Machine Learning Models for the Hearing Impairment Prediction in Workers Exposed to Complex Industrial Noise: A Pilot Study

Yanxia Zhao,[1] Jingsong Li,[1] Meibian Zhang,[2] Yao Lu,[1] Hongwei Xie,[2] Yu Tian,[1] Wei Qiu[3]

Objectives: To demonstrate the feasibility of developing machine learning models for the prediction of hearing impairment in humans exposed to complex non-Gaussian industrial noise.

Design: Audiometric and noise exposure data were collected on a population of screened workers (N = 1,113) from 17 factories located in Zhejiang province, China. All the subjects were exposed to complex noise. Each subject was given an otologic examination to determine their pure-tone hearing threshold levels and had their personal full-shift noise recorded. For each subject, the hearing loss was evaluated according to the hearing impairment definition of the National Institute for Occupational Safety and Health. Age, exposure duration, equivalent A-weighted SPL (L_{Aeq}), and median kurtosis were used as the input for four machine learning algorithms, that is, support vector machine, neural network multilayer perceptron, random forest, and adaptive boosting. Both classification and regression models were developed to predict noise-induced hearing loss applying these four machine learning algorithms. Two indexes, area under the curve and prediction accuracy, were used to assess the performances of the classification models for predicting hearing impairment of workers. Root mean square error was used to quantify the prediction performance of the regression models.

Results: A prediction accuracy between 78.6 and 80.1% indicated that the four classification models could be useful tools to assess noise-induced hearing impairment of workers exposed to various complex occupational noises. A comprehensive evaluation using both the area under the curve and prediction accuracy showed that the support vector machine model achieved the best score and thus should be selected as the tool with the highest potential for predicting hearing impairment from the occupational noise exposures in this study. The root mean square error performance indicated that the four regression models could be used to predict noise-induced hearing loss quantitatively and the multilayer perceptron regression model had the best performance.

Conclusions: This pilot study demonstrated that machine learning algorithms are potential tools for the evaluation and prediction of noise-induced hearing impairment in workers exposed to diverse complex industrial noises.

Key words: Complex noise exposure, Hearing impairment, Machine learning, Noise-induced hearing loss.

(Ear & Hearing 2019;40;690–699)

[1]Key Laboratory for Biomedical Engineering of Ministry of Education, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China; [2]Institute of Environmental and Occupational Health, Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, China; and [3]Auditory Research Laboratory, State University of New York at Plattsburgh, New York, USA.

INTRODUCTION

Hearing loss is a public health problem, and hearing difficulty usually leads to long-term defects in language and cognitive development, understanding, behavior, and social adaptation (Barker et al. 2009). Worldwide, it was estimated that approximately 16% of disabling hearing loss in adults is due to occupational noise, ranging from 7 to 21% in various countries (Lie et al. 2016). Studies have indicated that noise-induced hearing loss (NIHL) is a complex disorder influenced by genetic and environmental factors (Henderson et al. 1993; Mizoue et al. 2003; Nomura et al. 2005; Konings et al. 2009; Zhao et al. 2010; Pelegrin et al. 2015).

In real industrial noise environments, workers are not only exposed to continuous Gaussian noise, but often exposed to complex non-Gaussian noise as well. Complex noise refers to a non-Gaussian noise consisting of a steady or nonsteady state Gaussian noise that is punctuated by a temporally complex series of randomly occurring high-level noise transients. These transients can be brief high-level noise bursts or impacts. Complex noise is very common in industry and the military. Experimental and epidemiological data from animal and human noise exposures indicate that the current noise exposure criterion (ISO-1999 2013) underestimates the amount of NIHL acquired by workers exposed to complex industrial noises (Guberan et al. 1971; Atherley 1973; Hamernik & Qiu 2001; Hamernik et al. 2003; Qiu et al. 2007; Davis et al. 2009; Zhao et al. 2010; Qiu et al. 2013; Xie et al. 2016). These results emphasize the inadequacy of current methods of measuring and evaluating noise exposures for the purpose of hearing conservation. When NIHL is evaluated, noise characteristics must be taken into account.

An important goal in industrial health/hazard management and in applied NIHL research is the construction of a reliable model to predict NIHL from a set of given noise exposure parameters. Building a prediction model is a regression problem, that is, an estimation of an unknown continuous function from a finite set of noisy samples. ISO-1999 (2013) and other predictive schemes have applied the so-called "first-principle model" (Cherkassky & Mulier 2007). This approach is a distribution-based statistical technique that tries to mechanically fit a prespecified function to some data set. Most NIHL research results are based on a first-principle approach. This approach works well when the number of samples is large relative to the model complexity. However, even when the number of samples is very large, this method can produce a large bias if the model complexity is incorrectly selected or the system under study is too complex to be mathematically described. The prediction of NIHL under various noise exposure conditions is just such a problem.

<zdoi; 10.1097/AUD.0000000000000649>

New statistical approaches such as machine learning algorithms, which can develop models from large, complex, and information-rich data sets, are available and have been successfully used to solve complex nonlinear problems in science and engineering. Among the most attractive features of such approaches is that they are distribution-free models that learn directly from the input data. Machine learning algorithms could generate rules from data automatically and predict unknown data. Machine learning has been widely used and shown to be highly effective in predicting nonlinear and fuzzy information regarding complex issues (Chan & Paelinckx 2008), such as spam detection (Drucker et al. 1999), credit card fraud (Syeda, Reference Note 1), speech recognition (Graves et al. 2013), and face recognition (Mian et al. 2007). Therefore, they may be extremely useful in predicting the outcome of an exposure from the appropriate input parameters. However, little work has been done to predict NIHL from workers exposed to diverse complex industrial noises using machine learning algorithms.

In this study, four potential machine learning models based on a support vector machine (SVM) (Burges 1998), multilayer perceptron (MLP) (Basheer & Hajmeer 2000), adaptive boosting (Adaboost) (Settouti et al. 2016), and random forest (RF) (Breiman 2001) algorithms were investigated for predicting the hearing impairment in workers. A database with subjects (N = 1,113) exposed to a diverse set of complex noise exposures was used in this investigation. The aim of this study was to demonstrate the feasibility of developing machine learning models for the prediction of hearing impairment in humans exposed to complex industrial noise using both classification and regression approaches.

MATERIALS AND METHODS

Data Collection

Subjects • Industrial workers were recruited from 17 factories in the Zhejiang province of China. Subjects (N = 1,644) were introduced to the purpose and procedures to be followed in this study by an occupational physician and were asked to sign an informed consent form. The Institutional Review Boards for the protection of human subjects of the Zhejiang Provincial Center for Disease Control and Prevention approved the protocol for this study.

Questionnaire Survey • An occupational hygienist from the Zhejiang Provincial Center for Disease Control and Prevention administered a questionnaire to each subject in order to collect the following information: general personal information (age, sex, etc.); occupational history (factory, worksite, job description, length of employment, duration of daily noise exposure, and history of using hearing protection); personal life habits (e.g., smoking and alcohol use); and overall health (including history of ear disease and use of ototoxic drugs). Before the data collection, the hygienist interviewed the administrators of the investigated factories to verify that the working environment remained constant. An occupational physician entered all information into a database.

Noise Data Collection • Shift-long noise recording files were obtained for each noise-exposed subject using an ASV5910-R digital recorder (Hangzhou Aihua Instruments Co.). The ASV5910-R digital recorder is a specialized sound recording instrument that can be used for precision measurements and analysis of personal noise exposure. The instrument uses a 1/4-inch prepolarized condenser microphone having good stability, high upper measurement limit, and wide frequency response (20 Hz–20 kHz). The sensitivity level of the microphone is

−53 dB, and the measurement range is 40 to 141 dBA (the peak value upper limit is 144 dB). The ASV5910-R runs a self-calibration program each time the power is turned on and can continuously run for over 20 hr. The shift-long noise for each subject was continuously recorded by the ASV5910-R at 32-bit resolution at a 48-kHz sampling rate, saved in a 32 GB micro SD card and downloaded to a computer for subsequent analysis. The recorder weighs 85 g and was mounted on the shoulder of the subject using special clips.

Physical and Audiometric Evaluation • Each subject underwent a general physical and an otologic examination. Otoscopic and tympanometric screening were conducted to each subject to rule out possible conductive hearing loss. Pure-tone hearing threshold levels (HTLs) at 0.5, 1.0, 2.0, 3.0, 4.0, 6.0, and 8.0 kHz were measured in each ear by an experienced physician. The testing was conducted in an audiometric booth (baseline noise <30 dB SPL) using an audiometer (Madsen, OB40) calibrated according to the Chinese national standard (GB4854-84). Audiograms were measured at least 16 hr after the subject's last occupational noise exposure.

Definition of Hearing Impairment

In this study, the National Institute for Occupational Safety and Health (NIOSH) material hearing impairment definition was used, that is, the average HTLs at 1, 2, 3, and 4 kHz for the better ear exceeds 25 dB HL.

Data Analysis

Data sets were collected from 1,644 workers exposed to complex noises in 53 workshops of 17 factories in the Zhejiang province of China.

Feature Selection • Feature variables related to noise-induced hearing impairment were collected from three different data sources: (1) questionnaires, (2) shift-long noise records, and (3) audiogram tests. Features from the questionnaire survey included age, gender, the name of factory, the name of workshop, type of work in production, length of service, and daily noise exposure time.

Recent results from animal experiments (Hamernik et al. 2003; Qiu et al. 2006, 2007, 2013) and epidemiological studies (Zhao et al. 2010; Xie et al. 2016) have shown that, in addition to energy, a statistical metric of the noise amplitude distribution, the kurtosis, could order the extent of hearing and sensory cell loss from a variety of complex noise exposures. Thus, there is the possibility that the kurtosis, in combination with the energy, might be useful in evaluating a broad range of noise environments for hearing conservation purposes. The statistical metric, kurtosis, was defined as a ratio of the fourth-order central moment to the squared second-order moment of the amplitude distribution. Note that the kurtosis of a Gaussian distribution is 3. In this study, a median kurtosis of 4 or greater was used to identify a complex non-Gaussian noise exposure.

The kurtosis of the noise was computed over consecutive 40-s time windows of each shift-long noise record using MATLAB software (version 2015b, MathWorks), and the mean and median values were used to establish the kurtosis values for each noise record. The 8-hr equivalent A-weighted SPL ($L_{Aeq.8h}$), the most important feature of the noise, was obtained from shift-long noise record as well.

Features from the questionnaire survey and the noise exposure were used as input variables. For the classification model,

TABLE 1. Gender distribution of the hearing loss of all subjects in the study cohort

Qualitative Variables	Category	Hearing Loss [n (%)]	Normal [n (%)]	Total [n (%)]
Gender	Female	47 (15.1)	264 (84.9)	311
	Male	174 (21.7)	628 (78.3)	802
Total number	—	221 (19.8)	892 (80.2)	1113

the target variable was defined as a binary classification variable representing whether the average of HTLs of the better ear at 1 to 4 kHz exceeds 25 dB HL, and for the regression model, the target variable was a continuous variable defined as the mean age- and gender-adjusted HTLs at 1, 2, 3, and 4 kHz.

Data Inclusion • For inclusion in the study, all subjects had to satisfy six criteria: (1) a minimum of at least 1 year employment at their current task; (2) consistently worked within the same job category and worksite (noise exposure area) for their entire career; (3) no history of genetic or drug-related hearing loss, head wounds, or ear diseases; (4) no history of military service or shooting activities; (5) no history of using hearing protection; and (6) exposed to complex non-Gaussian noise. It was confirmed by

both the answers from questionnaires and the observation of a hygienist that the workers investigated rarely, if ever, used hearing protection device. Accordingly, a total of 1,113 workers were included from the original pool of 1,664 subjects. Of the included workers, 221 were hearing impaired and 892 had normal hearing, as shown in Table 1. Table 1 summarizes the descriptive statistical information of the hearing loss of all workers in the study. Table 2 provides a breakdown of average noise exposure level, duration of exposure, kurtosis, age, and sex, corresponding to the number of subjects exposed in each plant.

Figure 1 shows scatter plots of the relation between the quantitative independent variables and the average hearing thresholds of the better ear at 1 to 4 kHz for the workers included in the study. It can be seen from Figure 1A, C that the age and equivalent A-weighted SPL (L_{Aeq}) of the noise approximately obeyed normal distributions. Most workers were between 30 and 50 years old, and the exposure duration was between 1 and 20 years. The values of mean and median kurtosis were scattered in Figure 1D, E. It was noticed that the median kurtosis values were less than 25 and the mean kurtosis values were less than 50 for most workers. The average HTLs of the better ear at 1 to 4 kHz for most workers were between 10 and 30 dB HL. It can be seen from Figure 1C that most of the subjects were exposed to complex noise with the

TABLE 2. Descriptive statistical information of characteristic of workers of each factory in the study cohort

| Factory	Male (n)	Female (n)	Age (yr)	Duration (yr)	L_{Aeq} (dBA)	Median Kurtosis
Machinery plant	10	15	22–44	1–21	83.01–93.60	4.43–19.90
			36.16±5.09	12.80±7.31	88.22±2.62	11.04±4.44
Steel net rack plant	49	0	23–49	1–16	87.28–109.13	4.49–91.20
			36.61±6.03	6.02±3.91	94.98±4.14	27.25±20.56
Kitchen and bath factory	7	15	24–60	1–19	68.34–89.00	5.55–33.57
			43.45±9.63	8.32±5.30	80.46±4.82	18.36±8.55
Assembly plant	130	66	23–47	1–30	69.77–106.45	3.92–118.90
			35.86±4.44	14.59±5.36	90.79±4.96	12.55±12.24
Heavy truck engine factory	40	2	19–60	1–33	81.84–113.03	3.93–22.09
			31.67±11.01	8.40±8.83	89.87±5.83	7.33±4.41
Pipeline factory	37	2	20–52	1–5	80.47–97.90	4.10–38.14
			27.97±6.94	3.15±1.39	90.07±3.79	18.21±9.70
Auto fixture fastener plant	166	15	21–55	1–36	76.00–103.20	3.91–182.55
			38.94±6.81	18.83±7.36	90.08±5.16	14.15±18.09
Machinery and electric co.	38	112	22–50	1–29	77.68–98.03	4.72–145.13
			36.00±7.34	7.76±5.38	85.79±3.40	20.08±19.60
Pipe factory	32	3	21–56	1–35	73.85–92.95	4.21–23.64
			36.20±10.40	9.29±8.19	84.60±5.15	10.15±4.43
Woven bag factory	13	11	24–46	1–20	68.90–96.33	3.99–13.03
			38.88±5.21	10.00±3.55	87.94±6.93	5.99±2.27
Steel plant	71	0	20–53	1–33	74.15–99.83	3.91–18.03
			39.99±7.74	15.01±7.77	88.66±5.82	7.62±3.74
Vehicle frame plant	46	16	27–51	3–34	77.29–104.85	3.93–37.60
			39.21±6.15	17.86±7.47	90.97±5.65	10.83±7.02
Hardware factory	20	9	36–57	4–36	88.10–105.91	3.95–55.10
			46.41±5.46	20.69±8.51	96.69±4.27	11.65±12.17
Cement plant	28	7	22–57	2–37	73.05–90.33	4.01–25.90
			44.46±7.77	19.69±10.15	80.68±4.76	8.59±4.49
Automobile wheel Co.	30	21	30–48	2–29	81.75–102.12	3.95–40.05
			36.45±4.18	14.41±5.24	89.79±4.24	12.08±7.76
Vehicle body factory	41	16	24–49	1–30	80.71–99.17	3.95–55.66
			38.36±5.35	17.18±6.12	89.41±4.30	12.40±9.28
Steel rolling mill	44	0	22–52	2–30	92.27–109.20	4.00–20.74
			39.86±6.30	18.41±5.78	99.11±3.66	8.33±4.10

L_{Aeq}, equivalent A-weighted SPL.
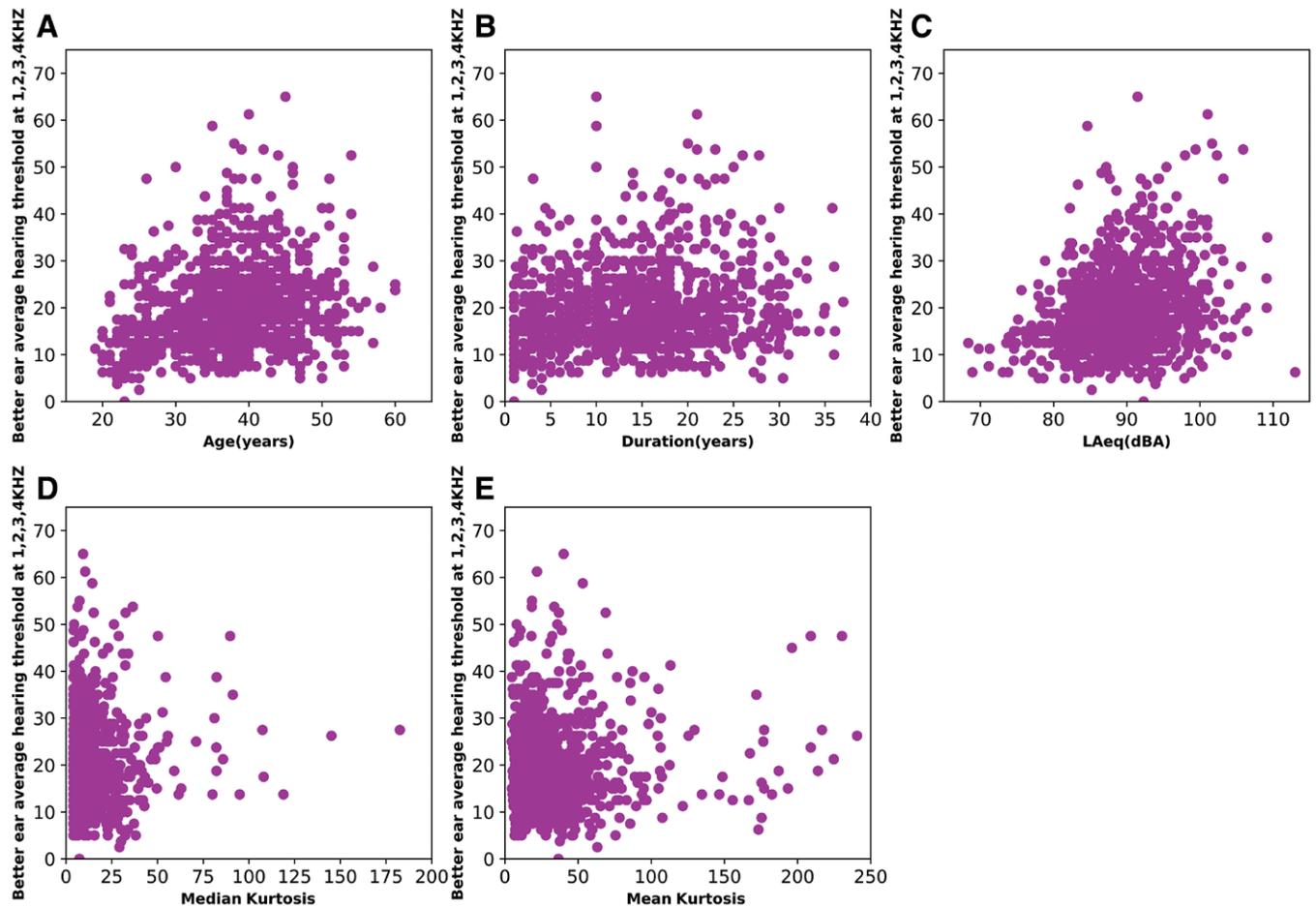
Fig. 1. Scatter plots of the relation between the various quantitative variables and the average of hearing threshold levels of the better ear at 1 to 4 kHz ($HTL_{1–4kHz}$) for all workers in the study cohort. A, Scatter plot of age vs. $HTL_{1–4kHz}$. B, Scatter plot of exposure duration vs. $HTL_{1–4kHz}$. C, Scatter plot of equivalent A-weighted SPL (L_{Aeq}) vs. $HTL_{1–4kHz}$. D, Scatter plot of median kurtosis vs. $HTL_{1–4kHz}$. E, Scatter plot of median kurtosis vs. $HTL_{1–4kHz}$.

L_{Aeq} >85 dBA, which exceeded the recommended exposure limit by NIOSH, that is, 85 dB, A-weighting, as 8-hr time-weighted average (NIOSH 1998).

Model Construction

Four potential machine learning algorithms were investigated in this study to build classification models for predicting hearing impairment and to build regression models for predicting NIHL quantitatively. Only the architecture of four machine learning classifiers is presented in this paper. The details of their classification and regression algorithms can be found in the references listed below. To train and validate the algorithms, 10 experimental runs were conducted through 10-fold cross-validation.

1. RF model: The RF classifier uses the bootstrap resampling technique to generate N new training sample sets. The new sample sets then are used to train N decision trees. The results of classification depend on the score formed by N decision trees voting. In general, an RF classifier consists of two modules: a training module and a classification module. The overall architecture is shown in Figure 2 (Breiman 2001).
2. Adaboost model: The Adaboost classifier is a highly accurate classifier that takes the single-layer decision tree as the base classification algorithm, training a number of

weak learners based on the weight update for the same training set and finally combining these weak learners through weighted fusion to obtain the final strong classifier. The Adaboost classifier architecture is shown in Figure 3 (Freund & Schapire 1997; Nasrabadi 2007; Witten et al. 2016).

3. MLP model: The MLP classifier is a feed-forward neural network. It utilizes a supervised learning technique called backpropagation for training. The algorithm has the advantage of being able to approximate any nonlinear function. In general, an MLP model consists of an input layer, at least one hidden layer, and an output layer. A schematic diagram of the MLP classifier is shown in Figure 4 (Basheer & Hajmeer 2000).
4. SVM model: The SVM classifier is based on establishing a separating hyperplane having maximum distance from the closest points of the training set. This has the advantage of having a strong ability to generalize and can perform very well in solving nonlinear and high-dimensional pattern recognition problems. The principle of an SVM classifier for linearly separable data is shown in Figure 5. For nonlinear problems, it is necessary to choose an appropriate kernel function. A polynomial function was used in this study (Burges 1998; Pontil & Verri 1998; Nasrabadi 2007).
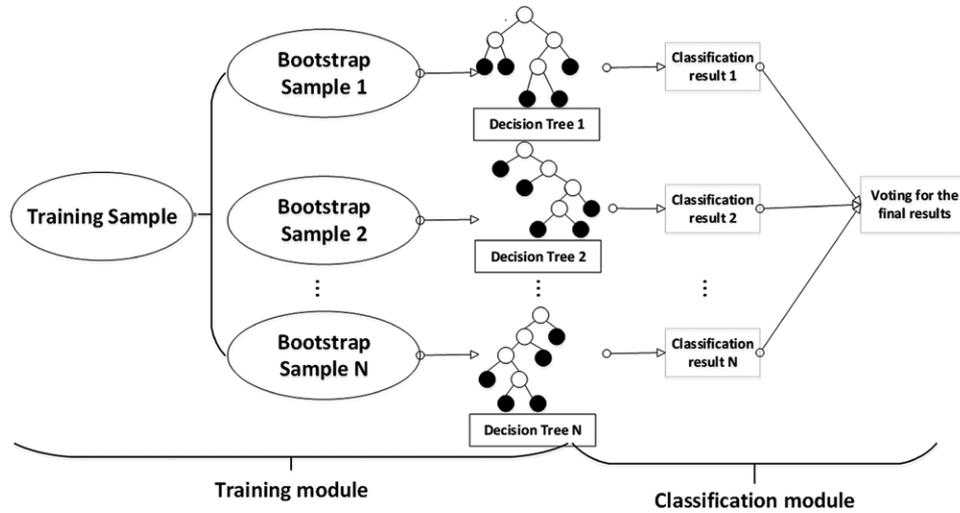
Fig. 2. Schematic diagram of the random forest (RF) classifier consisting of a training module and a classification module.

Risk Factor Identification

The most predictive features for hearing impairment were identified by feature selection. The selected data values were normalized by min-max normalization technique, and then the p value of each variable was calculated using a t test. Finally, the features for building the SVM, Adaboost, and MLP models were selected based on a significance level of $p <0.01$. The RF algorithm can handle high-dimensional data and does not require features to be preselected by a t test.

Indices for Model Evaluation

In this study, the indices of (1) the area under the receiver operating characteristic (ROC) curve (AUC) and (2) prediction accuracy were used to measure the performance of the classifiers. A root mean square error (RMSE) was used to evaluate the performance of the regression models. The RMSE is defined as follows:

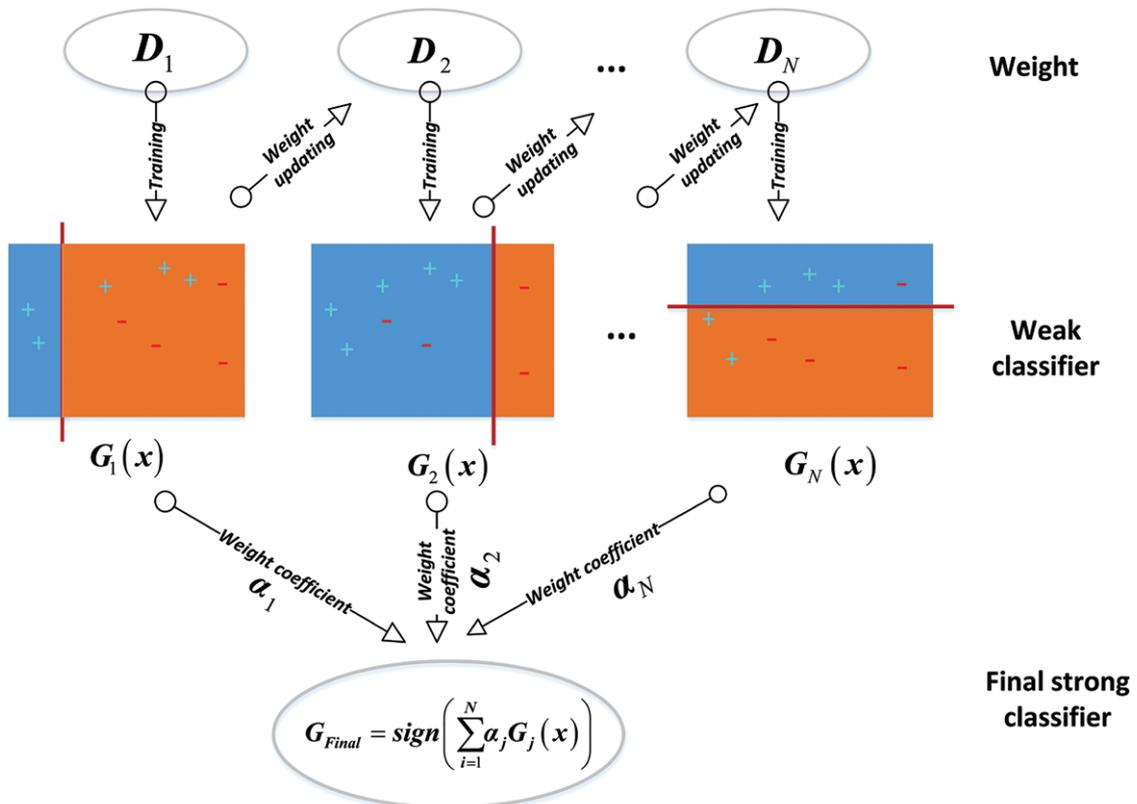$$RMSE = \sqrt{\frac{\sum_{j=1}^{n}\left(T_j - Y_j\right)^2}{n}} \tag{1}$$

Fig. 3. Schematic diagram of the adaptive boosting (Adaboost) classifier. Plus and minus signs represent two classes, respectively.

Input Layer **Hidden Layer** **Output Layer**
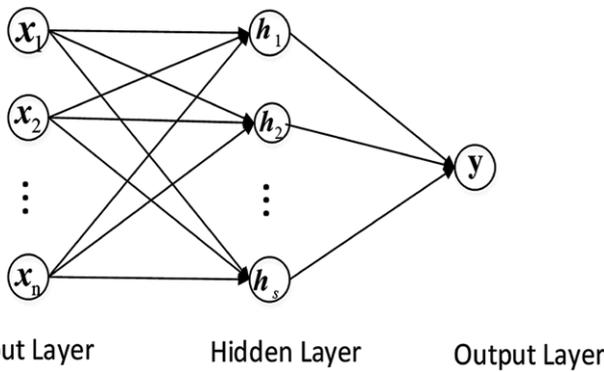
Fig. 4. Schematic diagram of the multilayer perceptron (MLP) classification model. Here, x represents sample input, h represents hidden layer, and y represents output layer. The transfer function of the hidden layer vector is $h = f\left(w^1 x + b^1\right)$, and the output layer neurons is $y = f\left(w^2 h + b^2\right)$, where w^1 is the weight matrix between x and h, b^1 is the bias vector of h, w^2 is the weight matrix between h and y, and b^2 denotes the bias vector of y. The activation function is $f(x) = \max(0, x)$.

where T_j and Y_j are the measured and the predicted values of data j, respectively, and n represents the number of measurements. The RMSE indicates the prediction accuracy of regression model. The lower the value of RMSE, the better predictive ability of the regression model.

RESULTS

Risk Factor Identification

It can be seen from Figure 1 that some of the risk factors appear to be related to hearing loss. Analysis of normalized values by t test revealed that age, exposure duration, L_{Aeq}, and median kurtosis were significantly associated with hearing loss ($p < 0.01$) (Table 3). These variables were used for building the three of the prediction models; as noted above the RF model does not require preselection of variables.
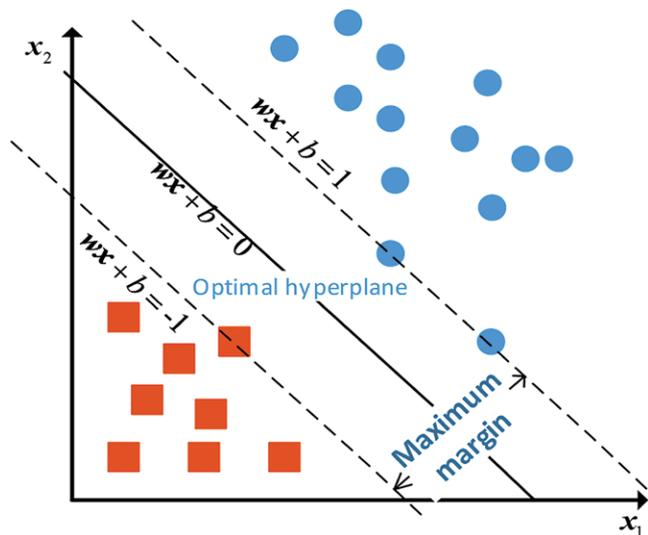

Fig. 5. A linear hyperplane learned by training support vector machine (SVM) in two dimensions (D = 2). Squares and circles represent class −1 and class 1, respectively. $wx + b = 0$ represents separating hyperplane, $wx + b = 1$ and $wx + b = -1$ represent margin boundaries, where w and b are the unknown weights and bias, x represents sample input.

TABLE 3. p values in the study cohort that reflect the relationship between the constructed feature variables and hearing impairment

Feature Variables	Gender	Age	Duration	L_{Aeq}	Mean Kurtosis	Median Kurtosis
p value	0.013	1.6×10^{-6}	5.0×10^{-7}	5.7×10^{-12}	0.011	9.7×10^{-4}

L_{Aeq}, equivalent A-weighted SPL.

Performance Evaluation of Different Hearing Impairment Prediction Models

In this study, 10-fold cross-validation was used to conduct the modeling and evaluation for both classification and regression analysis of hearing impairment. Each model was built by calling the Scikit-learn library (Pedregosa et al. 2011) which is one of the well-developed machine learning libraries based on NumPy, SciPy, and Matplotlib. The models were optimized using the grid search method by fine-tuning the parameters of models in the grid (Woodford & Phillips 1997).

The performance of Classification Models • The average ROC curve and the average AUC value of the cross-validation for all models are shown in Figure 6A. It can be reasonably assumed that the models did not over fit the data because the ROC curve corresponding to each model was smooth. However, the ROC curves could not directly reveal which classifier is the best one. Instead, the classifier corresponding to the largest AUC can be considered to have the best function. Thus, the AUC value was used to evaluate the performance of each model in this study. As shown in Figure 6B, the SVM model had the best performance compared with the other models, with an AUC value of 0.808.

In addition, the performance of the machine learning classifier could also be evaluated by the prediction accuracy. The classification accuracies of four models are shown in Figure 7. It can be seen from Figure 7 that SVM had a slightly higher prediction accuracy than the other three models, though the predictive abilities of four models were not significantly different.

In real applications of machine learning and data mining, the AUC is usually used as the single value for evaluating the predictive performance of a model. The study of Huang and Ling indicated that the AUC is a better measure than accuracy and should replace accuracy when comparing the performance of machine learning algorithms (Huang & Ling 2005). In the present study, both AUC and accuracy indexes showed that the SVM model was the best hearing impairment prediction model.

The Performance of Regression Models • NIHL can be quantitatively predicted by using the regression models with machine learning algorithms. In this study, Adaboost, RF, MLP, and SVM regression models were built to predict NIHL using the features of L_{Aeq}, exposure duration, and median kurtosis. Measured HTL at each frequency was adjusted by subtracting the median age- and gender-specific HTL from Annex A (ISO-1999 2013), and then the mean age- and gender-adjusted HTLs at 1, 2, 3, and 4 kHz were calculated and used as the definition of NIHL. The RMSE was used to quantify the prediction performances of the regression models. The prediction performances of the four regression models are shown in Table 4. The MLP regression model had the best performance because its RMSE value was the lowest. Therefore, the MLP model was chosen
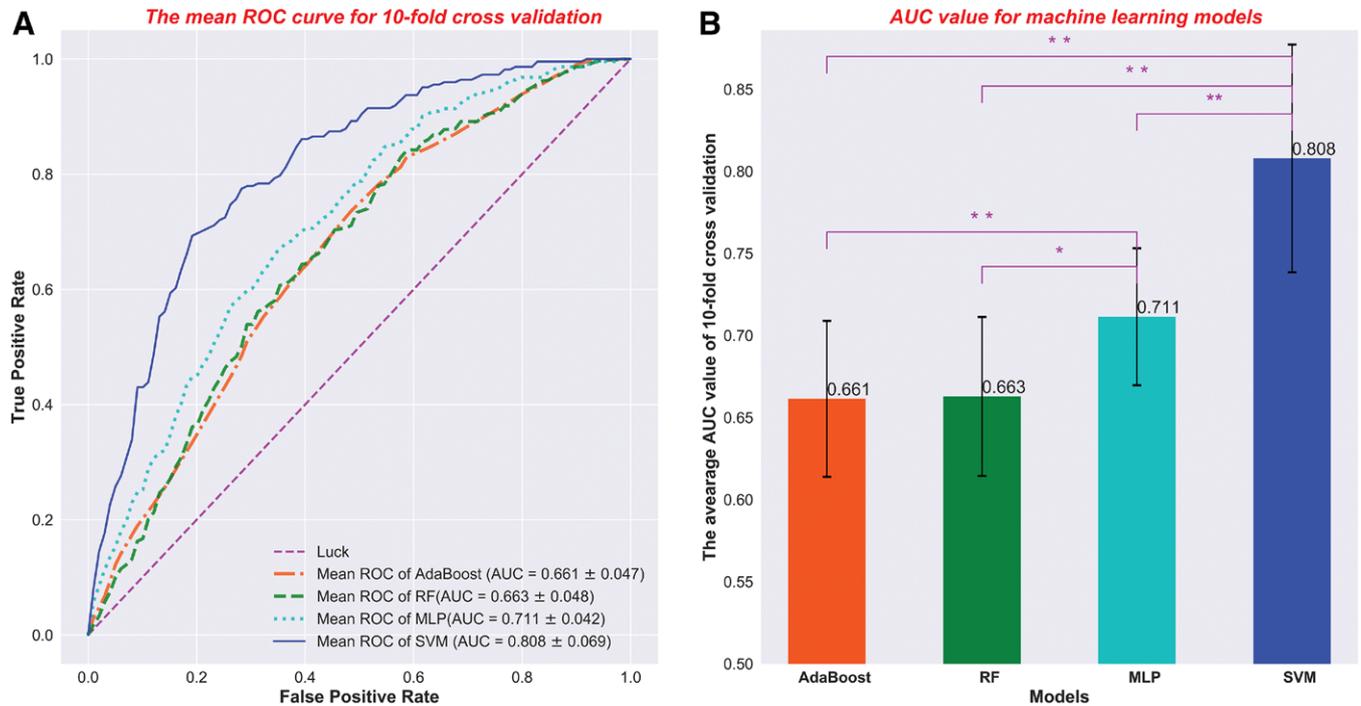
Fig. 6. Hearing impairment predictive performance of four machine learning classification models. Training and testing of all models were conducted with 10-fold cross-validation. A, The mean receiver operating characteristic (ROC) curves of 10-fold cross-validation for four machine learning classification models. B, The average area under the curve (AUC) value of 10-fold cross-validation for four machine learning classification models. Wilcoxon signed-rank test was used to compute p values for AUC of two classifiers using 10-fold cross-validation. The error bars represent the standard deviations of the 10-fold cross-validation. $*p < 0.01$, $**p < 0.001$.

to demonstrate the prediction ability of the developed machine learning regression models. The results are shown in Table 5. Because all audiometric data in this study were collected with a step size of 5 dB using the ascending technique, it is reasonable to assume that the prediction was correct if the difference between predicted and measured NIHL values was less than or equal to 2.5 dB. Comparing the mean predicted NIHL with the measured NIHL from 17 factories in Table 5, it can be seen in Table 5 that the MLP regression model could predict NIHL in all but three factories, that is, the woven bag factory, the hardware factory, and the cement factory.

DISCUSSION

Although machine learning algorithms have been widely applied, little work has been done to predict NIHL from workers exposed to diverse complex industrial noises using these algorithms. Aliabadi et al. (2015) and Farhadian et al. (2015) reported their work of developing NIHL prediction models using neural network algorithms. Using a database with 210 subjects from a steel factory, their results showed that the performance of prediction models was satisfactory. However, it may not be suitable to show the feasibility of the prediction models with such a small size of sample. In this study, four hearing
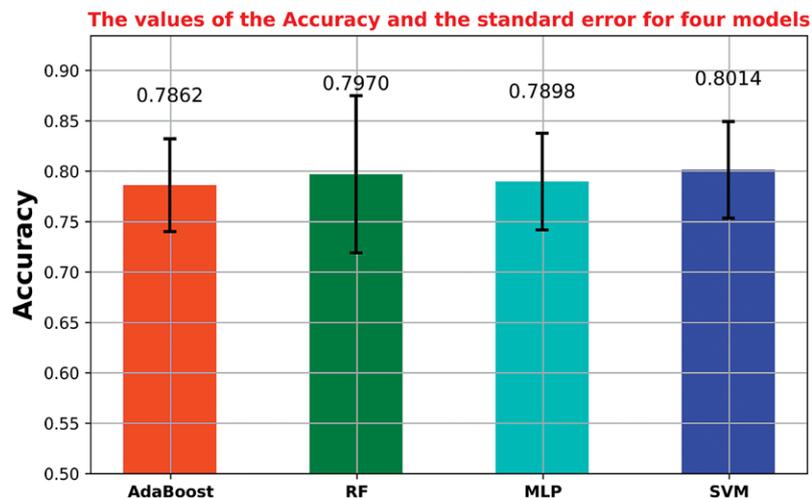

Fig. 7. Classification accuracy of various machine learning prediction models. Adaboost, adaptive boosting; MLP, multilayer perceptron; RF, random forest; SVM, support vector machine.

TABLE 4. RMSE of four regression models with 10-fold cross-validation

Performance Metric	Adaboost	RF	MLP	SVM
Mean RMSE (dB)	2.894	2.858	2.727	2.942

MLP, multilayer perceptron; RF, random forest; RMSE, root mean square error; SVM, support vector machine.

prediction models based on machine learning algorithms, using a noise database with 1,113 subjects from 17 factories, were investigated.

A univariate feature selection method, the *t* test method, was used in the present study. The *p* value can be used to rank the input variables according to the degree of correlation with the output variable. It can be seen in Table 3 that the *p* values of L_{Aeq} and duration were the smallest. It is not surprising that the equivalent A-weighted SPL and exposure duration are the most important risk factors for noise-induced hearing impairment. Because the definition of hearing impairment by NIOSH is not age and gender adjusted, age and gender were also considered as the risk factors. The kurtosis was used in this study to quantify the impulsiveness of the noise exposure. Because kurtosis was calculated over consecutive 40-s time windows of each shift-long noise record, there were 720 kurtosis values for an 8-hr noise record. The two metrics, mean and median kurtosis, were selected as candidates for risk factors. It can be seen in Table 3 that the importance of all the feature variables was ranked as follows: L_{Aeq}, duration, age, median kurtosis, gender, and mean kurtosis. Based on a significance level selected ($p < 0.01$), four risk factors, that is, L_{Aeq}, duration, age, and median kurtosis, were used in building the noise-induced hearing impairment prediction models.

Despite the large variation of worker's ages, noise levels, and exposure durations as shown in Figure 1, a prediction accuracy between 78.6 and 80.1% indicated that the four classifiers could be useful tools to assess hearing impairment of workers exposed to various complex occupational noises. A comprehensive evaluation using both the AUC and prediction accuracy showed that the SVM model achieved the best score in classification and thus should be selected as the best tool for predicting hearing impairment from occupational noise exposures in this study.

Although this work was based on the NIOSH hearing impairment definition at 1 to 4 kHz, the approaches mentioned above are universal and could be used to predict hearing impairment under different definitions. For example, applying the models to predict hearing impairment based on the Occupational Safety and Health Administration definition, that is, the average HTL of both ears exceeding 25 dB at 1, 2, and 3 kHz, prediction accuracies between 76.6 and 83% were achieved as shown in Table 6. The results in Table 6 verified that the SVM classification model produced the best performance with the highest prediction accuracy of 83%.

The results displayed in Table 5 indicated that the MLP regression model could correctly predict the mean NIHL except in three factories marked with an asterisk. Evaluating the noise data from these three failed factories, the sample sizes of these factories were small, indicating that the regression model might not be well trained with small size samples. Among these three failed factories, the cement factory showed the largest prediction error (6 dB). It was noticed that the workers of the cement factory were working in open fields, while the workers of other factories worked in closed workshops. Unlike the subjects working at a fixed position in a closed workshop, the working position for most workers in the cement factory was mobile. Thus, one noise recording for these workers may not be suitable to approximate the noise exposure over the lifetime of work.

In this pilot study, four machine learning algorithms were used to build both classification and regression models for noise-induced hearing impairment prediction. Although the SVM and MLP algorithms demonstrated their appropriateness for classification and regression analyses, Adaboost and RF algorithms also performed well in solving classification and regression problems. As shown in Figure 7, Adaboost

TABLE 5. Average predicted NIHL of each factory using multilayer perceptron regression model

Factory	n	Duration (yr)	L_{Aeq} (dBA)	Median Kurtosis	Measured NIHL (dB)	Predicted NIHL (dB)
Machinery plant	25	12.80±7.31	88.22±2.62	11.04±4.44	17.13±7.93	18.21±5.90
Steel net rack plant	49	6.02±3.91	94.98±4.14	27.25±20.56	20.57±10.10	19.80±8.19
Kitchen and bath factory	22	8.32±5.30	80.46±4.82	18.36±8.55	15.14±6.80	16.93±5.79
Assembly plant	196	14.59±5.36	90.79±4.96	12.55±12.24	18.49±9.21	19.42±7.77
Heavy truck engine factory	42	8.40±8.83	89.87±5.83	7.33±4.41	18.70±8.71	17.42±6.03
Pipeline factory	39	3.15±1.39	90.07±3.79	18.21±9.70	11.47±5.27	12.73±4.98
Auto fixture fastener plant	181	18.83±7.36	90.08±5.16	14.15±18.09	17.92±9.07	17.80±8.73
Machinery and electric co.	150	7.76±5.38	85.79±3.40	20.08±19.60	19.10±7.68	18.00±6.98
Pipe factory	35	9.29±8.19	84.60±5.15	10.15±4.43	16.16±5.74	16.58±3.94
Woven bag factory*	24	10.00±3.55	87.94±6.93	5.99±2.27	21.87±12.10	18.12±7.68
Steel plant	71	15.01±7.77	88.66±5.82	7.62±3.74	18.87±5.89	17.53±4.52
Vehicle frame plant	62	17.86±7.47	90.97±5.65	10.83±7.02	16.97±8.64	17.99±7.47
Hardware factory*	29	20.69±8.51	96.69±4.27	11.65±12.17	25.78±11.34	22.70±8.73
Cement plant*	35	19.69±10.15	80.68±4.76	8.59±4.49	7.05±5.29	13.05±4.61
Automobile wheel Co.	51	14.41±5.24	89.79±4.24	12.08±7.76	16.47±6.76	18.95±4.32
Vehicle body factory	57	17.18±6.12	89.41±4.30	12.40±9.28	17.17±8.05	18.53±7.06
Steel rolling mill	44	18.41±5.78	99.11±3.66	8.33±4.10	22.54±10.87	20.56±6.76

Factory with prediction error larger than 2.5 dB.
L_{Aeq}, *equivalent A-weighted SPL; NIHL, noise-induced hearing loss.*

TABLE 6. Performance of four machine learning classification models based on Occupational Safety and Health Administration definition using 10-fold cross-validation.

Performance metric	Adaboost	RF	MLP	SVM
Mean AUC	0.655	0.664	0.693	0.776
Mean accuracy	76.6%	76.9%	81.9%	83.0%

AUC, area under the curve; MLP, multilayer perceptron; RF, random forest; SVM, support vector machine

and RF could achieve a similar classification accuracy as SVM and MLP. In the regression study, the Adaboost and RF even scored smaller RMSE than SVM. Thus, Adaboost and RF may also have potential value in predicting hearing impairment.

Though the classification and regression models using the four machine learning algorithms performed well in predicting hearing impairment, the performance could be further improved by acquiring more data from a larger number of subjects with well-documented and diverse exposures. In addition to the feature variables used in the above prediction models, more information relating to hearing loss, such as the octave band noise level, kurtosis in frequency domain, smoking history, blood pressure, and genes related to susceptibility of hearing loss, should be collected. Another possible method to enhance performance would be to develop a prediction model for exposure groups that share very similar noise characteristics, thus producing a more homogeneous group of cohorts. With more noise data being collected and more relevant risk factors being considered, NIHL evaluation using machine learning can be very useful to hearing conservation practices.

CONCLUSIONS

This pilot study demonstrated that machine learning algorithms are potential tools for the prediction of noise-induced hearing impairment in workers exposed to diverse complex industrial noises.

ACKNOWLEDGMENTS

REFERENCES

Aliabadi, M., Farhadian, M., Darvishi, E. (2015). Prediction of hearing loss among the noise-exposed workers in a steel factory using artificial intelligence approach. *Int Arch Occup Environ Health*, 88, 779–787.

Atherley, G. R. (1973). Noise-induced hearing loss: The energy principle for recurrent impact noise and noise exposure close to the recommended limits. *Ann Occup Hyg*, 16, 183–194.

Barker, D. H., Quittner, A. L., Fink, N. E., et al.; CDaCI Investigative Team. (2009). Predicting behavior problems in deaf and hearing children: The influences of language, attention, and parent-child communication. *Dev Psychopathol*, 21, 373–392.

Basheer, I. A., & Hajmeer, M. (2000). Artificial neural networks: Fundamentals, computing, design, and application. *J Microbiol Methods*, 43, 3–31.

Breiman, L. (2001). Random forests. *Mach Learn*, 45(1), 5–32.

Burges, C. J. C. (1998). A tutorial on Support Vector Machines for pattern recognition. *Data Min Knowl Disc*, 2(2), 121–167.

Chan, J. C. W., & Paelinckx, D. (2008). Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sens Environ*, 112(6), 2999–3011.

Cherkassky, V., & Mulier, F. M. (2007). *Learning From Data: Concepts, Theory, and Methods*. New York, NY: John Wiley & Sons.

Davis, R. I., Qiu, W., Hamernik, R. P. (2009). Role of the kurtosis statistic in evaluating complex noise exposures for the protection of hearing. *Ear Hear*, 30, 628–634.

Drucker, H., Wu, D., Vapnik, V. N. (1999). Support vector machines for spam categorization. *IEEE Trans Neural Netw*, 10, 1048–1054.

Farhadian, M., Aliabadi, M., Darvishi, E. (2015). 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*, 19, 84–89.

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci*, 55(1), 119–139.

Graves, A., Mohamed, A. R., Hinton, G. (2013). Speech recognition with deep recurrent neural networks. *Proc. IEEE Int Conf Acoust Speech Signal Process*, 1, 6645–6649.

Guberan, E., Fernandez, J., Cardinet, J., et al. (1971). Hazardous exposure to industrial impact noise: Persistent effect on hearing. *Ann Occup Hyg*, 14, 345–350.

Hamernik, R. P., & Qiu, W. (2001). Energy-independent factors influencing noise-induced hearing loss in the chinchilla model. *J Acoust Soc Am*, 110, 3163–3168.

Hamernik, R. P., Qiu, W., Davis, B. (2003). The effects of the amplitude distribution of equal energy exposures on noise-induced hearing loss: The kurtosis metric. *J Acoust Soc Am*, 114, 386–395.

Henderson, D., Subramaniam, M., Boettcher, F. A. (1993). Individual susceptibility to noise-induced hearing loss: An old topic revisited. *Ear Hear*, 14, 152–168.

Huang, J., & Ling, C. X. (2005). Using AUC and accuracy in evaluating learning algorithms. *Ieee T Knowl Data En*, 17(3), 299–310.

ISO-1999. (2013). *Estimation of noise-induced hearing loss.* Geneva, Switzerland: International Organization for Standardization.

Konings, A., Van Laer, L., Van Camp, G. (2009). Genetic studies on noise-induced hearing loss: A review. *Ear Hear*, 30, 151–159.

Lie, A., Skogstad, M., Johannessen, H. A., et al. (2016). Occupational noise exposure and hearing: A systematic review. *Int Arch Occup Environ Health*, 89, 351–372.

Mian, A. S., Bennamoun, M., Owens, R. (2007). An efficient multimodal 2D-3D hybrid approach to automatic face recognition. *Ieee T Pattern Anal*, 29(11), 1927–1943.

Mizoue, T., Miyamoto, T., Shimizu, T. (2003). Combined effect of smoking and occupational exposure to noise on hearing loss in steel factory workers. *Occup Environ Med*, 60, 56–59.

Nasrabadi, N. M. (2007). Pattern recognition and machine learning. *J Electron Imaging*, 16(4), 049901.

NIOSH. (1998). *NIOSH -- Criteria for a Recommended Standard: Occupational Noise Exposure.* Cincinnati, Ohio: National Institute for Occupational Safety and Health.

Nomura, K., Nakao, M., Yano, E. (2005). Hearing loss associated with smoking and occupational noise exposure in a Japanese metal working company. *Int Arch Occup Environ Health*, 78, 178–184.

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine Learning in Python. *J Mach Learn Res*, 12, 2825–2830.

Pelegrin, A. C., Canuet, L., Rodríguez, Á. A., et al. (2015). Predictive factors of occupational noise-induced hearing loss in Spanish workers: A prospective study. *Noise Health*, 17, 343–349.

Pontil, M., & Verri, A. (1998). Properties of support vector machines. *Neural Comput*, 10, 955–974.

Qiu, W., Davis, B., Hamernik, R. P. (2007). Hearing loss from interrupted, intermittent, and time varying Gaussian noise exposures: the applicability of the equal energy hypothesis. *J Acoust Soc Am*, 121, 1613–1620.

Qiu, W., Hamernik, R. P., Davis, B. (2006). The kurtosis metric as an adjunct to energy in the prediction of trauma from continuous, nonGaussian noise exposures. *J Acoust Soc Am*, 120, 3901–3906.

Qiu, W., Hamernik, R. P., Davis, R. I. (2013). The value of a kurtosis metric in estimating the hazard to hearing of complex industrial noise exposures. *J Acoust Soc Am*, 133, 2856–2866.

Settouti, N., Bechar, M. E., Chikh, M. A. (2016). Statistical comparisons of the top 10 algorithms in data mining for classification task. *Int J Interact Multi*, 4(1), 46–51.

Witten, I. H., Frank, E., Hall, M. A., et al. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco, CA: Morgan Kaufmann.

Woodford, C., & Phillips, C. (1997). *Numerical Methods With Worked Examples*. London, United Kingdom: Chapman and Hall.

Xie, H. W., Qiu, W., Heyer, N. J., et al. (2016). The use of the kurtosis-adjusted cumulative noise exposure metric in evaluating the hearing loss risk for complex noise. *Ear Hear*, 37, 312–323.

Zhao, Y. M., Qiu, W., Zeng, L., et al. (2010). Application of the kurtosis statistic to the evaluation of the risk of hearing loss in workers exposed to high-level complex noise. *Ear Hear*, 31, 527–532.

REFERENCE NOTE

Syeda, M., Zhang, Y. Q., Pan, Y. (2002). Parallel granular neural networks for fast credit card fraud detection. *Proc. IEEE Int Conf Fuzzy Systems*, 1(2), 572–577.