

Preplanned Studies

Screening Carpal Tunnel Syndrome Risk in Occupational Populations — China, 2018–2023

Xinyao Hu¹; Xin Sun²; Zhongxu Wang²; Meibian Zhang²; Qing Xu²; Mimi Yang²;
Xinwei Guo²; Longlong Ma¹; Xingda Qu¹; Ning Jia^{2, #}

Summary

What is known about this topic?

Carpal tunnel syndrome is a common work-related disorder that is affected by personal, biomechanical, and psychosocial factors. Existing risk assessment methods often lack precision, hindering early intervention. As of August 1, 2025, occupational musculoskeletal disorders were recognized in China, emphasizing the need for effective preventive tools.

What has been added by this report?

This study developed machine learning models to identify carpal tunnel syndrome risk using a national dataset ($n=88,609$). The best model achieved 97.66% (3,130/3,131) accuracy, identifying key factors such as hand pain frequency and lifting load, and enhancing the early identification of high-risk workers.

What are the implications for public health practices?

These findings advocate for the use of data-driven screening tools in occupational health screening to enable early interventions. Employers should prioritize ergonomic improvements and symptom-reporting systems in alignment with China's proactive occupational disease policies.

gradient boosting demonstrated superior performance [recall: 99.97% (3,130/3,131); precision: 84.43% (3,130/3,707)] in identifying high-risk individuals. The SHapley Additive exPlanations (SHAP) analysis revealed that subjective hand pain characteristics were the primary factors, followed by psychosocial stressors, including fatigue and biomechanical exposure.

Conclusion: These findings support the transition toward data-driven public health practices, enabling targeted early prevention and ergonomic interventions to mitigate the burden of CTS in at-risk workforces.

Carpal tunnel syndrome (CTS) has emerged as a critical occupational health challenge owing to the increasing prevalence of repetitive strain injuries in diverse work environments (1). Globally, approximately 14.4% of the population is affected by CTS (1). In China, its prevalence among office workers is 9.6%, which is notably higher than that observed in many Western populations (2). On August 1, 2025, occupational musculoskeletal disorders were officially included in the updated "Classification and Catalog of Occupational Diseases" in China. This major policy milestone places CTS within the scope of statutory occupational disease management, creating an urgent imperative for enterprises and public health authorities to implement effective and legally compliant preventive and control measures.

The etiology of CTS is multifactorial and involves complex interactions among biomechanical exposure (including repetitive hand movements), personal demographics (including age and sex), and psychosocial stressors (3–5). Epidemiological studies have identified specific predictors. For instance, Lund et al. (6) quantified the impact of wrist velocity on the prevalence of CTS, whereas Lee et al. (7) highlighted demographic disparities and noted a high prevalence among middle-aged women. However, traditional

ABSTRACT

Introduction: Carpal tunnel syndrome (CTS) poses a substantial public health burden, necessitating precise early detection strategies that align with the prevention-focused goals of the evolving occupational health policy in China.

Methods: Machine learning algorithms were employed to integrate multidimensional risk factors for CTS screening using a large-scale dataset of 88,609 valid responses, including 10,438 reported CTS cases. The SHapley Additive exPlanations analysis identified the key factors, and ablation tests confirmed the importance of these features.

Results: Among the developed models, extreme

linear statistical methods often fail to capture the nonlinear and high-dimensional interactions among diverse risk factors (8). Consequently, the current screening approaches may lack the sensitivity required for early detection in large-scale occupational populations, representing a critical gap in meeting new prevention-oriented regulatory standards.

To address these limitations, this study applied advanced machine learning algorithms to synthesize multidimensional risk factors into a robust risk

screening framework. By enhancing screening accuracy beyond traditional methods, this study aimed to provide a data-driven foundation for early screening and targeted intervention, thereby supporting the implementation of new occupational health policies and reducing the burden of CTS in the workforce.

Figure 1 presents a schematic overview of the methodology of the study. This study used data from a large-scale national occupational health survey conducted in China between 2018 and 2023. The

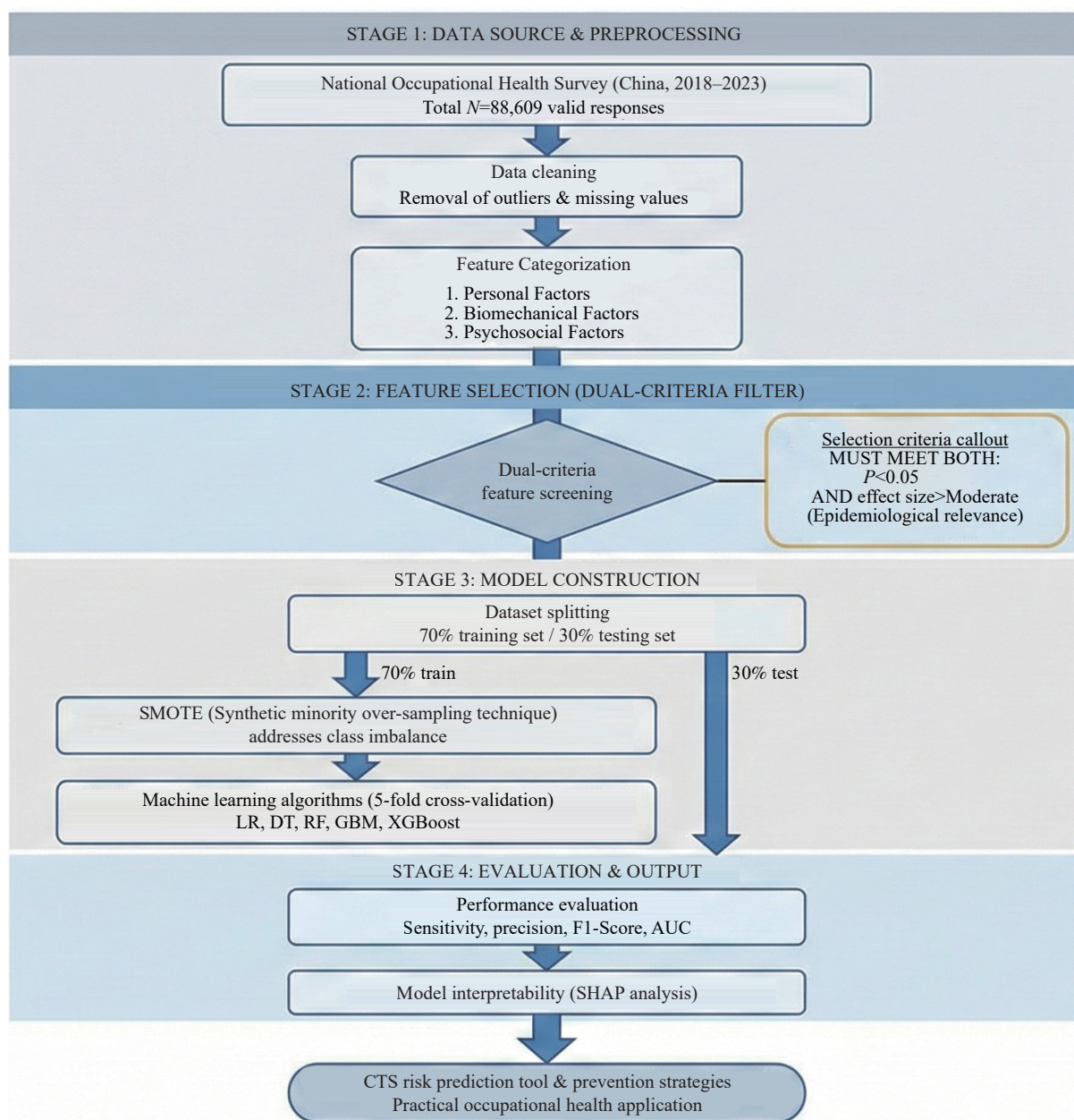


FIGURE 1. Schematic overview of the study methodology.

Abbreviation: LR=logistic regression; DT=decision tree; RF=random forest; GBM=gradient boosting machine; XGBoost=extreme gradient boosting; AUC=area under the ROC-curve; SHAP=SHapley Additive exPlanations.

dataset comprised 88,609 valid responses from workers across nine major industrial sectors and 29 distinct job types (9). A total of 10,438 participants reported CTS, corresponding to an overall prevalence of 11.78% in the study population. Data were collected using the electronic Chinese Musculoskeletal Disorder Questionnaire, integrated into the ergonomic assessment and WMSDs analysis system developed by the National Institute of Occupational Health and Poison Control, Chinese Center for Disease Control and Prevention (10). This study was approved by the Ethics Review Committee of the National Institute of Occupational Health and Poison Control, China. All participants provided informed consent prior to the study.

The primary outcome variable was self-reported CTS status (binary: CTS *vs.* non-CTS). The adoption of a standardized self-report-based CTS is fully aligned with the core objective of this study: to develop a scalable, high-efficiency risk-screening tool for large-scale occupational populations. To build a robust multidimensional screening model, potential risk factors were stratified into three domains: 1) personal factors (demographics, BMI, and exercise habits), 2) biomechanical factors (repetitive movements, force exertion, and posture), and 3) psychosocial factors (job demands and task load). Data preprocessing involved rigorous cleaning (removal of outliers and missing values) and encoding (standardization of continuous variables and ordinal or one-hot encoding of categorical variables). To ensure that the developed tool focused on practically meaningful risk factors rather than statistical noise, a dual-criteria feature-selection strategy was applied. Statistical tests (t-tests, chi-square tests, or nonparametric equivalents) were paired with effect size analyses. Only risk factors demonstrating both statistical significance ($P < 0.05$) and moderate to strong effect sizes ($\phi > 0.1$, $|r| > 0.25$, $\varepsilon^2 > 0.06$) were retained as input features for the machine learning models. Statistical analyses were performed using SPSS (version 19.0, IBM Corp., Armonk, NY, USA).

Five machine learning algorithms — binary logistic regression, decision tree, random forest, gradient boosting, and extreme gradient boosting (XGBoost) — were developed to screen CTS risk. These models were selected for their capacity to handle nonlinear interactions inherent in occupational exposure data. The dataset was partitioned into training [70% (62,026/88,609)] and testing [30% (26,583/88,609)] sets. To address typical class imbalance in occupational

disease data, the synthetic minority oversampling technique (SMOTE) was applied to the training set. The train-test split was performed at the beginning of the pipeline. All data-dependent procedures were performed exclusively within the training partitions. The held-out test set remained fully separate throughout all preprocessing and tuning steps to eliminate information leakage and ensure unbiased performance estimates.

SHapley Additive exPlanations (SHAP) values were used to identify the most relevant risk factors and to guide early intervention strategies. The model performance was evaluated using metrics prioritized for public health screening utility. While accuracy provides an overall baseline, emphasis is placed on sensitivity (recall) and the F1 score. High sensitivity is critical in occupational health surveillance to minimize false negatives (missed screenings) and ensure that high-risk workers receive timely ergonomic interventions. Precision was monitored to reduce false positives and unnecessary resource utilization. The area under the receiver operating characteristic curve was calculated to assess the discriminative power of the models across all classification thresholds.

SHAP analysis identified a clear hierarchy of risk factors and actionable targets for occupational CTS prevention (Table 1). Three risk tiers were proposed based on SHAP predictive weight, modifiability, and functional role in occupational CTS prevention. Tier 1: Core symptoms comprised the top-ranked hand pain and symptom variables with dominant explanatory power, serving as high-yield primary screening signals for identifying at-risk workers. Tier 2: Modifiable factors included psychosocial fatigue and biomechanical workload factors of moderate importance, representing the only intervention-actionable targets for workplace ergonomic optimization and fatigue management. Tier 3: Background covers stable personal and occupational factors, such as job tenure and general health, with small but consistent screening contributions used for baseline risk stratification rather than direct intervention. Subjective hand pain characteristics emerged as the strongest screening signals, with pain frequency and duration ranking as the top two contributors to model accuracy. Among the modifiable occupational risk factors, fatigue perception (psychosocial domain) and cumulative lifting load (biomechanical domain) demonstrated the highest importance, followed by repetitive hand movements and non-neutral wrist postures. Personal demographic

TABLE 1. Ranking of the top ten factors for carpal tunnel syndrome based on mean SHAP values and proposed risk tiers.

Rank	Factors	Category	Mean SHAP	Proposed risk tier
1	Frequency of hand pain/discomfort	Personal	2.029	Tier 1: Core Symptom
2	Hand pain/discomfort in past 12 months	Personal	1.356	Tier 1: Core Symptom
3	Tenure (years) in current job	Personal	0.053	Tier 3: Background
4	Physical health status	Personal	0.047	Tier 3: Background
5	Physical exercise	Personal	0.022	Tier 3: Background
6	Hand pain/discomfort in past 7 days	Personal	0.016	Tier 1: Core Symptom
7	Feel fatigue	Psychosocial	0.013	Tier 2: Modifiable
8	Hours until feeling fatigued	Psychosocial	0.013	Tier 2: Modifiable
9	Cumulative lifting load	Biomechanical	0.009	Tier 2: Modifiable
10	Duration of hand symptoms (Annual)	Personal	0.009	Tier 1: Core Symptom

Abbreviation: SHAP=SHapley Additive exPlanations.

factors, including age and sex, showed moderate but consistent contributions. This factor ranking directly informs workplace prevention strategies that should prioritize systematic symptom monitoring, fatigue management, and physical load reduction as core intervention components.

Statistical screening initially identified 26 significant factors spanning the personal, biomechanical, and psychosocial domains (Supplementary Table S1, available at <https://weekly.chinacdc.cn/>). Ablation testing further confirmed the fundamentally multidimensional nature of CTS risk. Removing any single factor domain, including psychosocial factors, resulted in a notable degradation in the model screening performance (Figure 2). This finding validates the necessity of integrated multidomain occupational health surveillance rather than single-factor risk monitoring.

Among the five machine learning algorithms

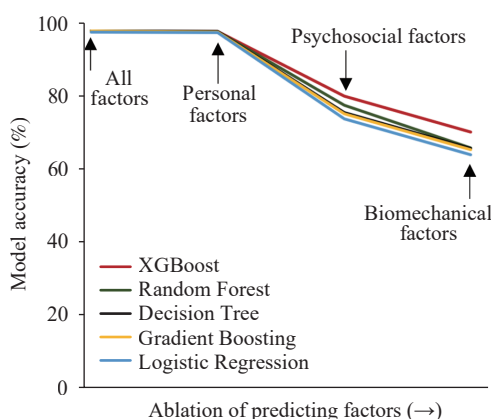


FIGURE 2. Model accuracy decreased with the ablation of the screening factors.

Abbreviation: XGBoost=extreme Gradient Boosting.

evaluated, XGBoost emerged as the optimal model for large-scale occupational CTS screening (Supplementary Table S2, available at <https://weekly.chinacdc.cn/>). It achieved near-perfect sensitivity [99.97% (3,130/3,131)], which is critical for minimizing false negatives in public health surveillance, while maintaining high precision [84.43% (3,130/3,707)] and robust discriminative power, with an area under the receiver operating characteristic curve of 0.992 (Supplementary Figure S1, available at <https://weekly.chinacdc.cn/>). Although the decision tree also achieved a high recall, XGBoost delivered the most balanced performance between case detection capability and screening reliability for population-level screening applications.

DISCUSSION

This study aimed to develop and validate a generalizable high-performance machine learning screening model for CTS risk screening to support large-scale universal occupational health surveillance. The analytical framework of this study is strictly aligned with this objective. By synthesizing multidimensional risk factors using machine learning, this study established a high-precision risk-screening framework for CTS in occupational populations. The XGBoost model demonstrated superior risk-screening performance, particularly in terms of sensitivity (99.97%), ensuring minimal false-negative results during mass screening. SHAP analysis further identified subjective hand pain characteristics, fatigue perception, and biomechanical load as primary risk factors, validating the multifactorial etiology of CTS.

These findings have profound public health

implications, particularly following the inclusion of musculoskeletal disorders in China's "Classification and Catalog of Occupational Diseases." First, the dominance of hand pain symptoms provided strong evidence for the establishment of symptom-based active sentinel monitoring. This approach offers a low-cost, direct pathway for early detection that is superior to passive reliance on clinical screening. Second, identifying fatigue as a critical factor bridges the gap between psychosocial stress and physical damage. This finding underscores the fact that effective CTS control requires the management of organizational factors, such as work rhythm, directly aligning with the Healthy China 2030 initiative's requirements for occupational stress management. Ultimately, this model serves as a quantifiable risk-grading tool, facilitating a paradigm shift in occupational health management from post-diagnosis compensation to pre-diagnosis prevention.

This study was limited by its cross-sectional design. The core screening variables of the established SHAP-based screening model were predominantly hand pain- and discomfort-related indicators extracted from the cross-sectional survey data, including the subjective pain frequency and duration. These symptomatic variables substantially overlapped with the symptomatic definition of self-reported CTS, which may partially account for the excellent screening performance of our model. This overlap indicates that the model primarily achieves symptom-based CTS identification rather than a definitive diagnosis. Furthermore, this study prioritized discrimination metrics for occupational health risk stratification and did not perform a formal calibration analysis. Future research should include calibration plots and Brier score evaluations to validate the reliability of the predicted risks and better support real-world decision-making.

From a practical application perspective, the proposed model was developed for the preliminary population-level screening of suspected carpal tunnel syndrome (CTS) in occupational health practice. It aims to support the early identification of at-risk workers and guide targeted follow-up arrangements and is not intended to serve as a standalone diagnostic tool. The definitive diagnosis of CTS still relies on formal specialist clinical assessment combined with objective electrophysiological examinations, such as nerve conduction studies, which remain the diagnostic gold standard.

The developed machine learning models can also

support targeted employer interventions. For example, employers should establish nonpunitive, rapid symptom-reporting mechanisms (such as "one-click reporting" via mobile applications). Reports of frequent hand pain should serve as immediate triggers for on-site ergonomic investigations. Furthermore, to address the critical risks associated with fatigue and insufficient rest, enterprises should implement structured microbreak policies and optimize work pacing to prevent continuous high-intensity exposure. For the risks associated with cumulative lifting loads, priority should be given to engineering controls, such as introducing assistive handling devices and optimizing the workstation height to minimize non-neutral wrist postures.

These findings provide a scientific basis for refining screening and prevention standards for occupational CTS and guide the allocation of work injury prevention funds. Future initiatives should focus on the longitudinal validation of these models across diverse industries to establish causal relationships. Policymakers can use these data-driven insights to formulate ergonomic exposure limits that are suitable for the Chinese industrial context.

Conflicts of interest: No conflicts of interest.

Ethical statement: Approved by the Ethics Review Committee of the National Institute of Occupational Health and Poison Control, China. All participants provided informed consent prior to their inclusion in the study.

Acknowledgements: All the participants involved in this study, from Chongqing, Shanghai, Jiangsu, Zhejiang, Tianjin, Beijing, Hubei, Ningxia Hui Autonomous Region, Sichuan, Shaanxi and Yunnan Provincial Centers for Disease Prevention and Control, Hubei Provincial Hospital of Integrated Chinese and Western Medicine, Guangzhou Twelfth People's Hospital Affiliated to Guangzhou Medical University, Liaoning Provincial Health Supervision Center, Shenyang, Liaoning, China, Guizhou Province Occupational Disease Prevention and Control Hospital, Shandong Academy of Occupational Health and Occupational Medicine, Guangdong Province Hospital for Occupational Disease Prevention and Treatment, Civil Aviation Medical Center of China Civil Aviation Administration, Tianjin Occupational Disease Prevention and Control Hospital, Fujian Province Occupational Disease and Chemical Poisoning Prevention and Control Center, and Institute of Occupational Medicine of Jiangxi.

Funding: Supported by the National Key R&D

Program of China (Project Number: 2022YFC2503205), the Project of Occupational Health Risk Assessment and the National Occupational Health Standard Formulation of the National Institute of Occupational Health and Poison Control (Project Number: 102393220020090000020), Capital's Funds for Health Improvement and Research (Capital Health Development 2024-1G-3101), and Key R&D Program of the Xinjiang Uygur Autonomous Region (2024B03040).

doi: 10.46234/ccdcw2026.163

* Corresponding author: Ning Jia, jianing@niohp.chinacdc.cn.

¹ Institute of Human Factors and Ergonomics, College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen City, Guangdong Province, China; ² National Institute of Occupational Health and Poison Control, Chinese Center for Disease Control and Prevention & Chinese Academy of Preventive Medicine, Beijing, China.

Copyright © 2026 by Chinese Center for Disease Control and Prevention & Chinese Academy of Preventive Medicine. All content is distributed under a Creative Commons Attribution Non Commercial License 4.0 (CC BY-NC).

Submitted: March 24, 2026

Accepted: July 31, 2026

Issued: August 07, 2026

REFERENCES

1. Padua L, Cuccagna C, Giovannini S, Coraci D, Pelosi L, Loreti C, et al. Carpal tunnel syndrome: updated evidence and new questions. *Lancet Neurol* 2023;22(3):255 – 67. [https://doi.org/10.1016/S1474-4422\(22\)00432-X](https://doi.org/10.1016/S1474-4422(22)00432-X).
2. Gebrye T, Jeans E, Yeowell G, Mbada C, Fatoye F. Global and regional prevalence of carpal tunnel syndrome: a meta-analysis based on a systematic review. *Musculoskeletal Care* 2024;22(4):e70024. <https://doi.org/10.1002/msc.70024>.
3. Feng BB, Chen KD, Zhu XX, Ip WY, Andersen LL, Page P, et al. Prevalence and risk factors of self-reported wrist and hand symptoms and clinically confirmed carpal tunnel syndrome among office workers in China: a cross-sectional study. *BMC Public Health* 2021;21(1):57. <https://doi.org/10.1186/s12889-020-10137-1>.
4. Trillos-Chacón MC, Castillo-M JA, Tolosa-Guzman I, Medina AFS, Ballesteros SM. Strategies for the prevention of carpal tunnel syndrome in the workplace: a systematic review. *Appl Ergon* 2021;93:103353. <https://doi.org/10.1016/j.apergo.2020.103353>.
5. Nowak W, Znamirska P, Szmigielska N, Zemsta K, Miśkiewicz J, Plata H, et al. Risk factors for carpal tunnel syndrome. *J Pre-Clin Clin Res* 2023;17(3):167 – 70. <https://doi.org/10.26444/jpcr/168559>.
6. Lund CB, Mikkelsen S, Thygesen LC, Hansson GÅ, Thomsen JF. Movements of the wrist and the risk of carpal tunnel syndrome: a nationwide cohort study using objective exposure measurements. *Occup Environ Med* 2019;76(8):519 – 26. <https://doi.org/10.1136/oemed-2018-105619>.
7. Lee IH, Kim YK, Kang DM, Kim SY, Kim IA, Kim EM. Distribution of age, gender, and occupation among individuals with carpal tunnel syndrome based on the National Health Insurance data and National Employment Insurance data. *Ann Occup Environ Med* 2019;31:e31. <https://doi.org/10.35371/aoem.2019.31.e31>.
8. van Rijn RM, Huisstede BMA, Koes BW, Burdorf A. Associations between work-related factors and the carpal tunnel syndrome—a systematic review. *Scand J Work Environ Health* 2009;35(1):19 – 36. <https://doi.org/10.5271/sjweh.1306>.
9. Jia N, Wang ZX, Zhang MB, Zhang HD, Ling RJ, Wang Z, et al. Prevalence and risk factors of lower extremity musculoskeletal disorders among occupational groups in key industries—China, 2018–2023. *China CDC Wkly* 2024;6(52):1388 – 95. <https://doi.org/10.46234/ccdcw2024.276>.
10. Jia N, Zhang MB, Zhang HD, Ling RJ, Liu YM, Li G, et al. Prevalence and risk factors analysis for low back pain among occupational groups in key industries of China. *BMC Public Health* 2022;22(1):1493. <https://doi.org/10.1186/s12889-022-13730-8>.

SUPPLEMENTARY MATERIALS

SUPPLEMENTARY TABLE S1. Predictors selected for machine learning model development: results of statistical screening ($n=88,609$).

Category	Risk factors	Data type	P	Effect size	Selected
Personal Factors	Gender	Binary	0.001	0.011	No
	Age	Continuous	<0.001	0.091	No
	Height	Continuous	0.136	0.028	No
	Weight	Continuous	0.003	0.016	No
	BMI	Continuous	0.020	0.003	No
	Physical exercise	Ordinal	<0.001	0.121	Yes
	Smoking history	Ordinal	<0.001	0.002	No
	Number of cigarettes smoked	Continuous	0.042	0.033	No
	Years since quitting smoking	Continuous	0.631	0.010	No
	Physical health status	Ordinal	<0.001	0.224	Yes
	Whether having traumatic disease/injury	Binary	<0.001	0.028	No
	Previous disease history	Binary	<0.010	0.004	No
	Tenure (years) in current job	Continuous	0.034	0.261	Yes
	Total work tenure (years)	Continuous	<0.001	0.140	No
	Hand pain/discomfort in past 7 days	Binary	0.011	0.390	Yes
	Hand pain/discomfort in past 12 months	Binary	<0.001	0.701	Yes
	Frequency of hand pain/discomfort	Ordinal	<0.001	0.814	Yes
	Duration of hand symptoms (Annual)	Ordinal	<0.001	0.509	Yes
	Hand pain score	Ordinal	<0.001	0.452	Yes
Biomechanical Factors	Lifting heavy objects (>5 kg each time)	Ordinal	<0.001	0.012	No
	Lifting very heavy objects (>20 kg each time)	Ordinal	<0.001	0.009	No
	Cumulative lifting load	Continuous	<0.001	0.290	Yes
	Work requiring upper limb/hand force	Ordinal	<0.001	0.031	No
	Using vibrating tools at work	Ordinal	<0.001	0.144	Yes
	Driving vehicles	Ordinal	0.142	0.081	No
	Working in uncomfortable postures	Ordinal	<0.001	0.212	Yes
	Work requires multiple repetitive motions per minute	Binary	<0.001	0.141	Yes
	Wrists frequently bend upward/downward	Binary	<0.001	0.163	Yes
	Wrists need to be bent for prolonged periods	Binary	<0.001	0.191	Yes
	Wrists frequently rest on hard, angled edges	Binary	<0.001	0.138	Yes
	Work requires pinching/gripping items/tools	Binary	<0.001	0.130	Yes
	Hand position during work	Binary	<0.001	0.025	No

Continued

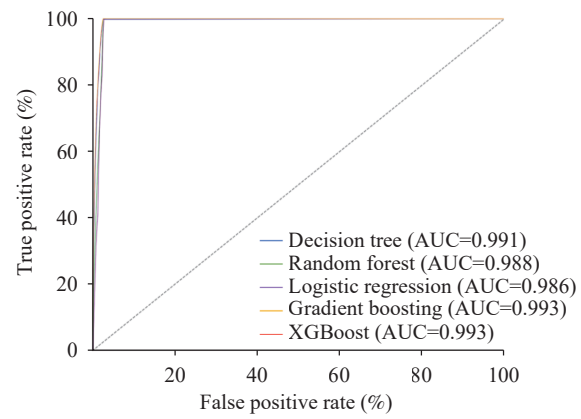
Category	Risk factors	Data type	P	Effect size	Selected
Psychosocial Job Factors	Doing almost the same work every day	Binary	0.010	0.071	No
	Work requirement changes daily	Binary	0.023	0.009	No
	Work rotation between you and colleagues	Binary	0.063	0.021	No
	Feel fatigue	Binary	0.011	0.202	Yes
	Levels of fatigue perception	Ordinal	<0.001	0.181	Yes
	Hours until feeling fatigued	Continuous	<0.001	0.381	Yes
	Number of work days per week	Continuous	<0.001	0.012	No
	Number of work hours per day	Continuous	<0.001	0.133	Yes
	Work shift conditions	Ordinal	<0.001	0.044	No
	Frequently working overtime	Binary	<0.001	0.083	No
	Number of work breaks per shift (times)	Continuous	0.271	0.112	Yes
	Total break time per shift (minutes)	Continuous	0.111	0.201	Yes
	Sufficient break time	Binary	<0.001	0.151	Yes
	Returning to work after breaks	Binary	<0.001	0.020	No
	Can decide when to start and end work	Binary	<0.001	0.033	No
	Can decide when to take breaks	Binary	<0.001	0.049	No
	Work department is understaffed	Binary	<0.001	0.110	No
	Frequently covering shifts for colleagues	Binary	<0.001	0.691	Yes
	Engaged in other jobs	Binary	<0.001	0.016	No
	Experiencing elbow fatigue	Binary	0.011	0.131	No
	Number of elbow fatigue per month	Continuous	<0.001	0.370	No
	Experiencing hand fatigue	Binary	<0.001	0.348	No
	Number of hand fatigue per month	Continuous	<0.001	0.743	No

Abbreviation: BMI=body mass index.

SUPPLEMENTARY TABLE S2. Ranking of the top ten predictors for carpal tunnel syndrome based on mean SHAP values and proposed risk tiers.

Rank	Predictor	Category	Mean SHAP	Proposed Risk Tier
1	Frequency of hand pain/discomfort	Personal	2.029	Tier 1: Core Symptom
2	Hand pain/discomfort in past 12 months	Personal	1.356	Tier 1: Core Symptom
3	Tenure (years) in current job	Personal	0.053	Tier 3: Background
4	Physical health status	Personal	0.047	Tier 3: Background
5	Physical exercise	Personal	0.022	Tier 3: Background
6	Hand pain/discomfort in past 7 days	Personal	0.016	Tier 1: Core Symptom
7	Feel fatigue	Psychosocial	0.013	Tier 2: Modifiable
8	Hours until feeling fatigued	Psychosocial	0.013	Tier 2: Modifiable
9	Cumulative lifting load	Biomechanical	0.009	Tier 2: Modifiable
10	Duration of hand symptoms (Annual)	Personal	0.009	Tier 1: Core Symptom

Abbreviation: SHAP=SHapley Additive exPlanations.



SUPPLEMENTARY FIGURE S1. ROC curves and AUC of the machine learning models.

Abbreviation: ROC=receiver operating characteristic; AUC=area under the ROC-curve; XGBoost=extreme Gradient Boosting.