BMJ Open Construction of a risk prediction model for occupational noise-induced hearing loss using routine blood and biochemical indicators in Shenzhen, China: a predictive modelling study

Caiping Li ⁽¹⁾, ¹ Liuwei Shi ⁽¹⁾, ^{2,3} Linlin Chen ⁽¹⁾, ¹ Dafeng Lin ⁽¹⁾, ² Xiangli Yang ⁽¹⁾, ² Peimao Li ⁽¹⁾, ² Wen Zhang ⁽¹⁾, ² Wenting Feng ⁽¹⁾, ² Yan Guo ⁽¹⁾, ² Liang Zhou ⁽¹⁾, ¹ Naixing Zhang ⁽¹⁾, ² Dianpeng Wang ⁽¹⁾, ²

ABSTRACT

To cite: Li C. Shi L. Chen L. et al. Construction of a risk prediction model for occupational noise-induced hearing loss using routine blood and biochemical indicators in Shenzhen, China: a predictive modelling study. BMJ Open 2025;15:e097249. doi:10.1136/ bmjopen-2024-097249

Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (https://doi.org/10.1136/ bmjopen-2024-097249).

CL and LS are joint first authors.

Received 28 November 2024 Accepted 08 April 2025

Check for updates

C Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

¹School of Public Health, Southern Medical University, Guangzhou, Guangdong, China ²Shenzhen Prevention and Treatment Center for Occupational Diseases, Shenzhen, Guangdong, China ³Jilin University, Changchun, Jilin, China

Correspondence to

Dr Dianpeng Wang; szpcr@126.com and Dr Naixing Zhang; zhanghealth@126.com

Objectives Occupational noise-induced hearing loss (ONIHL) represents a prevalent occupational health condition, traditionally necessitating multiple puretone audiometry assessments. We have developed and validated a machine learning model leveraging routine haematological and biochemical parameters, thereby offering novel insights into the risk prediction of ONIHL. Design, setting and participants This study analysed data from 3297 noise-exposed workers in Shenzhen, including 160 ONIHL cases, with the data set divided into D1 (2868 samples, 107 ONIHL cases) and D2 (429 samples, 53 ONIHL cases). The inclusion criteria were formulated based on the GBZ49-2014 Diagnosis of Occupational Noise-Induced Hearing Loss. Model training was performed using D1, and model validation was conducted using D2. Routine blood and biochemical indicators were extracted from the case data, and a range of machine learning algorithms including extreme gradient boosting (XGBoost) were employed to construct predictive models. The model underwent refinement to identify the most representative variables, and decision curve analysis was conducted to evaluate the net benefit of the model across various threshold levels.

Primary outcome measures Model creation data set and validation data sets: ONIHL.

Results The prediction model, developed using XGBoost, demonstrated exceptional performance, achieving an area under the receiver operating characteristic curve (AUC) of 0.942, a sensitivity of 0.875 and a specificity of 0.936 on the validation data set. On the test data set, the model achieved an AUC of 0.990. After implementing feature selection, the model was refined to include only 16 features, while maintaining strong performance on a newly acquired independent data set, with an AUC of 0.872, a balanced accuracy of 0.798, a sensitivity of 0.755 and a specificity of 0.840. The analysis of feature importance revealed that serum albumin (ALB), platelet distribution width (PDW), coefficient of variation in red cell distribution width (RDW-CV), serum creatinine (Scr) and lymphocyte percentage (LYMPHP) are critical factors for risk stratification in patients with ONIHL.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- \Rightarrow The model predicts occupational noise-induced hearing loss (ONIHL) using routine blood and biochemical indicators, eliminating the need for audiometric tests or direct noise exposure data.
- \Rightarrow It simplifies the diagnostic process, reducing time. costs and manpower requirements.
- \Rightarrow It provides an accessible and efficient alternative for early screening and prevention of ONIHL.
- \Rightarrow The study is limited to the Shenzhen population, and the model's generalisability to other groups and settings remains uncertain.
- \Rightarrow The positive-to-negative sample ratio exceeds

tings remains uncertain. ⇒ The positive-to-negative sample ratio exceeds 1:20, mirroring real-world conditions but limiting predictive accuracy; future integration of additional biomarkers, such as DNA methylation, may improve performance. Conclusion The analysis of feature importance identified ALB, PDW, RDW-CV, Scr and LYMPHP as pivotal factors for risk stratification in patients with ONIHL. The machine learning model, using XGBoost, effectively distinguishes patients with ONIHLamong individuals exposed to noise, thereby facilitating early diagnosis and intervention. INTRODUCTION Occupational noise-induced hearing loss (ONIHL) is characterised as a progressive sensorineural hearing impairment predomi-nantly attributed to damage of the hair cells

nantly attributed to damage of the hair cells **3** within the inner ear, consequent to prolonged exposure to high-intensity noise environments.¹ As reported by WHO, approximately 10% of the global workforce is impacted by elevated noise levels, with occupational noise exposure accounting for 7-21% of hearing loss among workers.² A national occupational research agenda says that ONIHL has the highest prevalence of occupational diseases

and

Protected by copyright, including for uses related to text

in the USA.³ About 22 million US workers are currently exposed to hazardous occupational noise.⁴ This incidence is notably higher in developing countries.⁵ As the largest developing nation, China has witnessed an increasing trend in the incidence of occupational ONIHL in recent years. The prevalence of ONIHL has been reported to be over 20% among noise-exposed workers in China.⁶ Such hearing loss can result in communication challenges, social isolation, loneliness and depression, thereby adversely impacting patients' quality of life and leading to indirect economic losses for society.⁷ However, despite being a major global public health issue, early screening methods for ONIHL remain limited.

Currently, pure-tone audiometry (PTA) is regarded as the gold standard for diagnosing ONIHL.⁸ However, its reliance on costly audiological equipment and the necessity for highly trained professionals restrict its practicality for large-scale ONIHL screening among noiseexposed occupational groups.⁹ Additionally, PTA relies on subjective auditory feedback and may be influenced by individual auditory adaptation. Consequently, there is a pressing need to develop a practical and user-friendly screening tool specifically designed for patients with ONIHL to prevent the advancement to clinically significant ONIHL. Numerous instances of ONIHL are characterised by an initial deterioration in high-frequency hearing, which gradually progresses to impairments in low-frequency or speech frequency hearing.¹⁰ The early identification of individuals at high risk is essential for effective prevention and intervention strategies.

Consequently, the development of predictive models to screen high-risk populations for further evaluation represents a viable alternative approach. The growing volume of data has facilitated the application of machine learning (ML) techniques in the context of ONIHL. At present, a variety of methodologies employing either traditional statistical analysis or ML techniques are used to predict the risk of ONIHL. These methodologies frequently necessitate substantial human resources and present challenges in manual definition.¹¹ The integration of ML within the field of audiology has demonstrated potential, particularly in its capacity to effectively analyse non-linear relationships within data, such as forecasting hearing thresholds for individuals exposed to specific risk factors.¹² Abdollahi *et al*¹³ constructed eight ML models to forecast sensorineural hearing loss following radiotherapy and chemotherapy, with five of these models demonstrating accuracy and precision exceeding 70%. Comparable levels of accuracy have been reported in other investigations employing ML models to predict sudden sensorineural hearing loss (SSNHL) and ototoxic hearing loss.^{14 15} Additionally, various studies have documented accuracy rates between 0.64 and 0.99 when using diverse ML algorithms and input parameters to predict ONIHL risk factors.^{16–20} Among the diverse array of ML techniques, support vector machines (SVM) models, random forest (RF) models and extreme gradient boosting (XGBoost) models have demonstrated superior

<page-header><text><text><text><text><text>

group and the noise-exposed normal hearing group. After preprocessing, a total of 3297 samples were retained, with D1 and D2 consisting of 2868 and 429 samples, respectively. Among them, there were 107 and 53 cases of noiseinduced hearing loss, representing the positive samples. We then applied random sampling to split D1 into a training set and a test set at a 7:3 ratio. D2 was used as an independent test set.

All data sets included the following variables: sex, age, total protein (TP), albumin (ALB), glucose (GLU), cholesterol (CHO), triglycerides (TG), HDL, LDL, total bilirubin (TBIL), direct bilirubin (DBIL), indirect bilirubin (IBIL), alanine aminotransferase (ALT), aspartate aminotransferase (AST), blood urea nitrogen (BUN), serum creatinine (Scr), uric acid (UA), globulin (GLB), haemoglobin (Hb), red blood cell count (RBC), haematocrit (HCT), mean corpuscular volume (MCV), mean corpuscular haemoglobin (MCH), mean corpuscular haemoglobin concentration (MCHC), WBC, eosinophil count (EOC), basophil count (BAC), lymphocyte count (LYMPHC), monocyte count (MOC), platelet count (PLT), neutrophil count (GRANC), eosinophil percentage (EOP), basophil percentage (BAP), coefficient of variation in red cell distribution width (RDW-CV), mean platelet volume (MPV), platelet distribution width (PDW), plateletcrit (PCT), neutrophil percentage (GRANP), lymphocyte percentage (LYMPHP), monocyte percentage (MOP), SD in red cell distribution width (RDW-SD), platelet-to-HDL ratio (PLT/HDL), glucoseto-HDL ratio (GLU/HDL), platelet-to-lymphocyte ratio (PLT/LYMPHC), albumin-to-globulin ratio (A/G), neutrophil-to-lymphocyte ratio (S/L), triglycerideglucose index (TyG) and estimated glomerular filtration rate (eGFR) (the calculation formulas for TyG and eGFR are detailed in online supplemental additional 1).

In light of the pronounced class imbalance present across all data sets, we employed oversampling of the positive instances within the training set using the 'ovun. sample()' function from the ROSE package. This function randomly replicates samples from the minority class, thereby equalising the number of positive and negative samples in the training set and achieving a balanced class distribution.²⁸ This approach effectively increases the sample size of the minority class, mitigating the effects of class imbalance during model training. All data sets underwent Z-score normalisation, using the mean and SD derived from the training set data.

Framework

Employing occupational health examination data, we introduce an integrated framework for the identification of patients with noise-induced hearing loss, as illustrated in figure 1. Initially, we preprocessed two data sets, desig-9 nated as D1 and D2. Data set D1 was partitioned into training and validation subsets in a 7:3 ratio, while data 8 set D2 served as an independent test set for the evaluation of the final model. Due to the class imbalance present in the data set, we employed an oversampling technique on the training set. Subsequently, we used a comprehensive array of ML algorithms, including XGBoost, logistic regression (LR), RF, SVM and k-nearest neighbour (KNN), to construct predictive models. We then applied feature selection methods to the most optimal predictor uses rela among the five to enhance the tool's feasibility. The performance of the refined model was evaluated using an independent test set. We conducted a feature importance analysis to identify variables correlated with the incidence of noise-induced hearing loss. Additionally, we optimised ç the model to select the most representative variables and text and employed decision curve analysis (DCA) to evaluate the net benefit of the model across various threshold levels.

Model construction

In order to construct predictive models, we employed five ML algorithms: LR, RF, SVM, KNN and XGBoost. LR is a form of linear regression that uses the Sigmoid function to convert outputs into probabilities for classification purposes.²⁹ RF comprises an ensemble of independently trained decision trees, with the ultimate prediction being derived through a voting mechanism Model evaluation (Feture selection (Validation set) (Pecision Curve Analysis)



Figure 1 A combined framework for identifying patients with occupational noise-induced hearing loss. AUC, area under the receiver operating characteristic curve; KNN, k-nearest neighbours; LR, logistic regression; mRMR, maximum relevance minimum redundancy; PCA, principal component analysis; PR-AUC, area under the precision–recall curve; RF, random forest; SVM, support vector machine; XGBoost, extreme gradient boosting.

data

among these trees, thereby mitigating the risk of overfitting.³⁰ SVM algorithm classifies samples by identifying an optimal hyperplane within the feature space, and it is capable of managing nonlinearly separable data.³¹ KNN algorithm, an instance-based learning method, classifies samples according to the proximity of their KNN, making it particularly suitable for small data sets and straightforward to implement.³² XGBoost is an ensemble method based on decision trees that enhances model performance through a gradient boosting framework. It constructs decision trees in an iterative manner to minimise model error, demonstrating particular efficacy in handling large-scale, high-dimensional data sets due to its robust generalisation capabilities and computational efficiency.³³ All models were developed in R (V.4.3.1) using a standardised fivefold cross-validation framework, with performance evaluated by the area under the receiver operating characteristic curve (AUC). Hyperparameter optimisation was performed via grid search to maximise validation AUC, supported by a heatmap illustrating key parameter interactions in XGBoost (online supplemental figure S1) and boxplots comparing crossvalidation stability across models (online supplemental figure S2). For XGBoost, critical parameters included tree depth (max_depth), learning rate (eta) and subsampling ratios, optimised to max_depth=7, eta=0.1 and subsampling ratios of 0.6. RF, SVM and KNN employed targeted tuning strategies-such as feature subset selection, regularisation balancing and dynamic neighbour selection-while LR used L2 regularisation. To ensure reproducibility, data splitting and randomisation were controlled by a global seed (set.seed(123)), with parallel processing (four threads) accelerating computations. We chose the model with the best performance on the validation set for further optimisation.

Model evaluation

To evaluate model performance, considering the class imbalance in the validation and test sets, we used the following metrics to comprehensively assess model performance: sensitivity, specificity, balanced accuracy, AUC, area under the precision-recall curve (PR-AUC), F1 score and precision. These metrics are defined as follows:

> Sensitivity = Recall = TPR = $\frac{TP}{TP+FN}$ Specificity = TNR = $\frac{TN}{TN+FP}$ Balanced Accuracy = $\frac{TPR+TNR}{2}$ $\begin{aligned} Precision &= \frac{TP}{TP+FP} \\ F1 \ score &= 2 \times \frac{Precision \times Recall}{Precision+Recall} \end{aligned}$

The performance of all models was assessed using the 'pROC' package in R to calculate AUC and PR-AUC values.

TP, that is, true positive, is the number of cases of noiseinduced hearing loss. FP, that is, false positive, denotes the number of normal subjects incorrectly predicted as having ONIHL. TN, that is, true negative, indicates the

BMJ Open: first published as 10.1136/bmjopen-2024-097249 on 28 April 2025. Downloaded from http://bmjopen.bmj.com/ on June 7, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES). 8 use úr (A Al training, and

number of healthy subjects correctly classified as normal. FN, that is, false negative, refers to the number of cases with ONIHL incorrectly classified as normal. All the above metrics range from 0 to 1.

Feature selection and feature importance analysis

Despite the relatively high performance of the prediction model using 48 features, there remains the possibility of redundant information or noise features that could adversely affect the decision-making process. To enhance the effective utilisation of features and streamline the model, we employed a combination of manual curation, principal component analysis (PCA) and maximum relevance minimum redundancy (mRMR) methods to extract essential features for the final model.³⁴ In the manual curation process, we initially identified features that exhibited significant differences between positive and negative samples. To improve the stability of the predictive model, we eliminated features that contributed to significant collinearity.³⁵ As a result, 16 features were retained. To ensure consistency, the number of feature subsets was also fixed at 16 during the application of PCA and mRMR analysis. PCA selected principal components based on cumulative explained variance, retaining those accounting for up to 80% of the variance retaining those accounting for up to 80% of the variance to balance dimensionality reduction and information preservation. Meanwhile, mRMR leveraged mutual information to maximise feature relevance while minimising 5 redundancy, ensuring an optimal feature subset. Furthermore, feature selection was conducted on the training set to mitigate the risk of overfitting. The analysis of feature importance facilitates the interpretation of the predictive model and aids in identifying the features most closely a associated with ONIHL. In this context, Feature importance was assessed using XGBoost's weight coefficients, with Gain (an internal statistical metric of XGBoost) highlighting features that maximally improve model performance while minimising redundancy.

Patient and public involvement

Patients and the public were not involved in the design, RESULTS We initially gathered occupational health examination **g**

data from the Shenzhen Occupational Disease Prevention and Control Institute for the period spanning 2023-2024, with subgroup D1 comprising 2868 noise-exposed workers. Of these, 107 participants were diagnosed with ONIHL. Table 1 provides a detailed description of the characteristics of both noise-exposed individuals and patients with ONIHL. The five most prominent features exhibiting significant differences between the ONIHL and non-ONIHL samples include ALB, TP, Age, RDW-CV and PDW (online supplemental figure S3).

Statistical characteristics of noise-exposed Table 1 hearing normal individuals and patients with ONIHL

Characteristics	Control	Case	P value
Ν	2761	107	
Sex			<0.001**
Female	25 (0.91%)	21 (19.6%)	
Male	2736 (99.1%)	86 (80.4%)	
Age, year	38.5±7.65	43.5±7.17	<0.001**
TP, g/L	72.7±3.91	68.1±4.69	<0.001**
ALB, g/L	46.7±2.59	42.7±2.95	<0.001**
GLU, mmol/L	5.23±0.63	5.42±0.86	0.023*
CHO, mmol/L	4.89±0.86	4.90±1.02	0.930
TG, mmol/L	1.64±1.21	1.97±1.28	0.011*
LDL, mmol/L	1.32±0.26	1.31±0.31	0.717
HDL, mmol/L	3.03±0.62	3.02±0.77	0.833
TBIL, µmol/L	16.5±6.26	15.7±6.02	0.188
DBIL, µmol/L	3.02±1.24	2.80±1.16	0.061
IBIL, µmol/L	13.4±5.20	12.6±4.90	0.098
ALT, U/L	29.5±22.1	27.2±21.8	0.294
AST, U/L	24.7±9.79	27.5±42.3	0.503
BUN, mmol/L	5.04±1.18	5.06±5.07	0.966
Scr, µmol/L	85.1±11.2	76.7±16.9	<0.001**
UA, µmol/L	401±79.2	482±893	0.351
GLB, g/L	26.0±3.36	25.4±3.37	0.086
Hb, g/L	154±10.1	144±17.9	<0.001**
RBC, ×10 ¹² /L	5.02±0.37	4.84±0.49	<0.001**
HCT, L/L	0.45±0.03	0.42±0.04	<0.001**
MCV, fL	89.6±4.77	87.6±8.47	0.015*
MCH, pg	30.8±1.86	29.9±3.47	0.007**
MCHC, g/L	344±6.82	341±12.3	0.013*
WBC, ×10 ⁹ /L	6.09±1.53	6.81±1.61	<0.001**
EOC, ×10 ⁹ /L	0.16±0.14	0.21±0.14	<0.001**
BAC, ×10 ⁹ /L	0.03±0.02	0.03±0.02	0.019*
LYMPHC, ×10 ⁹ /L	2.10±0.57	2.12±0.53	0.615
MOC, ×10 ⁹ /L	0.37±0.12	0.44±0.13	<0.001**
PLT, ×10 ⁹ /L	241±50.4	252±63.3	0.064
GRANC, ×10 ⁹ /L	3.44±1.15	4.00±1.15	<0.001**
EOP, %	2.59±1.91	3.07±1.72	0.005**
BAP, %	0.51±0.27	0.38±0.23	<0.001**
RDW-CV, %	12.9±0.65	13.9±1.54	<0.001**
MPV, fL	10.2±1.05	10.2±1.07	0.758
PDW, fL	16.2±0.35	14.9±1.97	<0.001**
PCT, %	0.24±0.04	0.26±0.06	0.041*
GRANP, %	55.8±7.74	58.4±6.22	<0.001**
LYMPHP, %	35.0±7.32	31.5±5.71	<0.001**
MOP, %	6.09±1.41	6.46±1.35	0.006**
RDW-SD, %	41.9±1.96	44.1±3.91	<0.001**

Continued

⊳

trair

<u>0</u>

Table 1 Continued

Characteristics	Control	Case	P value
PLT/HDL	189±55.0	202±65.5	0.046*
GLU/HDL	4.11±0.98	4.38±1.32	0.038*
PLT/LYMPHC	122±38.0	124±37.6	0.571
A/G	1.83±0.28	3.27±16.3	0.361
S/L	1.02±0.43	2.09±10.4	0.289
TyG	8.67±0.56	8.87±0.63	0.001**
eGFR, mL/ (min×1.73 m ²)	91.8±3.84	92.3±5.18	0.349

Data are presented as N (%) or mean±SE.

P values were based on χ^2 tests or t-test: *p<0.05, **p<0.01. A/G, albumin-to-globulin ratio; ALB, albumin; ALT, alanine aminotransferase; AST, aspartate aminotransferase; BAC, basophil count; BAP, basophil percentage; BUN, blood urea nitrogen; CHO, cholesterol; DBIL, direct bilirubin; eGFR, estimated glomerular filtration rate; EOC, eosinophil count; EOP, eosinophil percentage; GLB, globulin; GLU, glucose; GRANC, neutrophil count; GRANP, neutrophil percentage; Hb, haemoglobin; HCT, haematocrit; HDL, high-density lipoprotein; IBIL, indirect bilirubin; LDL, low-density lipoprotein; LYMPH, lymphocyte count; LYMPHP, lymphocyte percentage; MCH, mean corpuscular haemoglobin; MCHC, mean corpuscular haemoglobin concentration; MCV, mean corpuscular volume; MOC, monocyte count; MOP, monocyte percentage; MPV, mean platelet volume; ONIHL, occupational noise-induced hearing loss; PCT, plateletcrit; PDW, platelet distribution width; PLT, platelet count; RBC, red blood cell; RDW-CV, coefficient of variation in red cell distribution width; RDW-SD, SD in red cell distribution width; Scr, serum creatinine; S/L, neutrophil-to-lymphocyte ratio; TBIL, total bilirubin; TG, triglyceride; TP, total protein; TyG, triglycerideglucose index; UA, uric acid; WBC, white blood cell.

Performance comparison of the five machine learning methods

Protected by copyright, including for uses related to text and data mining Table 2 and online supplemental figure S4A, B show that the XGBoost algorithm has the highest AUC (0.942) and PR-AUC (0.791), as well as high recall (0.875) and balanced accuracy (0.905). The F1 score is only second to that of RF, making its overall performance excellent. The RF algorithm also performs well, with AUC (0.921) and PR-AUC (0.690), and a high balanced accuracy (0.872), indicating an outstanding overall performance. To maximise the identification of patients with ONIHL, the XGBoost algorithm was ultimately selected to further build the prediction model.

The results of the fivefold cross-validation on the training set show an AUC of 0.999, sensitivity of 0.995 and model demonstrates reliable performance on the test set set (AUC=0.900, PR-AUC=0.648), as shown in fi

Feature selection for the final model

Several pairs of features were observed to have high correlations, such as MCH and RDW-CV, and GRANP and LYMPHP, which may introduce redundant information and affect the model's decision-making and stability (online supplemental figure S5 for related heat maps). Therefore, we used manual curation, PCA and mRMR

Table 2 Performance comparison of the five machine learning methods										
	AUC	PR-AUC	Recall	Precision	Balanced accuracy	F1 score				
Logistic regression	0.923	0.683	0.938	0.200	0.896	0.330				
Random forest	0.921	0.690	0.781	0.446	0.872	0.568				
Support vector machine	0.797	0.369	0.750	0.140	0.786	0.235				
k-Nearest neighbours	0.829	0.521	0.688	0.333	0.817	0.449				
Extreme gradient boosting	0.942	0.791	0.875	0.346	0.905	0.496				

The values of AUC, PR-AUC, recall, precision, balanced accuracy, and F1 score range from 0 to 1, with higher values indicating better performance.

AUC, area under the receiver operating characteristic curve; PR-AUC, area under the precision-recall curve.

methods to identify the optimal features. As a result, 16 features were used to reconstruct the XGBoost model from each of manual curation, PCA and mRMR (online supplemental table S1). The PCA and mRMR feature selection methods identified seven shared features, with three of the top five selected features overlapping: ALB, RDW-CV and Scr. The model constructed using 16 features selected by mRMR and PCA showed a slight improvement on the validation set compared with the model built with all 48 features. Specifically, the PCA model achieved an AUC of 0.957 and a PR-AUC of 0.741, while the mRMR model had an AUC of 0.957 and a PR-AUC of 0.720. In contrast, the model based on manual feature selection exhibited a decline in performance, with an AUC of 0.919 and a PR-AUC of 0.540 (figure 3A,B). Similarly, the models constructed using mRMR and PCA demonstrated improvements in sensitivity and balanced accuracy on the validation set, with maximum increases of 29.2% and 3.7%, respectively, under the condition that the threshold of the full features model was set to 0.1. In comparison, the model based on manual curation showed suboptimal performance across all evaluation metrics (figure 3C).

Protected by copyright. We further evaluated the models using an independent test set (D2). In this test set, the manual curation model achieved an AUC of 0.830, the PCA model had an AUC of 0.837 and the mRMR model outperformed both with an AUC of 0.872 (figure 3D). The PR-AUC values were 0.524, 0.540 and 0.594 for the manual curation, PCA and mRMR models, respectively, with mRMR again demonрq strating the best performance (figure 3E). Regarding ğ sensitivity, specificity and balanced accuracy, the mRMR model exhibited the highest performance in the test See set evaluation (figure 3F). Notably, the lowest sensitivity observed for the mRMR model on D2 was 75.5%, while ttec all specificity scores remained above 78.0%. Overall, the model demonstrated strong performance on the inde-5 pendent test set, indicating that the selected core features text are sufficient for detecting noise-induced hearing loss and data mini among noise-exposed workers.

Feature importance ranking

To investigate which features contribute the most to the risk of ONIHL, we first used mRMR to select 16 important features and then built an XGBoost model based on



Figure 2 Performance of the prediction model on the validation set of data set D1 and the test set of data set D2. (A) receiver operating characteristic (ROC) curves; (B) precision-recall curves. AUC, area under the ROC curve; PR-AUC, area under the precision-recall curve.



Figure 3 Feature selection for the final model using principal component analysis (PCA), manual curation and maximum relevance minimum redundancy (mRMR). (A) Receiver operating characteristic (ROC) curves for models constructed with PCA-selected, manual curation-selected and mRMR-selected features on the validation set of data set D1. (B) Precision–recall curves of the above models. (C) Comparison of sensitivity, specificity and balanced accuracy on the validation set of data set between the model constructed before and after feature selection. (D) ROC curve of models using selected features on the test set of data set D2. (E) Precision–recall curve of the above models. (F) Comparison of other metrics on the independent test set D2. AUC, area under the POC curve; PR-AUC, area under the precision–recall curve.



Figure 4 Feature importance ranking for the model built using features selected by maximum relevance minimum redundancy.

these features. Subsequently, we ranked these features according to their weights in the XGBoost model, as shown in figure 4. The feature importance of the predictors based on PCA and manual curation is shown in online supplemental figures S6 and S7. The results indicated that the top five features, in order of importance, were ALB, PDW, RDW-CV, Scr and LYMPHP. Further comparisons between the ONIHL and normal samples revealed significant differences in ALB, TP, Age, RDW-CV and PDW. These findings are highly consistent with the top-ranked results in the XGBoost model, indicating a strong correlation between these indicators (ALB, PDW, RDW-CV, Scr and LYMPHP) and ONIHL.

Decision curve analysis

DCA (figure 5) revealed distinct net benefit patterns across threshold probabilities for the three models. The mRMR model demonstrated superior performance, achieving the highest net benefit over a broad threshold range (0.0–0.6), with particularly pronounced advantages at lower thresholds (0.0–0.2). In contrast, the PCA model exhibited competitive efficacy within the moderate threshold interval (0.2–0.4). Notably, both the manual curation model and the 'All/None' strategy underperformed: manual curation yielded consistently lower net benefits across all thresholds, and 'All/None' resulted



Figure 5 Decision curve analysis curves for models built using three different feature selection methods. mRMR, maximum relevance minimum redundancy; PCA, principal component analysis.

in negative net benefits at thresholds below 0.3, indicating clinical impracticality. These findings support a threshold-adaptive selection strategy: prioritising mRMR for thresholds≤0.4 (maximising robustness) and PCA for 0.2–0.4 thresholds (balancing accuracy and efficiency). This approach optimises clinical utility by aligning model strengths with context-specific decision risks.

DISCUSSION

ONIHL represents a significant global public health concern.^{2 36} Despite its complexity, ONIHL is a preventable condition.^{37 38} The Occupational Safety and Health Administration requires the implementation of hearing conservation programmes for workers exposed to noise levels of 85 decibels or higher, with the objective of safeguarding auditory health in noisy occupational environments.³⁹ Consequently, the development of a risk screening tool for ONIHL is crucial as a primary strategy for screening and prevention among workers exposed to occupational noise. In this study, we employed five ML algorithms using haematological test results to construct an ONIHL risk screening model. The models demonstrated AUC values exceeding 0.85, with accuracy and sensitivity surpassing 0.75 in both validation and independent test data sets. These results suggest that ML models are capable of accurately identifying patients with ONIHL within the population of noise-exposed workers.

In an evaluation of model performance on the validation set, the XGBoost model exhibited superior efficacy compared with all other algorithms assessed, achieving an AUC of 0.942 and a PR-AUC of 0.791. The specificity and balanced accuracy metrics for the XGBoost model all exceeded 0.8 on the validation set. Furthermore, the XGBoost model maintained consistent performance on the test set, with an AUC of 0.900 and a PR-AUC of 0.648. XGBoost is recognised as an ML technique that efficiently and flexibly manages missing data and integrates weak predictive models into a robust predictive framework.⁴⁰ As an open-source package, XGBoost has gained significant recognition in various ML and data mining competitions. For example, in 2015, 17 out of the 29 winning solutions featured on Kaggle's blog used XGBoost, and all of the top 10 winning teams in the 2015 KDD Cup also incorporated XGBoost into their solutions.⁴¹ In neurology, XGBoost achieved AUC values of 0.950 (mortality) and 0.958 (functional outcomes) in patients with aneurysmal subarachnoid haemorrhage, outperforming LR.42 XGBoost has been applied to predict 5-year survival in elderly patients with intrahepatic cholangiocarcinoma (AUC=0.713, SEER database) and type 2 diabetes risk (accuracy=89.09%, AUC=0.9182 in Beijing residents).^{43 44} These advancements highlight XGBoost's utility in high-dimensional clinical data sets with interpretable feature insights. Furthermore, our findings indicate that the predictive efficacy of the XGBoost model surpasses that of LR, RF, SVM and KNN. This aligns with previous research demonstrating that traditional LR frequently exhibits comparatively lower AUC values in ROC curve analyses, alongside higher prediction errors and

inferior performance relative to more contemporary methodologies.^{45 46}

Screening for ONIHL is important, and various methods have been explored for this purpose. Otoacoustic emissions (OAE) testing, particularly distortion product OAE, is a sensitive tool for detecting early cochlear damage before significant hearing threshold shifts appear in PTA.⁴⁷ It can identify subtle outer hair cell dysfunction in noise-exposed individuals, making it valuable for early intervention and monitoring.⁴⁸ Auditory brainstem response (ABR) testing, T another physiological method, assesses neural integrity and can detect hidden hearing loss even when audiometric thresholds remain normal.49 Despite their advantages, the large-scale application of OAE and ABR in occupaŝ 8 tional screening is limited by high costs, equipment availability and the need for trained operators. In contrast, our model predicts ONIHL risk solely from routine blood and biochemical indicators, eliminating the need for specialised audiometric assessments or noise exposure data. By analysing markers linked to inflammation, oxidative stress and immune response, it provides a cost-effective, scalable alternative for early screening. Integrating this approach with existing methods like OAE or ABR could further . uses enhance ONIHL risk assessment, enabling earlier interventions before irreversible damage occurs.

Consequently, the early identification and intervention of risk factors identified in our model could have substantial implications for the prevention of ONIHL among workers exposed to noise. The risk factors contributing to the devele opment of ONIHL are varied. We have developed a risk assessment model for ONIHL using clinical data and routine physical examination indicators, employing an ML algorithm. This approach contrasts with most existing methods for predicting ONIHL risk, which predominantly depend on variables such as age, sex, medical history (including conditions like hypertension and diabetes), history of noise exposure and behavioural factors such as smoking and physical activity.^{16 50 51} For instance, prior research has developed able predictive outcomes. These models primarily incorexposure and median peak intensity, which contrast with the physical examination indicators used in our study.²⁰ Wang *et al*¹⁰ formulated an ML-based risk assessment model for high-frequency hearing loss employing routine physical examination data, attaining an AUC of 0.868. This model, however, was principally designed for community residents and incorporated risk factors, including 13 blood **g** test indicators, demographic characteristics, disease-related features, behavioural factors, environmental exposure and auditory cognitive factors, which differ from the population of noise-exposed workers in our study. Our model offers a more comprehensive approach than previous research by integrating a wide range of biochemical and routine blood indicators to assess the risk of ONIHL from multiple dimensions. Unlike models that rely on hearing assessments and direct noise exposure measurements, our model focuses on routine blood and biochemical indicators, reducing the

need for specialised equipment and resources. This makes it a more efficient, cost-effective alternative for early detection and prevention of ONIHL, offering personalised risk assessments without the reliance on extensive testing.

Routine blood tests administered at occupational disease prevention clinics are typically conducted on an annual basis. Based on these tests, the application of these indicators can enhance early screening and provide warnings for prevalent occupational diseases. In our study, the developed model demonstrates the significance of haematological test data in screening for ONIHL. This includes variables such as age, sex, inflammatory and immune markers (eg, WBC, LYMPHP, MOC, BAP, EOC and GRANC), as well as oxidative stress and metabolic markers (eg, ALB, Scr, RDW-CV and RDW-SD). Noise exposure influences haematological parameters through complex immunoinflammatory pathways, which may both reflect and exacerbate cochlear damage. Studies have shown that ONIHL is closely associated with systemic immune and inflammatory responses, with WBC serving as a key inflammatory marker linked to ONIHL. A study analysing health examination data from 3508 noise-exposed workers found that WBC levels were significantly higher in the NIHL group compared with those with normal hearing.⁵² This suggests that noise exposure may trigger chronic inflammatory responses in the body. At the cellular level, noise activates resident macrophages in the cochlea, triggering the release of pro-inflammatory cytokines such as interleukin-1 β and tumour necrosis factor- α .⁵³

⁵⁴ This leads to increased permeability of the blood-labyrinth barrier, facilitating the infiltration of systemic immune cells-including MOC, GRANC and adaptive immune lymphocytes—into the inner ear.⁵⁵ This immune influx amplifies local inflammation, creating a microenvironment that promotes sensory cell apoptosis and spiral ganglion degeneration. Notably, our study identified significant elevations in WBC, MOC and GRANC, aligning with these immunological responses. Chronic noise stress further disrupts systemic immune homeostasis, as demonstrated in animal models.⁵⁵ Prolonged exposure induces immunosuppressive changes, including a decrease in LYMPHP and a reduced CD4+/CD8+ T cell ratio, which may impair anti-inflammatory responses and regenerative capacity. This systemic immune imbalance is consistent with our findings of decreased LYMPHP in noise-exposed individuals. Additionally, our findings indicate that increased EOC levels may also serve as a risk factor for ONIHL. EOC may play a pathogenic role in noise-induced inner ear vasculitis, a process increasingly recognised as a critical mediator of sensorineural damage. Elevated EOC levels have been associated with SSNHL, suggesting their potential as prognostic indicators in inflammatory hearing disorders.⁵⁶

Oxidative stress is a key mechanism underlying ONIHL.⁵⁷ RDW-CV/SD is a crucial marker of RBC oxidative damage, with increased RDW levels indicating decreased membrane stability.^{58 59} This instability shares a common pathological basis with noise-induced hair cell apoptosis. A positive correlation between RDW parameters (CV and SD) and the average hearing threshold further highlights the role of

oxidative stress in ONIHL.⁶⁰ These findings underscore the need to identify inflammatory conditions when screening workers at risk for chronic inflammation and ONIHL. ALB, a critical antioxidant protein, plays a protective role in maintaining blood-labyrinth barrier integrity.⁶¹ Low ALB levels may weaken antioxidant defences and increase susceptibility to hearing loss. Notably, ALB levels in patients with SSNHL were significantly lower than those in the control group (p<0.001).⁶² Furthermore, studies have shown a positive correlation between reduced eGFR and hearing loss, suggesting that impaired kidney function may contribute to cochlear microcirculatory dysfunction via inflammationmediated mechanisms.⁶³ Since Scr is a key component in calculating eGFR, our observed reduction in Scr could 9 reflect a broader metabolic or physiological shift rather than 8 direct renal impairment alone. PDW, an indicator of platelet activation, may also play a role in ONIHL by promoting microvascular inflammation. Research has demonstrated microvascular inflammation. Research has demonstrated a significant association between PDW and the severity of SSNHL.⁶⁴ This finding is consistent with our study, as we also observed an association between PDW and ONIHL. Given that PDW is an indicator of platelet activation, its potential role in promoting microvascular inflammation may uses r contribute to the pathophysiological mechanisms underlying ONIHL. Our results further support the notion that vascular and inflammatory responses play a crucial role in noise-induced cochlear damage. Age and male gender have been identified as risk factors for hearing loss.⁶⁵ Leveraging artificial intelligence and big data analysis, haematological e parameters can serve as predictive markers for ONIHL. An ML model based on XGBoost integrates inflammatory, oxidative stress and metabolic-related indicators to enhance risk assessment. Feature importance analysis highlights a ALB, PDW, RDW-CV, Scr and LYMPHP as key predictors of ONIHL, reinforcing their potential role in early detection and risk stratification.

Although haematological indicators provide a low-cost training and accessible approach for ONIHL prediction, their specificity remains limited, necessitating integration with objective auditory assessments such as ABR and OAE to enhance predictive accuracy. Additionally, the generalisability of our model requires further validation as it is currently based on a Shenzhen population and may not fully represent other demographic and occupational groups. The model's precision and F1 score are also relatively low, primarily due to the severe class imbalance, with ONIHL cases being far less frequent than noise-exposed individuals with normal hearing. Despite these limitations, future studies can address these challenges by conducting large-scale, multicentre validations, employing advanced data-balancing techniques, and incorporating multi-omics data-such as metabolomics and transcriptomics-to unravel the molecular mechanisms linking inflammation, oxidative stress and immune dysregulation in ONIHL. Such advancements will not only optimise predictive models but also facilitate their clinical application in occupational health screening and early intervention, ultimately improving hearing loss prevention strategies.

Protected by copyright, including for uses related

CONCLUSION

In this study, we developed five ML models to construct a risk screening model for ONIHL, with the XGBoost-based model demonstrating superior performance. By integrating biochemical and haematological indicators with ML techniques, this model effectively identifies individuals at high risk for ONIHL. This approach not only introduces a novel tool for the early screening of hearing loss but also lays the groundwork for the development of personalised intervention strategies. In the future, the integration of additional biological data is anticipated to further augment the model's predictive capabilities. Furthermore, this model holds potential for extension to forecast risks associated with other occupational or chronic diseases, thereby offering substantial support for the maintenance and enhancement of public health.

Acknowledgements We thank the Shenzhen Prevention and Treatment Center for Occupational Diseases for the approval of the ethical clearance. We also extend our warm gratitude to the different hospital stakeholders and participants for their valuable contribution during data collection.

Contributors The authors made substantial contributions to the acquisition, analysis and interpretation of the data and the drafting and revision of the manuscript. All authors also approved the final version of the article and agreed to be accountable for all aspects of the work. CL and DW: writing-original draft, Investigation, data curation and conceptualsation. CL, LS and LC: methodology and data curation. DL: data curation, writing-original draft, supervision, project administration, formal analysis and conceptualisation, XY and LZ: investigation, PL: validation and investigation. WZ: validation. YG and NZ: supervision, project administration and conceptualisation. CL is the guarantor of this article.

Funding This work was supported by Science and Technology Planning Project of Shenzhen. Municipality (No.KCXFZ20201221173602007, No.JCYJ20240813162306009), Shenzhen Fund for Guangdong Provincial Highlevel Clinical Key Specialties (No.SZGSP015).

Disclaimer The content of this study is solely the responsibility of the authors and does not necessarily represent the official views of the Science and Technology Planning Project of Shenzhen Municipality or the Shenzhen Fund for Guangdong Provincial High-Level Clinical Key Specialties.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study was approved by the Ethics Committee of Shenzhen Prevention and Treatment Center for Occupational Diseases (approval number: LL2020-34; date: 14 December 2020). All methods were carried out in accordance with relevant ethical guidelines and regulations.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Data are available upon reasonable request. Original data collected within this study are not publicly available as they might contain sensitive information. De-identified data can be shared based on a reasonable request by sending an email to szpcr@126. com

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which

permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs

Caiping Li http://orcid.org/0009-0006-1431-2058 Liuwei Shi http://orcid.org/0009-0003-9811-1103 Linlin Chen http://orcid.org/0009-0002-6967-4997 Dafeng Lin http://orcid.org/0000-0002-3810-9875 Xiangli Yang http://orcid.org/0009-0003-6219-6088 Peimao Li http://orcid.org/0009-0002-2494-4236 Wen Zhang http://orcid.org/0009-0005-0655-7824 Wenting Feng http://orcid.org/0009-0009-5302-1175 Yan Guo http://orcid.org/0009-0002-1347-3942 Liang Zhou http://orcid.org/0000-0001-5973-7873 Naixing Zhang http://orcid.org/0009-0004-5362-7103 Dianpeng Wang http://orcid.org/0000-0002-1886-2886

REFERENCES

- Ding T, Yan A, Liu K. What is noise-induced hearing loss? Br J Hosp Med 2019;80:525-9.
- Nelson DI, Nelson RY, Concha-Barrientos M, et al. The global burden of occupational noise-induced hearing loss. Am J Ind Med 2005:48:446-58
- Themann C, Suter A, Stephenson M. National Research Agenda for З the Prevention of Occupational Hearing Loss-Part 1. Semin Hear 2013;34:145-207.
- Themann CL, Masterson EA. Occupational noise exposure: A review of its effects, epidemiology, and impact with recommendations for reducing its burden. J Acoust Soc Am 2019;146:3879.
- Távora-Vieira D, Acharya A, Rajan GP. What can we learn from adult cochlear implant recipients with single-sided deafness who became elective non-users? Cochlear Implants Int 2020;21:220-7.
- 6 Shi Z, Zhou J, Huang Y, et al. Occupational Hearing Loss Associated With Non-Gaussian Noise: A Systematic Review and Meta-analysis. Ear Hear 2021;42:1472–84.
- Vlaming MSMG, MacKinnon RC, Jansen M, et al. Automated 7 screening for high-frequency hearing loss. Ear Hear 2014;35:667-79.
- 8 Bhatt IS, Washnik N, Torkamani A. Suprathreshold Auditory Measures for Detecting Early-Stage Noise-Induced Hearing Loss in Young Adults. *J Am Acad Audiol* 2022;33:185–95.
- Cunningham LL, Tucci DL. Hearing Loss in Adults. N Engl J Med 9 2017;377:2465-73.
- 10 Wang Y, Yao X, Wang D, et al. A machine learning screening model for identifying the risk of high-frequency hearing impairment in a general population. BMC Public Health 2024;24:1160.
- McKearney RM, MacKinnon RC. Objective auditory brainstem 11 response classification using machine learning. Int J Audiol 2019;58:224-30.
- 12 Chang Y-S, Park H, Hong SH, et al. Predicting cochlear dead regions in patients with hearing loss through a machine learning-based approach: A preliminary study. PLoS One 2019;14:e0217790.
- 13 Abdollahi H, Mostafaei S, Cheraghi S, et al. Cochlea CT radiomics predicts chemoradiotherapy induced sensorineural hearing loss in head and neck cancer patients: A machine learning and multivariable modelling study. Phys Med 2018;45:192-7.
- 14 Tomiazzi JS, Pereira DR, Judai MA, et al. Performance of machinelearning algorithms to pattern recognition and classification of hearing impairment in Brazilian farmers exposed to pesticide and/or cigarette smoke. Environ Sci Pollut Res Int 2019;26:6481-91.
- 15 Bing D, Ying J, Miao J, et al. Predicting the hearing outcome in sudden sensorineural hearing loss via machine learning models. Clin Otolaryngol 2018;43:868-74.
- Aliabadi M, Farhadian M, Darvishi E. Prediction of hearing loss 16 among the noise-exposed workers in a steel factory using artificial intelligence approach. Int Arch Occup Environ Health 2015:88:779-87
- Farhadian M, Aliabadi M, Darvishi E. Empirical estimation of the 17 grades of hearing impairment among industrial workers based on new artificial neural networks and classical regression methods. Indian J Occup Environ Med 2015;19:84-9.
- 18 Kim YS, Cho YH, Kwon OJ, et al. The Risk Rating System for Noiseinduced Hearing Loss in Korean Manufacturing Sites Based on the 2009 Survey on Work Environments. Saf Health Work 2011;2:336-47.
- 19 Nawi NM, Rehman MZ, Ghazali MI. Noise-induced hearing loss prediction in Malaysian industrial workers using gradient descent

with adaptive momentum algorithm. *Int Rev Comput Softw* 2011;6:740–8. Available: https://api.semanticscholar.org/CorpusID: 110206172

- 20 Zhao Y, Li J, Zhang M, *et al.* Machine Learning Models for the Hearing Impairment Prediction in Workers Exposed to Complex Industrial Noise: A Pilot Study. *Ear Hear* 2019;40:690–9.
- 21 Li P, Pang K, Zhang R, et al. Prevalence and risk factors of hearing loss among the middle-aged and older population in China: a systematic review and meta-analysis. *Eur Arch Otorhinolaryngol* 2023;280:4723–37.
- 22 Tsimpida D, Kontopantelis E, Ashcroft D, *et al.* Socioeconomic and lifestyle factors associated with hearing loss in older adults: a cross-sectional study of the English Longitudinal Study of Ageing (ELSA). *BMJ Open* 2019;9:e031030.
- 23 Baiduc RR, Sun JW, Berry CM, et al. Relationship of cardiovascular disease risk and hearing loss in a clinical population. Sci Rep 2023;13:1642.
- 24 Soylemez E, Avci I, Yildirim E, et al. Predicting noise-induced hearing loss with machine learning: the influence of tinnitus as a predictive factor. J Laryngol Otol 2024;138:1030–5.
- 25 Jung DJ, Do JY, Cho KH, et al. Association between triglyceride/ high-density lipoprotein ratio and hearing impairment in a Korean population. *Postgrad Med* 2017;129:943–8.
- 26 Verschuur CA, Dowell A, Syddall HE, et al. Markers of inflammatory status are associated with hearing threshold in older people: findings from the Hertfordshire Ageing Study. Age Ageing 2012;41:92–7.
- 27 Nonoyama H, Tanigawa T, Shibata R, et al. Red blood cell distribution width predicts prognosis in idiopathic sudden sensorineural hearing loss. Acta Otolaryngol 2016;136:1137–40.
- 28 Lunardon N, Menardi G, Torelli N. ROSE: a Package for Binary Imbalanced Learning. R J 2014;6:79.
- 29 Elkahwagy D, Kiriacos CJ, Mansour M. Logistic regression and other statistical tools in diagnostic biomarker studies. *Clin Transl Oncol* 2024;26:2172–80.
- 30 Schauberger G, Klug SJ, Berger M. Random forests for the analysis of matched case-control studies. *BMC Bioinformatics* 2024;25:253.
- 31 Valkenborg D, Rousseau A-J, Geubbelmans M, et al. Support vector machines. Am J Orthod Dentofacial Orthop 2023;164:754–7.
- 32 Garcia-Carretero R, Vigil-Medina L, Mora-Jimenez I, et al. Use of a K-nearest neighbors model to predict the development of type 2 diabetes within 2 years in an obese, hypertensive population. *Med Biol Eng Comput* 2020;58:991–1002.
- 33 Bridgelall R, Tolliver DD. Railroad accident analysis using extreme gradient boosting. Accid Anal Prev 2021;156:106126.
- 34 Xia Z, Chen Y, Xu C. Multiview PCA: A Methodology of Feature Extraction and Dimension Reduction for High-Order Data. *IEEE Trans Cybern* 2022;52:11068–80.
- 35 Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Machine Intell* 2005;27:1226–38.
- 36 Śliwińska-Kowalska M, Zaborowski K. WHO Environmental Noise Guidelines for the European Region: A Systematic Review on Environmental Noise and Permanent Hearing Loss and Tinnitus. Int J Environ Res Public Health 2017;14:1139.
- 37 Seixas NS, Neitzel R, Stover B, et al. A multi-component intervention to promote hearing protector use among construction workers. Int J Audiol 2011;50 Suppl 1:S46–56.
- 38 Amjad-Sardrudi H, Dormohammadi A, Golmohammadi R, et al. Effect of noise exposure on occupational injuries: a cross-sectional study. J Res Health Sci 2012;12:101–4.
- 39 Park S, Johnson MD, Hong O. Analysis of Occupational Safety and Health Administration (OSHA) noise standard violations over 50 years: 1972 to 2019. *American J Industrial Med* 2020;63:616–23.
- 40 Yuan K-C, Tsai L-W, Lee K-H, *et al*. The development an artificial intelligence algorithm for early sepsis diagnosis in the intensive care unit. *Int J Med Inform* 2020;141:104176.
- 41 Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA: ACM; 2016:785–94.
- 42 Wang R, Zhang J, Shan B, *et al.* XGBoost Machine Learning Algorithm for Prediction of Outcome in Aneurysmal Subarachnoid Hemorrhage. *Neuropsychiatr Dis Treat* 2022;18:659–67.

- 43 Xu Q, Lu X. Development and validation of an XGBoost model to predict 5-year survival in elderly patients with intrahepatic cholangiocarcinoma after surgery: a SEER-based study. *J Gastrointest Oncol* 2022;13:3290–9.
- 44 Wang L, Wang X, Chen A, *et al.* Prediction of Type 2 Diabetes Risk and Its Effect Evaluation Based on the XGBoost Model. *Healthcare* (*Basel*) 2020;8:247.
- 45 Xiao J, Ding R, Xu X, *et al.* Comparison and development of machine learning tools in the prediction of chronic kidney disease progression. *J Transl Med* 2019;17:119.
- 46 Li Y-M, Li Z-L, Chen F, et al. A LASSO-derived risk model for longterm mortality in Chinese patients with acute coronary syndrome. *J Transl Med* 2020;18:157.
- 47 Soylemez E, Mujdeci B. Evaluation of auditory disability and cochlear functions in industrial workers exposed to occupational noise. *Hearing Balance Commun* 2021;19:16–20.
- 48 Kapoor N, Mani KV, Shukla M. Distortion product oto-acoustic emission: a superior tool for hearing assessment than pure tone audiometry. *Noise Health* 2019;21:164–8.
- 49 Mehraei G, Gallardo AP, Shinn-Cunningham BG, et al. Auditory brainstem response latency in forward masking, a marker of sensory deficits in listeners with normal hearing thresholds. *Hear Res* 2017;346:34–44.
- 50 Chen F, Cao Z, Grais EM, et al. Contributions and limitations of using machine learning to predict noise-induced hearing loss. Int Arch Occup Environ Health 2021;94:1097–111.
- 51 Sun R, Shang W, Cao Y, et al. A risk model and nomogram for highfrequency hearing loss in noise-exposed workers. *BMC Public Health* 2021;21:747.
- 52 Yingjun L, Weisen Z, Hao Z, *et al.* Study on the correlation between noise-induced hearing loss and white blood cell count. *International Medicine and Health Guidance News* 2023;29:115–8.
- 53 Frye MD, Ryan AF, Kurabi A. Inflammation associated with noiseinduced hearing loss. *J Acoust Soc Am* 2019;146:4020–32.
- 54 Rai V, Wood MB, Feng H, et al. The immune response after noise damage in the cochlea is characterized by a heterogeneous mix of adaptive and innate immune cells. Sci Rep 2020;10:15167.
- 55 Aguas AP, Esaguy N, Grande N, et al. Effect low frequency noise exposure on BALB/c mice splenic lymphocytes. Aviat Space Environ Med 1999;70:A128–31.
- 56 Zhiwei W, Yuewen L, Xiaohui D, *et al.* Analysis of related factors between sudden sensorineural hearing loss and serum indices base on artificial intelligence and big data. *Lin Chuang Er Bi Yan Hou Tou Jing Wai Ke Za Zhi* 2020;34:977–80.
- 57 Zhou Y, Fang C, Yuan L, *et al*. Redox homeostasis dysregulation in noise-induced hearing loss: oxidative stress and antioxidant treatment. *J Otolaryngol Head Neck Surg* 2023;52:78.
- 58 Shi X. Cochlear vascular pathology and hearing loss. In: Ramkumar V, Rybak LP, eds. *Inflammatory mechanisms in mediating hearing loss*. Cham: Springer International Publishing, 2018: 61–90.
- 59 Jung DJ, Yoo MH, Lee K-Y. Red cell distribution width is associated with hearing impairment in chronic kidney disease population: a retrospective cross-sectional study. *Eur Arch Otorhinolaryngol* 2020;277:1925–30.
- 60 Natarajan N, Batts S, Stankovic KM. Noise-Induced Hearing Loss. *J Clin Med* 2023;12:2347.
- 61 Belinskaia DA, Voronina PA, Shmurak VI, et al. Serum Albumin in Health and Disease: Esterase, Antioxidant, Transporting and Signaling Properties. Int J Mol Sci 2021;22:10318.
- 62 Zheng Z, Liu C, Shen Y, et al. Serum Albumin Levels as a Potential Marker for the Predictive and Prognostic Factor in Sudden Sensorineural Hearing Loss: A Prospective Cohort Study. Front Neurol 2021;12:747561.
- 63 Liu W, Meng Q, Wang Y, *et al.* The association between reduced kidney function and hearing loss: a cross-sectional study. *BMC Nephrol* 2020;21:145.
- 64 Mirvakili A, Dadgarnia MH, Baradaranfar MH, *et al.* Role of Platelet Parameters on Sudden Sensorineural Hearing Loss: A Case-Control Study in Iran. *PLoS One* 2016;11:e0148149.
- 65 Chou C-F, Beckles GLA, Zhang X, et al. Association of Socioeconomic Position With Sensory Impairment Among US Working-Aged Adults. Am J Public Health 2015;105:1262–8.