This is an Accepted Manuscript for *The Journal of Laryngology & Otology*
DOI: 10.1017/S002221512400094X

Predicting noise-induced hearing loss with machine learning: The influence of tinnitus as a predictive factor

Emre Soylemez[1,2,7], Isa Avci[3], Elif Yildirim[3], Engin Karaboya[4], Nihat Yilmaz[5], Süha Ertugrul[5], Suna Tokgoz-Yilmaz[6,7]

1- Department of Audiometry, Vocational School of Health Services, Karabuk University, Karabuk, Turkey.
2- Ankara University, Health Sciences Institute, Audiology and Speech Pathology Ph.D. Program, Ankara, Turkey.
3- Department of Computer Engineering, Karabuk University Turkey.
4- Department of Audiology, Karabuk Training and Research Hospital, Turkey.
5- Department of Otorhinolaryngology, Karabuk University, Turkey.
6- Department of Audiology, Faculty of Health Sciences, Ankara University, Ankara, Turkey.
7- Audiology, Balance and Speech Disorders Unit, Medical Faculty, Ankara University, Ankara, Turkey.

Authors

Emre Soylemez (Lecturer, ***Corresponding Author***); ORCID ID: 0000-0002-7554-3048

e-mail: emresoylemez@karabuk.edu.tr, Phone: +905523952511, Address: Karabük University Vocational School of Health Services, Audiometry Program

Isa Avci (Assistant professor);

e-mail: isaavci@karabuk.edu.tr, ORCID ID: 0000-0001-7032-8018

Elif Yıldırım (MSc Student);

e-mail: eliff.celikel@gmail.com ORCID ID: 0000-0001-6336-6390

Engin Karaboya (Audiologist);

e-mail: karaboya122112@gmail.com ORCID ID: 0000-0003-2458-2052

Nihat Yilmaz (Associate Professor),

e-mail: nihatyilmaz@karabuk.edu.tr ORCID ID: 0000-0003-1575-1280

Suha Ertugrul (Associate Professor);

e-mail: drsuhaertugrul@hotmail.com ORCID ID: 0000-0001-9743-6924

Suna Tokgoz-Yilmaz (Professor);

e-mail: sunayilmaz11@gmail.com ORCID ID: 0000-0002-4656-099X

https://doi.org/10.1017/S002221512400094X Published online by Cambridge University Press

Abstract

Objective: This study aims to determine which machine learning (ML) model is most suitable for predicting noise-induced hearing loss (NIHL) and the effect of tinnitus on the models' accuracy.

Method: Two hundred workers employed in a metal industry were selected for this study and tested using pure tone audiometry. Their occupational exposure histories were collected, analysed, and used to create a dataset. Eighty percent of the data collected was used to train six ML models, and the remaining 20% was used to test the models.

Results: Eight (40.5%) workers had bilaterally normal hearing, and 119 (59.5%) had hearing loss. Tinnitus was the second most important indicator after age for NIHL. The support vector machine (SVM) was the best-performing algorithm with 90% accuracy, 91% F1-score, 95% precision, and 88% recall.

Conclusion: The use of tinnitus as a risk factor in the SVM model may increase the success of occupational health and safety programs.

Keywords: Machine learning, noise-induced hearing loss, occupational disease, workers

Introduction

Despite being preventable, noise-induced hearing loss (NIHL) is one of the most common types of sensorineural hearing loss. NIHL refers to damage to the inner ear caused by prolonged exposure to high levels of noise. The estimated worldwide prevalence of NIHL is 16%, with a 7% prevalence in Western countries and 21% in developing countries.[1] After presbycusis, NIHL is the second most common cause of sensorineural hearing loss.[2] The severity of NIHL depends on both the intensity and duration of exposure to noise, as well as individual factors. Although hearing loss typically progresses slowly, it can eventually reach moderate or even severe levels over time. NIHL can have negative impacts on workers' communication skills, work performance, and quality of life. Additionally, exceeding the hearing level of 40 dB is classified as a disability.[3] As an occupational disease, NIHL affects not only workers and employers but also government budgets. Since there is currently no medical or surgical treatment available for NIHL,[4,5] early diagnosis is critical in preventing some of the adverse effects. The potential usefulness of methods such as otoacoustic emissions in the early detection of NIHL is currently a topic of research.

Machine learning algorithms (MLAs), a new group of statistical methods primarily used in software and engineering fields, are preferred for analysing nonlinear multidimensional complex events and uncertain information.[6] MLAs have the ability to automatically generate new rules based on input data and can estimate unknown data that may be difficult to define manually.[7] In other words, a dataset with known risk factors (input) and outcomes (output) can be taught to MLAs. After training, MLAs can estimate outputs for new inputs that are presented to them. Some studies have investigated the use of MLAs in the detection of occupational diseases.[6,8] Environmental and individual factors (such as age and noise intensity) that play a role in the formation of NIHL can be used to predict hearing loss using MLAs.[8,9]

Few studies in the literature predict NIHL with MLAs.[6,8,9] Unlike these studies, we included tinnitus as an input in our study. This study aims to determine which MLA is more suitable for predicting NIHL and the effect of tinnitus on the models' accuracy. According to our hypothesis, using tinnitus, one of the early markers of NIHL, as an input may increase the accuracy of MLA models.

Materials and methods

Participant selection

https://doi.org/10.1017/S002221512400094X Published online by Cambridge University Press

This prospective study was carried out on metal industry workers who presented at the otolaryngology outpatient clinic and were referred for hearing tests. All workers had been working in the machinery area for at least one year and were exposed to noise at a minimum level of 85 dB(*A*). During the audiological examination, a detailed anamnesis was obtained from each participant. The questionnaire included the following questions: age (years), duration of exposure to noisy environments (years), frequency of ear protection equipment (EPE) use (never, sometimes, or continuously), smoking status (yes or no, and if yes, how many years), and the presence of tinnitus (right ear, left ear, or bilateral). Workers with perforation of the eardrum, type *B* and *C* tympanograms, conductive and combined hearing loss, and hearing loss due to another reason (congenital hearing loss, sudden hearing loss, etc.) were not included in the study. A pure tone audiometry test was administered to all 200 male workers to determine their hearing thresholds. The anamnesis data, which included risk factors for NIHL, were used as inputs to train the MLAs to estimate the probability of hearing loss (output) in these workers (as shown in Figure 1). We obtained both verbal and written consent from all participants in accordance with the Declaration of Helsinki. The study was approved by the ethics committee of the university (approval number: 2022/838).

Audiological evaluation

A pure-tone audiometry test was administered bilaterally to all workers using the Madsen Astera (GN Otometrics Taastrup, Denmark) in a soundproof room. The air conduction hearing thresholds in the range of 250-6000 Hz were determined using TDH 39 supraaural headphones, while the bone conduction hearing thresholds in the range of 500-4000 Hz were determined using the Radioear B71 bone vibrator. The tympanometric examination was performed with the Interacoustics AZ 26 (Middelfart, Denmark) with a 226 Hz probe tone. Pure tone average (PTA) was calculated by taking the arithmetic mean of the frequency band thresholds (500, 1000, 2000, and 4000 Hz). A PTA greater than 20 dB in at least one ear was considered to indicate hearing loss.

Statistical analysis and machine learning models

The Inernational Business Machines Statistical Package for the Social Science 21 (IBM SPSS Corp.; Armonk, NY, USA) was used for statistical analysis. Variables that met the normality assumption were presented as mean±SD, and variables that did not meet the normality assumption were presented as median (min-max). The compliance of the variables with normality distribution was checked with the Shapiro-Wilk test. Mann Whitney U test and χ^2

were used to compare hearing loss groups and risk factors. In all statistical analyses, $p<0.05$ was accepted as the statistical significance level.

Python programming language (Version 3.7) was used to develop the machine learning algorithm. Machine learning algorithms can be classified as supervised, unsupervised, and reinforced reinforcement learning types. Supervised learning algorithms are used in classification and regression problems.[10] In this study, K-nearest neighbours (KNN), decision trees (DT), random forest (RF), support vector machine (SVM), logistic regression (LR), and XGBoost, which are considered supervised algorithms, were used. The performance of these models was evaluated using accuracy, precision, F1-score, recall, and area under the receiver operating characteristic curve (ROC) Curve (ROC-AUC).[11]

1- $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$

2- $Precision = \frac{TP}{TP+FP}$

3- $F1\ Score = 2 \times \frac{Recall \times Precision}{Recall + Precision}$

4- $Recall = \frac{TP}{TP+FN}$

TP: True positive, TN: True negative, FN: False negative, FP: False positive

These algorithms were trained using 160 data points (80% of the data), and the success of the algorithms was tested with the remaining 40 data points (20% of the data).

K-nearest neighbours (KNN)

KNN classifies newly obtained data by assigning it to the class of the nearest similar neighbours. It uses two basic metrics: distance and k neighborhood ratios.[12]

Decision tree (DT)

It is a supervised learning algorithm used in classification and regression problems. The decision tree algorithm tries to solve the problem by representing the data in tree form. Each decision node corresponds to a variable, and each leaf node corresponds to its target tag. The following sequence is followed while creating the tree.[13]

- The most suitable variable is put at the root of the tree. It should be as simple as possible.

- The dataset is divided into subsets.

- The above operations are continued until the leaf reaches the nodes.

https://doi.org/10.1017/S002221512400094X Published online by Cambridge University Press

Random forest (RF)

This ensemble learning algorithm combines the decisions of many independent multivariate trees. RF is a classification model that attempts to create more accurate and compatible models using multiple decision trees.[14] It uses averaging to improve forecast accuracy and control overfitting.

Support vector machine (SVM)

SVM is a learning algorithm that can be used for classification and regression analysis. Data points are separated by a line or hyperplane to divide them into two or more classes. The gap between the two classes should be as large as possible to reduce errors during classification.[14]

Logistic regression (LR)

LR is a regression method for classification. LR is used to reveal the effect of one or more variables on the overall outcome. Because of this feature, it is the preferred research approach to find the strongest variable among the independent variables and to predict the estimating the output variable. While performing LR, attention should be paid to the variables' independence and the validity of assumptions to ensure appropriate modelling.[15]

XGBoost

XGBoost is a reinforced tree algorithm model based on gradient-boosting principles. Compared to other approaches, XGBoost applies more systematic model reinforcement to control overfitting, thus aiming to improve performance.[16] This algorithm makes corrections to errors after making predictions. The performance of XGBoost depends on parallelism and hardware optimization.

Results

The workers had a mean age of 39.96±10.96 (range 19-57) years. Eighty-one workers (40.5%) had a normal bilateral hearing, and 119 (59.5%) had hearing loss. Among those with hearing loss, 93 (78.15%) had bilateral hearing loss, and 26 (21.84%) had unilateral hearing loss. Hearing thresholds of workers with and without hearing loss are presented in Table 1.

We compared age, exposure duration to industrial noise, smoking, use of EPE, and the presence of tinnitus between workers with and without hearing loss. Ageing, duration of exposure to industrial noise, smoking, not using EPE, and tinnitus were risk factors for NIHL ($p<0.05$). Conditions that are risk factors for NIHL are presented in Table 2.

The SVM was the best-performing algorithm with 90% accuracy, 91% F1-score, 95% precision, 88% recall, and 90.6% ROC-AUC. The confusion matrix and the ROC curve of the SVM model are presented in Figures 2a and 2b, respectively. Our study showed that age and tinnitus contributed the highest to the overall result in the SVM model (Figure 2c). The DT algorithm was the second-best-performing algorithm with 87.5% accuracy, 89% F1-score, 100% precision, 79% recall, and 84.8% ROC-AUC. On the other hand, the KNN algorithm was the worst-performing algorithm with 80% accuracy, 83% F1-score, 86% precision, 79% recall, and 80.4% ROC-AUC. The performances of the LR, RF, SVM, DT, KNN and XGBoost models in test sets are presented in Table 3.

Discussion

Long-term exposure to workplace noise affects the inner ear in primarily three stages.[17,18] The first stage involves minor damage to hair cells that occurs with the first exposure to noise. This damage cannot be detected with a pure tone audiometry test. However, individuals may experience auditory disturbances, such as tinnitus and hyperacusis, as well as non-auditory disorders, including headaches, fatigue, and stress.[17] The second stage occurs when the noise exposure continues for months or years and damages the basal part of the cochlea due to the resonance frequency effect in the external auditory canal. This damage can be detected as acoustic notches on the audiogram at 3, 4, or 6 kHz. Speech intelligibility is usually not severely affected at this stage, and the damage may go unnoticed without a hearing test. The severity of NIHL may rapidly increase and reach a plateau at the end of this stage. The third stage occurs with long-term exposure to chronic noise, which often leads to a decline in communication skills, and seeking treatment for hearing loss becomes necessary.[17] The goal of workplace hearing screenings, which is mandated by occupational health and safety regulations, is to detect NIHL in its early stages and implement necessary measures promptly. This study utilised data from 200 workers, with NIHL risk factors as input and hearing test results as output, to predict the likelihood of NIHL using MLAs. Of the 200 workers, the data of 160 (80%) were used as training data, and the data of 40 (20%) as test data. Six MLAs (KNN, DT, RF, SVM, LR, and XGBoost) were trained using the training data. The overall accuracy of the six models ranged from 80% to 90%, with the SVM performing best at accuracy 90%.

NIHL is a multifactorial disease that arises from the interplay of genetic, individual, and environmental factors. Nevertheless, the biological damage incurred by individuals is linked to the total amount of noise (the fundamental energy level).[18] The equal-energy principle (Leq) posits that equal energy exposure leads to an equal amount of biological damage, which is determined by the sound pressure level and duration of noise exposure. Therefore, sound pressure level and exposure time are crucial risk factors for hearing loss. In our study, all participants were machinery area workers in the metal industry, and their exposure to noise levels was similar. So, we evaluated the use of EPE (input) as a potential risk factor. EPE is designed to reduce the intensity of noise before it reaches the inner ear. Regular and continuous use of EPE can prevent up to 30% of hearing loss.[19] Ramakers et al., showed that individuals who used EPE during outdoor music festivals reported less temporary hearing loss and tinnitus than those who did not.[20]

Other risk factors that we used as inputs in our study to train MLAs and predict NIHL are age and smoking. Ageing causes degenerations in the peripheral and central auditory systems as well as in all tissues and cells. Chronic workplace noise does not directly damage the cochlea but leads to the production of reactive oxygen species and other free radical molecules in the cochlea, the possible cause of which is metabolically overactive cochlear mitochondria, ionic fluxes, and ischemic-reperfusion.[21] Nicotine in cigarettes increases the amount of free radicals and reactive oxygen species, triggering oxidative damage similar to the effect of chronic noise exposure. It also stimulates the production of nuclear factor kappa B (NF-kB), which plays a role in inflammatory processes and cell damage.[22] Consequently, smoking and noise exposure act synergistically to increase the risk of hearing loss. Tao et al. reported that the mean hearing thresholds at 4 and 6 kHz were higher in smokers than non-smokers, and the incidence of high-frequency hearing loss was higher in smokers (48.9%) than non-smokers (33.8%).[23]

There are several studies in the literature that have used machine learning algorithms to predict NIHL.[6,24] Zhao et al. estimated the hearing test results of 1113 workers in 17 different factories using four machine learning models (SVM, neural network multilayer perceptron, RF, and adaptive boosting).[6] The researchers used the age of the workers, exposure time to noise, A-weighted equivalent SPL (LAeq), and median kurtosis as inputs. The best performing algorithm in the study was the SVM model, with an accuracy of 80.1%, while the other three algorithms had accuracies ranging from 78% to 79%. Similarly, Farhadian et al., estimated the hearing test results of 210 workers in a steel factory using artificial neural networks and logistic regression.[24] In this study, the age of the workers, noise exposure level, work experience, use of EPE, and smoking status were used as inputs. The authors reported that the accuracy of artificial neural networks was 88.6% in predicting hearing loss and was better than logistic regression. In our study, we aimed to detect hearing loss in metal industry workers using KNN, DT, RF, SVM, LR, and XGBoost algorithms. Similar to Zhao et al.'s[6] findings, SVM performed best, and the accuracy rate was 90%. The accuracy of the other algorithms we used were between 80% and 87.5%. The performance of the algorithms can vary based on factors such as the number and type of inputs used, the weight ratios of the inputs, and their correlation with the outputs. Unlike previous studies, we also used the presence of tinnitus in workers as an input. We found hearing loss in 61 (95.3%) of 64 (32%) workers with tinnitus. This finding is consistent with previous reports that the prevalence of tinnitus is higher in workers exposed to excessive noise and can reach up to 80% in military personnel.[25] Indeed, tinnitus was the second most significant variable affecting the success rate in our study, and adding tinnitus as an input may have increased the accuracy rate of the SVM model.

Another study estimated the NIHL degree with the C5 algorithm, and factors affecting hearing loss were investigated.[26] The authors stated that 4 kHz had the highest effect, with a 22% in estimating the degree of hearing loss in the C5 algorithm. In our study, age had the most significant effect weight in predicting NIHL at 24%, while working time in noisy environments had the lowest effect weight at 0.7%.

Recently developed MLAs and artificial neural networks have become very interesting when applied to occupational diseases, such as NIHL. NIHL is one of the most common occupational diseases and is mainly shaped by the influence of environmental factors. Our study demonstrated that the risk of NIHL can be predicted cheaply and quickly using environmental factors and workers' characteristics in MLA (SVM model). Furthermore, it is shown that the onset age of NIHL can be detected approximately when the existing risk factors are implemented in MLA. Future studies could incorporate hearing screening scales and other diseases, such as metabolic diseases that may affect NIHL, as inputs and investigate the accuracy of the algorithms.

Summary

- Machine learning can be used to predict diseases.
- It has been stated that noise-induced hearing loss can also be predicted with machine learning.
- In the studies, the workers' age, working duration, smoking and earplug usage status were used as inputs.
- These studies achieved 80.1% and 88.6% accuracy with SVM and neural networks, respectively.

https://doi.org/10.1017/S002221512400094X Published online by Cambridge University Press

- We also used tinnitus as an input in our study, and we achieved 90% accuracy with the SVM model.

Conclusion

Incorporating early markers of hearing loss, such as tinnitus, into MLAs may enhance the prediction ratio of the models. The SVM algorithm, which holds the highest accuracy, can be used in the early detection of NIHL. Thus, the success of occupational health and safety programs for employees exposed to occupational noise can be increased.

Acknowledgment

We acknowledge the volunteers for their contribution to this study.

References

1. Nelson DI, Nelson RY, Concha-Barrientos M, Fingerhut M. The global burden of occupational noise-induced hearing loss. *Am J Ind Med* 2005;48:446-58
2. Rabinowitz PM. Noise-induced hearing loss. *Am Fam Physician* 2000;61:2749- 56
3. Schmucker C, Kapp P, Motschall E, Loehler J, Meerpohl JJ. Prevalence of hearing loss and use of hearing aids among children and adolescents in Germany: a systematic review. *BMC Public Health* 2019;19:1277
4. Manar MK, Shukla SP, Mohan U, Singh SK, Verma V. Validation of a questionnaire to identify noise-induced hearing loss among drivers. *J Family Med Prim Care* 2019;8:1196-201
5. Lapsley Miller JA, Marshall L, Heller LM, Hughes LM. Low-level otoacoustic emissions may predict susceptibility to noise-induced hearing loss. *J Acoust Soc Am* 2006;120:280-96.
6. Zhao Y, Li J, Zhang M, Lu Y, Xie H, Tian Y, *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
7. McKearney RM, MacKinnon RC. Objective auditory brainstem response classification using machine learning. *Int J Audiol* 2019;58:224-30
8. Chen F, Cao Z, Grais EM, Zhao F. Contributions and limitations of using machine learning to predict noise-induced hearing loss. *Int Arch Occup Environ Health* 2021;94:1097-111

9. Mrena R, Ylikoski M, Mäkitie A, Pirvola U, Ylikoski J. Occupational noise-induced hearing loss reports and tinnitus in Finland. *Acta Otolaryngol* 2007;127:729-35
10. Hasan M. Balbahaith Z. Tarique M. Detection of SQL injection attacks: A machine learning approach. In 2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA). *IEEE* 2019;1-6
11. Hosam, E. Hosny H. Ashraf W. Kaseb AS. SQL Injection Detection Using Machine Learning Techniques. In 2021 8th International Conference on Soft Computing & Machine Intelligence (ISCMI). *IEEE* 2021;15-20
12. Bui XN, Jaroonpattanapong P, Nguyen H, Tran QH, Long NQ. A Novel Hybrid Model for Predicting Blast-Induced Ground Vibration Based on k-Nearest Neighbors and Particle Swarm Optimization. *Sci Rep* 2019;9:13971
13. Song YY, Lu Y. Decision tree methods: applications for classification and prediction. *Shanghai Arch Psychiatry* 2015;27:130-5
14. Bandyopadhyay N, Jadhav A. Churn Prediction of Employees Using Machine Learning Techniques. *Tehnički glasnik* 2021;15:51-9
15. Stoltzfus JC. Logistic regression: a brief primer. *Acad Emerg Med* 2011;18:1099- 104
16. Jain R, Nayyar A. Predicting Employee Attrition using XGBoost Machine Learning Approach. International Conference on System Modeling & Advancement in Research Trends (SMART). *IEEE* 2018;113-20
17. Hirose K, Discolo CM, Keasler JR, Ransohoff R. Mononuclear phagocytes migrate into the murine cochlea after acoustic trauma. *J Comp Neurol* 2005;489:180-94
18. Le TN, Straatman LV, Lea J, Westerberg B. Current insights in noise-induced hearing loss: a literature review of the underlying mechanism, pathophysiology, asymmetry, and management options. *J Otolaryngol Head Neck Surg* 2017;46:41
19. Davies H, Marion S, Teschke K. The impact of hearing conservation programs on incidence of noise-induced hearing loss in Canadian workers. *Am J Ind Med* 2008;51:923–31
20. Ramakers GG, Kraaijenga VJ, Cattani G, van Zanten GA, Grolman W. Effectiveness of Earplugs in Preventing Recreational Noise-Induced Hearing Loss: A Randomized Clinical Trial. *JAMA Otolaryngol Head Neck Surg* 2016;142:551-8
21. Kamogashira T, Fujimoto C, Yamasoba T. Reactive oxygen species, apoptosis, and mitochondrial dysfunction in hearing loss. *Biomed Res Int* 2015;2015:617207

22. Istrate M, Hasbei-Popa M, Iliescu DA, Ghita AC, Ghita AM. Effects of cigarette smoking on sensorineural hearing impairment and age related macular degeneration. *Tob Prev Cessat* 2021;7:55
23. Tao L, Davis R, Heyer N, Yang Q, Qiu W, Zhu L, *et al.* Effect of cigarette smoking on noise-induced hearing loss in workers exposed to occupational noise in China. *Noise Health* 2013;15:67-72
24. Farhadian M, Aliabadi M, Darvishi E. Empirical estimation of the 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
25. Yankaskas K. Prelude: noise-induced tinnitus and hearing loss in the military. *Hear Res* 2013;295:3–8
26. Zare S, Ghotbi-Ravandi MR, ElahiShirvan H, Ahsaee MG, Rostami M. Predicting and Weighting the Factors Affecting Workers' Hearing Loss Based on Audiometric Data Using C5 Algorithm. *Ann Glob Health* 2019;85:88

Table 1. Pure tone hearing thresholds (Mean±Sd.) for the right and left ears according to frequencies (*N*=200).

Frequency	**with HL**	**without HL**
250 Hz		
Left (dB)	17.26±7.55	12.67±5.70
Right (dB)	18.36±7.78	12.40±4.40
500 Hz		
Left (dB)	18.36±9.09	11.23±4.14
Right (dB)	18.52±8.52	11.48±4.21
1000 Hz		
Left (dB)	18.27±10.42	9.81±4.83
Right (dB)	17.81±10.20	9.25±5.13
2000 Hz		
Left (dB)	23.41±17.44	10.18±6.29
Right (dB)	21.42±15.10	8.02±4.78
4000 Hz		
Left (dB)	55.50±18.66	21.23±8.85
Right (dB)	52.31±20.47	21.11±11.56
6000 Hz		

Left (dB)	51.42±23.08	20.61±11.49
Right (dB)	51.68±24.29	21.79±13.63
Pure Tone Avarage		
Left (dB)	28.88±10.00	13.11±3.46
Right (dB)	27.51±10.13	12.46±3.86

Note: HL= hearing loss

https://doi.org/10.1017/S002221512400094X Published online by Cambridge University Press

Table 2. Conditions that are risk factors for NIHL (*N*=200).

Risk Factors	with HL	without HL	*p*
Age (years)	48.0 (19.0-57.0)	30.0 (19.0-50.0)	<0.001[a]
Working duration (years)	16.0 (1.0-37.0)	5.0 (1.0-30.0)	<0.001[a]
Smoking (years)	10.0 (0-37.0)	0.5 (0-23.0)	0.002[a]
Using hearing protection apparatus (n)			<0.001[b]
Never	64 (53.8%)	32 (39.5%)	
Sometimes	45 (37.8%)	24 (29.6%)	
Continuously	10 (8.4%)	25 (30.9%)	
Tinnitus (n)	61 (%51.3)	3 (3.7%)	<0.001[b]

Note: HL= hearing loss, a= Mann Whitney-U test, b= χ^2

https://doi.org/10.1017/S002221512400094X Published online by Cambridge University Press

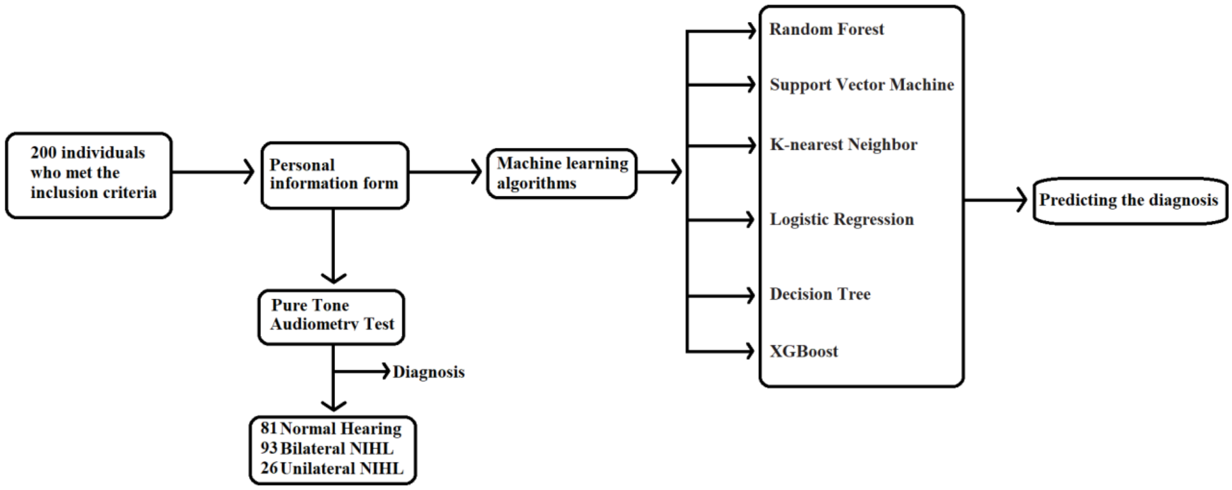
Table 3. Test performances of logistic regression, random forest, support vector machine, decision tree, K-nearest neighbors and XGBoost.

Models	Precision	Recall	F1-score	Accuracy	AUC- ROC
Logistic Regression	0.91	0.83	0.87	0.85	0.885
Random Forest	0.91	0.83	0.87	0.85	0.863
Support Vector Machine	0.95	0.88	0.91	0.90	0.906
Decision Tree	1.00	0.79	0.89	0.875	0.848
K-Nearest Neighbors	0.86	0.79	0.83	0.80	0.804
XGBoost	0.84	0.88	0.86	0.825	0.899

Note: AUC-ROC= Area under the receiver operating characteristic curve

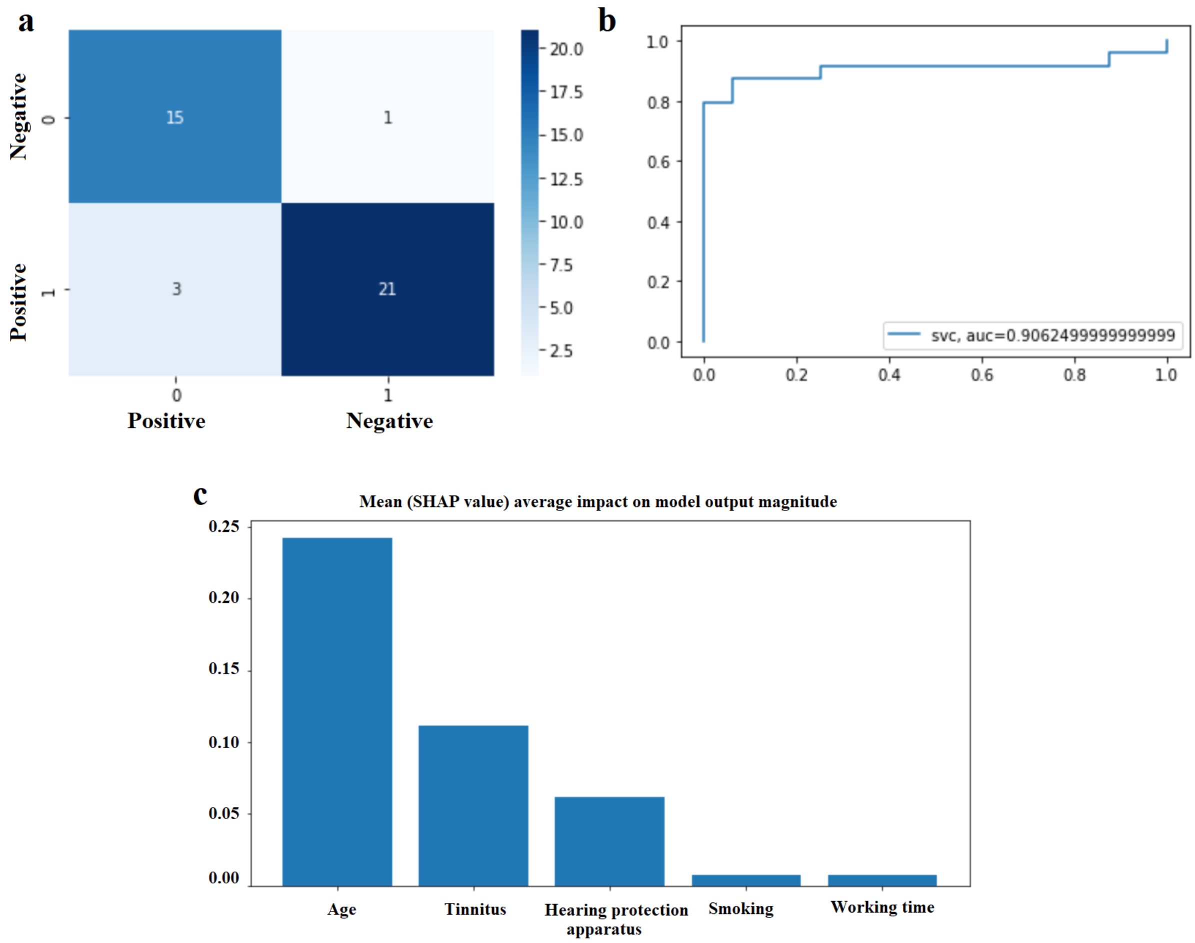
https://doi.org/10.1017/S002221512400094X Published online by Cambridge University Press

Figure 1. Flowchart detailing test steps and machine learning algorithms applied to workers. Two hundred workers were included in the study. A personal information form was applied to these workers. Questions in the fact sheet included risk factors for noise-induced hearing loss; age (years), working duration in noisy environments (years), using hearing protection apparatus (never, sometimes, continuously), smoking status (yes-no, if yes, how many years) and tinnitus (right, left or bilateral).



https://doi.org/10.1017/S002221512400094X Published online by Cambridge University Press

Figure 2. Figure 2a: The confusion matrix of the support vector machine model. Figure 2b: ROC curve of the support vector machine model. Figure 2c: SHAP analysis of the support vector machine



https://doi.org/10.1017/S002221512400094X Published online by Cambridge University Press