;9:20552076231184054. https://doi.org/10.1177/20552076231184054
25. Xiong Y, Ma Y, Ruan L, Li D, Lu C, Huang L. Comparing different machine learning techniques for predicting COVID-19 severity. Infect Dis Poverty. 2022;11(1):19. https://doi.org/10.1186/s40249-022-00946-4
26. Quangang Petrochemical Industrial Park. http://www.enquanzhou.com/2019-11/19/c_425637.htm#:~:text=QuangangPetrochemical%20Industrial%20Park%20is,base%20in%20Quanzhou%2 C%20Fujian%20province.
27. Wang B, Wu C, Kang LG, Huang L, Pan W. What are the new challenges, goals, and tasks of occupational health in China's Thirteenth five-year plan (13th FYP) period? J Occup Health. 2018;60(3):208–28. https://doi.org/10.1539/joh.2017-0275-RA
28. de Vet HCW, Mokkink LB, Mosmuller DG, Terwee CB. Spearman-Brown prophecy formula and Cronbach's alpha: different faces of reliability and opportunities for new applications. J Clin Epidemiol. 2017;85:45–9. https://doi.org/10.1016/j.jclinepi.2017.01.013
29. Zhang Y, Xu J, Zhang C, Zhang X, Yuan XL, Ni WQ, Zhang HM, Zheng YJ, Zhao ZG. Community screening for dementia among older adults in China: a machine learning-based strategy. BMC Public Health. 2024;24(1). https://doi.org/10.1186/s12889-024-18692-7
30. Mao YK, Weng JY, Xie QY, Wu LD, Xuan YL, Zhang J, Han J. Association between dietary inflammatory index and stroke in the US population: evidence from NHANES 1999–2018. BMC Public Health. 2024;24(1). https://doi.org/10.1186/s12889-023-17556-w
31. Mueller-Using S, Feldt T, Sarfo FS, Eberhardt KA. Factors associated with performing tuberculosis screening of HIV-positive patients in Ghana: LASSO-based predictor selection in a large public health data set. BMC Public Health. 2016;16. https://doi.org/10.1186/s12889-016-3239-y
32. Teng F, Fan W, Luo Y, Xu S, Gong H, Ge R, Zhang X, Wang X, Ma L. A risk prediction model by LASSO for Radiation-Induced Xerostomia in patients with nasopharyngeal carcinoma treated with Comprehensive Salivary gland-sparing helical tomotherapy technique. Front Oncol. 2021;11:633556. https://doi.org/10.3389/fonc.2021.633556
33. Tay JK, Narasimhan B, Hastie T. Elastic Net Regularization paths for all generalized Linear models. J Stat Softw. 2023;106(1):1–31. https://doi.org/10.18637/jss.v106.i01
34. Gorji HT, Wilson N, VanBree J, Hoffmann B, Petros T, Tavakolian K. Using machine learning methods and EEG to discriminate aircraft pilot cognitive workload during flight. Sci Rep. 2023;13(1). https://doi.org/10.1038/s41598-023-29647-0
35. Chen M, Yin Z. Classification of Cardiotocography based on the Apriori Algorithm and Multi-model Ensemble Classifier. Front Cell Dev Biol. 2022;10:888859. https://doi.org/10.3389/fcell.2022.888859
36. Bustos A, Payá A, Torrubia A, Jover R, Llor X, Bessa X, Castells A, Carracedo Á, Alenda C. xDEEP-MSI: Explainable Bias-rejecting microsatellite instability Deep Learning System in Colorectal Cancer. Biomolecules. 2021;11(12). https://doi.org/10.3390/biom11121786
37. Wang HJ, Liang QX, Hancock JT, Khoshgoftaar TM. Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods. J Big Data. 2024;11(1). https://doi.org/10.1186/s40537-024-00905-w
38. Wu CD, Zhu JJ, Hsu CY, Shie RH. Quantifying source contributions to ambient NH3 using Geo-AI with time lag and parcel tracking functions. Environment International. 2024; 185. httpshttps://doi.org/10.1016/j.envint.2024.108520
39. Chung YM, Heshmati A. Measurement of environmentally sensitive productivity growth in Korean industries. J Clean Prod. 2015;104:380–91. https://doi.org/10.1016/j.jclepro.2014.06.030

40. Taghizadeh E, Heydarheydari S, Saberi A, JafarpoorNesheli S, Rezaeijo SM. Breast cancer prediction with transcriptome profiling using feature selection and machine learning methods. BMC Bioinformatics. 2022;23(1):410. https://doi.org/10.1186/s12859-022-04965-8
41. Lip GYH, Genaidy A, Tran G, Marroquin P, Estes C, Sloop S. Effects of multimorbidity on incident COVID-19 events and its interplay with COVID-19 event status on subsequent incident myocardial infarction (MI). Eur J Clin Invest. 2022;52(5):e13760. https://doi.org/10.1111/eci.13760
42. Marien L, Valizadeh M, Castell Wz, Nam C, Rechid D, Schneider A, Meisinger C, Linseisen J, Wolf K, Bouwer LM. Machine learning models to predict myocardial infarctions from past climatic and environmental conditions. Nat Hazards Earth Syst Sci. 2022;22(1561–8633):3015–39. https://doi.org/10.5194/nhess-22-3015-2022
43. Stretch R, Ryden A, Fung CH, Martires J, Liu S, Balasubramanian V, Saedi B, Hwang D, Martin JL, Della Penna N, et al. Predicting Nondiagnostic Home Sleep Apnea tests using machine learning. J Clin Sleep Med. 2019;15(11):1599–608. https://doi.org/10.5664/jcsm.8020
44. Nunfam VF, Adusei-Asante K, Frimpong K, Van Etten EJ, Oosthuizen J. Barriers to occupational heat stress risk adaptation of mining workers in Ghana. Int J Biometeorol. 2020;64(7):1085–101. https://doi.org/10.1007/s00484-020-01882-4
45. Weinberger KR, Tamburic L, Peters CE, McLeod CB. Heat-related illness among workers in British Columbia, 2001–2020. J Occup Environ Med. 2023;65(2):E88–92. https://doi.org/10.1097/jom.0000000000002761
46. Buller M, Fellin R, Bursey M, Galer M, Atkinson E, Beidleman BA, Marcello MJ, Driver K, Mesite T, Seay J, et al. Gait instability and estimated core temperature predict exertional heat stroke. Br J Sports Med. 2022;56(8):446–. https://doi.org/10.1136/bjsports-2021-104081
47. Lu IC, Yang CC, Huang CH, Chen SY, Lin CW, Lin CH, Chuang HY. The risk factors for Radiolucent Nephrolithiasis among workers in High-Temperature workplaces in the Steel Industry. Int J Environ Res Public Health. 2022;19(23). https://doi.org/10.3390/ijerph192315720
48. Harsini AZ, Ghofranipour F, Sanaeinasab H, Shokravi FA. A randomised controlled trial of an educational intervention to promote safe behaviours in petrochemical workers: a study protocol. BMC Public Health. 2019;19. https://doi.org/10.1186/s12889-019-7126-1
49. Marinaccio A, Scortichini M, Gariazzo C, Leva A, Bonafede M, Donato FKD, Stafoggia M, Viegi G, Michelozzi P, Carla A et al. Nationwide epidemiological study for estimating the effect of extreme outdoor temperature on occupational injuries in Italy. Environment International. 2019; 133. httpshttps://doi.org/10.1016/j.envint.2019.105176
50. Li XD, Chow KH, Zhu YM, Lin Y. Evaluating the impacts of high-temperature outdoor working environments on construction labor productivity in China: a case study of rebar workers. Build Environ. 2016;95:42–52. https://doi.org/10.1016/j.buildenv.2015.09.005
51. Liu J, Varghese BM, Hansen A, Xiang J, Zhang Y, Dear K, Gourley M, Driscoll T, Morgan G, Capon A, et al. Is there an association between hot weather and poor mental health outcomes? A systematic review and meta-analysis. Environ Int. 2021;153:106533. https://doi.org/10.1016/j.envint.2021.106533
52. Jingesi M, Lan S, Hu J, Dai M, Huang S, Chen S, Liu N, Lv Z, Ji J, Li X, et al. Association between thermal stress and cardiovascular mortality in the subtropics. Int J Biometeorol. 2023;67(12):2093–106. https://doi.org/10.1007/s00484-023-02565-6
53. Shin S, Choi JH, Lee KE, Yoon JH, Lee W. Risk and status of Gastrointestinal Cancer according to the International Standard Industrial classification in Korean workers. Cancers (Basel). 2022;14(20). https://doi.org/10.3390/cancers14205164
54. Si S, Lewkowski K, Fritschi L, Heyworth J, Liew D, Li IA. Productivity Burden of Occupational noise-Induced hearing loss in Australia: a life table modelling study. Int J Environ Res Public Health. 2020;17(13). https://doi.org/10.3390/ijerph17134667
55. Neri F, Laschi A, Foderi C, Fabiano F, Bertuzzi L, Marchi E. Determining noise and vibration exposure in Conifer Cross-cutting Operations by using Li-Ion batteries and Electric Chainsaws. Forests. 2018;9(8). https://doi.org/10.3390/f9080501
56. Mette J, Velasco Garrido M, Harth V, Preisser AM, Mache S. Healthy offshore workforce? A qualitative study on offshore wind employees' occupational strain, health, and coping. BMC Public Health. 2018;18(1):172. https://doi.org/10.1186/s12889-018-5079-4
57. Ebi KL, Capon A, Berry P, Broderick C, de Dear R, Havenith G, Honda Y, Kovats RS, Ma W, Malik A, et al. Hot weather and heat extremes: health risks. Lancet. 2021;398(10301):698–708.
58. Nguyen THY, Bertin M, Bodin J, Fouquet N, Bonvallot N, Roquelaure Y. Multiple exposures and coexposures to Occupational Hazards among Agricultural Workers: a systematic review of Observational studies. Saf Health Work. 2018;9(3):239–48. https://doi.org/10.1016/j.shaw.2018.04.002
59. Cattaneo I, Kalian AD, Di Nicola MR, Dujardin B, Levorato S, Mohimont L, Nathanail AV, Carnesseschi E, Astuto MC, Tarazona JV, et al. Risk Assessment of Combined exposure to multiple chemicals at the European Food Safety Authority: principles, Guidance documents, applications and Future challenges. Toxins. 2023;15(1). https://doi.org/10.3390/toxins15010040
60. McInnes JA, Akram M, MacFarlane EM, Keegel T, Sim MR, Smith P. Association between high ambient temperature and acute work-related injury: a case-crossover analysis using workers' compensation claims data. Scandinavian J Work Environ Health. 2017;43(1):86–94. https://doi.org/10.5271/sjweh.3602
61. Braun J, Baraliakos X, Bülow R, Schmidt CO, Richter A. Striking sex differences in magnetic resonance imaging findings in the sacroiliac joints in the population. Arthritis Res Ther. 2022;24(1):29. https://doi.org/10.1186/s13075-021-02712-7
62. He BJ, Zhao DX, Dong X, Xiong K, Feng C, Qi QL, Darko A, Sharifi A, Pathak M. Perception, physiological and psychological impacts, adaptive awareness and knowledge, and climate justice under urban heat: A study in extremely hot-humid Chongqing, China. Sustainable Cities and Society. 2022; 79. https://doi.org/10.1016/j.scs.2022.103685
63. Mathee A, Oba J, Rose A. Climate change impacts on working people (the HOTHAPS initiative): findings of the South African pilot study. Glob Health Action. 2010;3. https://doi.org/10.3402/gha.v3i0.5612
64. Chen Y, Zhang CK, Lu L, Zheng XH, Chang SQ. Dynamic of upper body sweat distribution in young males wearing fully encapsulated chemical protective ensembles. Sci Rep. 2022;12(1). https://doi.org/10.1038/s41598-022-04974-w
65. Smyth B, Maunder E, Meyler S, Hunter B, Muniz-Pumares D. Decoupling of Internal and External Workload during a Marathon: an analysis of durability in 82,303 recreational runners. Sports Med. 2022;52(9):2283–95. https://doi.org/10.1007/s40279-022-01680-5
66. Giersch GEW, Taylor KM, Caldwell AR, Charkoudian N. Body mass index, but not sex, influences exertional heat stroke risk in young healthy men and women. Am J Physiology-Regulatory Integr Comp Physiol. 2023;324(1):R15–9. https://doi.org/10.1152/ajpregu.00168.2022
67. Kim D, Kim HJ, Song TJ. Association of body composition indices with cardiovascular outcomes: a nationwide cohort study. Am J Clin Nutr. 2024;119(4):876–84. https://doi.org/10.1016/j.ajcnut.2024.02.015
68. Park SE, So WY, Kang YS, Yang JH. Relationship between perceived stress, obesity, and hypertension in Korean adults and older adults. Healthcare. 2023;11(16). https://doi.org/10.3390/healthcare11162271
69. Gervasoni E, Bertoni R, Anastasi D, Solaro C, Di Giovanni R, Grange E, Gunga HC, Rovaris M, Cattaneo D, Maggioni MA et al. Acute Thermoregulatory and Cardiovascular Response to Submaximal Exercise in People With Multiple Sclerosis. Frontiers in Immunology. 2022; 13. httpshttps://doi.org/10.3389/fimmu.2022.842269
70. Spech C, Paponetti M, Mansfield C, Schmitt L, Briggs M. Biomechanical variations in children who are overweight and obese during high-impact activities: a systematic review and meta-analysis. Obes Rev. 2022;23(6). https://doi.org/10.1111/obr.13431
71. Stephens D, Brearley M, Vermeulen L. Heat Health Management in a Quarantine and isolation facility in the tropics. Prehosp Disaster Med. 2022;37(2):259–64. https://doi.org/10.1017/s1049023x22000255
72. Tawatsupa B, Yiengprugsawan V, Kjellstrom T, Berecki-Gisolf J, Seubsman SA, Sleigh A. Association between Heat Stress and Occupational Injury among Thai workers: findings of the Thai Cohort Study. Ind Health. 2013;51(1):34–46. https://doi.org/10.2486/indhealth.2012-0138
73. Sui X, Wang Y, Jin M, Li K, Jiang G, Song A, He Z, Yin C, Zhao J, Wang L, et al. The effects of dexmedetomidine for patient-controlled analgesia on post-operative sleep quality and gastrointestinal motility function after surgery: a prospective, randomized, double-blind, and controlled trial. Front Pharmacol. 2022;13:990358. https://doi.org/10.3389/fphar.2022.990358
74. Elhadi M, Alsoufi A, Msherghi A, Alshareea E, Ashini A, Nagib T, Abuzid N, Abodabos S, Alrifai H, Gresea E, et al. Psychological Health, Sleep Quality, Behavior, and Internet Use among people during the COVID-19 pandemic: a cross-sectional study. Front Psychiatry. 2021;12:632496. https://doi.org/10.3389/fpsyt.2021.632496

75. Wee J, Tan XR, Gunther SH, Ihsan M, Leow MKS, Tan DSY, Eriksson JG, Lee JKW. Effects of medications on Heat loss capacity in Chronic Disease patients: Health implications amidst global warming. Pharmacol Rev. 2023;75(6):1140–66. https://doi.org/10.1124/pharmrev.122.000782

Publisher's note