APRIL 26 2019

Development of an automatic classifier for the prediction of hearing impairment from industrial noise exposure

Yanxia Zhao; Yu Tian; Meibian Zhang; Jingsong Li; Wei Qiu

Check for updates

J. Acoust. Soc. Am. 145, 2388–2400 (2019)
https://doi.org/10.1121/1.5096643

CrossMark

View Online

Export Citation

Related Content

Analysis of correlation between window duration for kurtosis computation and accuracy of noise-induced hearing loss prediction

J. Acoust. Soc. Am. (April 2021)

Use of the kurtosis statistic in an evaluation of the effects of noise and solvent exposures on the hearing thresholds of workers: An exploratory study

J. Acoust. Soc. Am. (March 2018)

04 October 2023 14:05:49

Development of an automatic classifier for the prediction of hearing impairment from industrial noise exposure

Yanxia Zhao,[1,a)] Yu Tian,[1,a)] Meibian Zhang,[2] Jingsong Li,[1,a),b)] and Wei Qiu[3]
[1]*Key Laboratory for Biomedical Engineering of Ministry of Education, 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*
[3]*Auditory Research Laboratory, State University of New York at Plattsburgh, Plattsburgh, New York 12901, USA*

(Received 19 November 2018; revised 15 February 2019; accepted 8 March 2019; published online 26 April 2019)

The ISO-1999 [(2013). International Organization for Standardization, Geneva, Switzerland] standard is the most commonly used approach for estimating noise-induced hearing trauma. However, its insensitivity to noise characteristics limits its practical application. In this study, an automatic classification method using the support vector machine (SVM) was developed to predict hearing impairment in workers exposed to both Gaussian (G) and non-Gaussian (non-G) industrial noises. A recently collected human database ($N = 2{,}110$) from industrial workers in China was used in the present study. A statistical metric, kurtosis, was used to characterize the industrial noise. In addition to using all the data as one group, the data were also broken down into the following four subgroups based on the level of kurtosis: G/quasi-G, low-kurtosis, middle-kurtosis, and high-kurtosis groups. The performance of the ISO-1999 and the SVM models was compared over these five groups. The results showed that: (1) The performance of the SVM model significantly outperformed the ISO-1999 model in all five groups. (2) The ISO-1999 model could not properly predict hearing impairment for the high-kurtosis group. Moreover, the ISO-1999 model is likely to underestimate hearing impairment caused by both G and non-G noise exposures. (3) The SVM model is a potential tool to predict hearing impairment caused by diverse noise exposures.
© *2019 Acoustical Society of America*. https://doi.org/10.1121/1.5096643

[AKCL] Pages: 2388–2400

04 October 2023 14:05:49

I. INTRODUCTION

The commonly accepted noise damage-risk criterion (DRC) is currently the ISO-1999 (2013) document. However, the ISO-1999 standard is based on data collected in the 1950s and 1960s (Johnson, 1994). Epidemiological data used to prepare the ISO-1999 standard mainly concern exposure to steady-state industrial noises (Burns and Robinson, 1970; Passchier-Vermeer, 1974). The data used to form the ISO-1999 standard are old, and the noise assessment was primitive compared to what is available today. According to the National Institute for Occupational Safety and Health (NIOSH) document (Kardous *et al.*, 2005), the ISO-1999 standard is based on experimental data of mixed quality, which contributes to significant limitations and constrains the ability to establish a widely accepted DRC for impulsive and other types of sounds. Significant controversy exists over the ISO-1999 standard's ability to estimate hearing loss in individuals. For example, an examination of the epidemiological data that was incorporated into ISO 1999 reveals hearing losses from presumably comparable exposures that span more than 70 dB hearing level (HL) at most audiometric test frequencies (Burns and Robinson, 1970; Martin, 1976; Passchier-Vermeer, 1977; Mills *et al.*, 1996). At the time the demographic data were incorporated into ISO-1999, energy metrics were the only metrics that could be easily measured from a noisy environment with noise survey equipment. Thus, the ISO-1999 is suitable for continuous or steady-state Gaussian (G) noise, but is less applicable to complex noises (Ahroon *et al.*, 1993; Zhao *et al.*, 2010). A complex noise is defined as a non-Gaussian (non-G) noise consisting of a background G noise with embedded high-intensity transients, such as impacts/impulses or noise bursts.

The number of animal model experiments has increased dramatically since the early 1970s. These studies have effectively served to demonstrate the complexity of noise-induced hearing loss (NIHL) and have, as a consequence, altered our perspective on NIHL. Conclusions drawn from animal model experiments show that characterizing noise exposure by an energy metric is inadequate in many industrial situations. One of the most important findings is the significant role of temporal variables in the production of NIHL. Since the energy approach, inherent to most criteria, including the ISO-1999 standard, integrates across time, an energy metric is insensitive to the temporal structure and presentation sequence of an exposure (Lei *et al.*, 1994; Qiu *et al.*, 2006; Qiu *et al.*, 2007; Qiu *et al.*, 2013).

One key to much of the problem lies in the role of temporal variables and how they can be quantified. Recent results from animal experiments (Hamernik and Qiu, 2001;

[a)]Also at: Connected Healthcare Big Data Research Center, Zhejiang Lab, Hangzhou, China.
[b)]Electronic mail: ljs@zju.edu.cn

Hamernik *et al.*, 2003; Lei *et al.*, 1994; Qiu *et al.*, 2006; Qiu *et al.*, 2007; Qiu *et al.*, 2013) have shown in a series of equal energy noise exposures that a statistical metric, the kurtosis, computed on both the time and frequency domain signal is strongly correlated with both the magnitude and frequency specificity of noise-induced trauma. That is, at the same equivalent continuous sound level (L_{eq}), there exists a consistent relation between kurtosis and hearing trauma; as kurtosis increases so does hearing trauma. The kurtosis incorporates both the peak and interpeak interval histogram characteristics of a time varying signal. Thus, it can quantify the impulsiveness of a noise exposure. These results suggest that a statistically based metric, the kurtosis, in combination with the energy metric, may be more predictive of NIHL. It is worth mentioning that the kurtosis value of the G noise is equal to 3.

Davis *et al.* (2012) examined the accuracy of the ISO-1999 model on the median noise-induced permanent threshold shift (NIPTS) estimates using 240 highly screened workers exposed to both G and non-G industrial noises in China. Their results showed that the kurtosis could differentiate the hazardous effects produced by G and non-G noise environments, and the ISO-1999 predictive model did not accurately estimate the degree of median NIPTS incurred from high level non-G noise exposures. This result suggests that, instead of the ISO-1999 model, a better predictive model should be developed to accurately evaluate hearing trauma induced by high level complex noise exposure.

In summary, there is agreement on the need to be able to predict the effects of a noise exposure, but there remains no current agreement on the appropriate method to predict hearing hazard in individuals resulting from complex occupational noise exposures (CHABA, 1992; Dunn *et al.*, 1991; Kardous *et al.*, 2005; Patterson *et al.*, 1993; Thiery and Meyer-Bisch, 1988). The main reason for this is the absence of published data on human NIHL, which would need to include a sufficient number of workers with well-documented exposures. To develop a consistent method to predict hearing trauma, a large body of detailed personal noise exposure and audiometric data from exposed workers need to be collected, and a new approach to measure and analyze noise needs to be developed. Currently, our research team collected data from a large number of workers ($N = 2{,}110$) exposed to various industrial environments in China. The database consists of (1) digital recordings of complete shift-long (8 h) temporal waveforms of the noise that the individual workers were exposed to in a variety of industries; and (2) audiograms from the workers exposed to each of the specific noise environments. Using this data a support vector machine (SVM) model was developed to predict the noise-induced auditory system effects. The model was trained and validated using the human database.

In this study, a SVM classifier with a supervised machine learning algorithm was developed to predict hearing impairment caused by a variety of industrial noise exposures. The ability to generate rules from data automatically and predict unknown data make machine learning a promising tool to predict hearing trauma from any industrial noise exposure. The recently collected database with subjects ($N = 2{,}110$) exposed to both G and non-G complex industrial noises was used in modelling the SVM classifier. The prediction performances of the SVM and the ISO-1999 models were compared. The aim of this study was to demonstrate the feasibility of developing a machine learning model for the prediction of hearing impairment in humans exposed to diverse industrial noises.

II. MATERIALS AND METHODS

A. Data collection

1. Subjects

Industrial workers were recruited from 24 factories in the Zhejiang province of China between 2010 and 2017. Subjects ($N = 3{,}242$) were introduced to the study purpose and design by an occupational physician and were asked to sign an informed consent form. The Zhejiang Provincial Center for Disease Control and Prevention (ZJCDC) institutional committee for the protection of human subjects approved the study protocol (approval reference number: ZJCDC-T-043-R).

For inclusion in the study, all subjects had to satisfy the following five criteria: (1) a minimum of at least one year of employment in their current position; (2) consistently worked in the same job category and at the same work site (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; and (5) no history of using hearing protection. Accordingly, a total of 2110 workers were included from the original pool of 3242 subjects. Table I provides a breakdown of average noise exposure level, duration of exposure, kurtosis, age, and sex corresponding to the number of subjects exposed by the factory.

While the implementation of hearing conservation programs is an important and necessary practice, with the use of HPDs, each worker's occupational exposure is attenuated in an unpredictable fashion, interfering with the effective estimation of noise exposure over time for experimental studies. Whether a worker has used a hearing protection device (HPD) in his/her industrial noise exposure history was determined from the noise exposure questionnaire and interview. The subjects with the use of HPD were excluded in this study. For those workers who have never used HPD, the members of the research team recommended them to use adequate HPDs according to the individual results of the noise data analysis. The workers were also given training of how to use HPD properly during the investigation.

2. Questionnaire survey

An occupational hygienist from ZJCDC administered a questionnaire to each subject in order to collect the following information: general personal information (age, sex, etc.), occupational history (factory, work site, 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 conditions (including history of ear disease and use of ototoxic drugs). An occupational physician entered all information into a database.

04 October 2023 14:05:49

TABLE I. A breakdown of average noise exposure level, duration of exposure, kurtosis, age, and sex, corresponding to the number of subjects exposed by factory. (*n*, number of subjects in each group; ±, plus/minus 1 standard deviation; –, minimum to maximum.)

Factory	Male (n)	Female (n)	Age (year)	Duration (year)	L_{Aeq} (dBA)	Median Kurtosis
Spandex plant	100	12	20–55	1–17	61–104	3–107
			33.4±8.2	6.2±3.4	88.6±9.3	18.7±15.7
Furniture factory #1	88	2	19–45	1–12	77–94	25–926
			27.6±5.5	2.7±2.2	86.2±4.01	233.8±180.4
Furniture factory #2	64	11	21–63	1–23	84–98	21–562
			39.2±9.1	8.1±5.6	91.5±3.3	176.6±136.2
Furniture factory #3	47	0	18–48	1–10	85–97	13–777
			33.0±8.2	4.6±2.6	90.3±3.0	187.9±160.8
Furniture factory #4	69	18	20–52	1–20	75–101	14–374
			32.8±7.7	3.5±3.6	88.0±4.6	77.2±65.0
Furniture factory #5	50	7	19–62	1–18	76–99	33–863
			34.6±9.8	3.7±3.2	90.5±3.5	273.0±164.4
Furniture factory #6	59	16	20–70	1–37	66–95	15–307
			43.2±11.8	9.9±8.3	83.4±6.1	78.8±67.9
Auto parts factory	332	58	19–62	1–30	50–98	4–82
			34.2±8.3	6.1±4.7	87.50±4.22	14.7±10.8
Woven bag manufacturing company	37	49	20–58	1–23	69–102	4–139
			39.3±7.0	10.2±3.9	90.4±6.2	15.1±16.4
Hardware factory	48	15	19–56	1–20	71–94	6–128
			31.3±8.7	5.3±4.5	86.2±5.7	20.8±19.5
Electronic equipment factory	45	23	19–46	1–19	68–101	4–46
			27.6±5.6	3.5±3.8	84.1±9.3	15.8±9.1
Paper plant #1	31	13	20–67	1–35	75–97	3–52
			47.8±11.1	10.5±9.9	88.7±4.4	12.1±10.0
Paper plant #2	35	22	25–63	1–33	71–101	4–70
			47.5±8.7	12.2±6.7	88.7±4.5	10.2±10.0
Machinery plant #1	16	7	33–64	5–20	63–104	4–169
			46.8±7.7	11.0±3.9	89.0±10.6	31.3±33.4
Machinery plant #2	37	5	22–60	2–37	66–102	13–175
			39.7±10.6	11.7±9.9	82.5±8.1	53.4±37.5
Brake pads factory	112	51	19–49	1–9	75–104	6–647
			30.4±6.9	2.9±1.7	87.4±5.1	51.5±75.8
Hydroelectric plant	131	6	20–69	1–40	68–105	5–154
			42.6±12.0	12.0±8.5	86.9±8.2	35.1±25.3
Stroller factory	42	37	23–55	1–20	86–102	6–206
			39.9±8.4	4.5±3.9	93.6±3.8	20.3±23.5
Electrical appliance factory	43	11	18–36	1–17	71–96	6–739
			25.0±4.2	2.8±3.1	81.5±5.9	66.1±119.3
Kitchen and bath manufacturing company	7	15	24–60	1–19	68–89	17–177
			43.5±9.4	8.3±5.2	80.5±4.7	71.9±54.2
Yarn–dyed fabric factory	10	74	18–48	1–30	93–105	3–21
			30.8±7.4	5.2±4.5	97.4±1.8	5.5±2.9
Cement plant	56	9	22–57	2–38	69–110	3–67
			44.3±7.0	19.4±9.6	84.5±8.5	21.5±15.8
Machinery and electric company	38	116	22–50	1–29	78–98	7–241
			36.1±7.2	7.8±5.3	85.9±3.5	38.4±30.7
Pipe manufacturing company	33	3	21–56	1–35	74–93	8–64
			36.1±10.1	9.2±8.0	84.2±4.9	29.1±13.7

3. Noise data collection

Shift-long noise recording files were obtained for each noise-exposed subject at the 24 factories using an ASV5910-R digital recorder (Hangzhou Aihua Instruments Co., Hangzhou, China). The ASV5910-R digital recorder is a specialized sound recording device that can be used for precision measurements and analysis of personal noise exposure. The instrument uses a $\frac{1}{4}$-in. pre-polarized condenser microphone characterized by 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 40–141 dBA. The ASV5910-R runs a self-calibration program each time the power is turned on. The shift-long noise for each subject was continuously recorded by the ASV5910-R at 32-bit resolution with a 48-kHz sampling rate. The noise record was saved in a 32 GB micro secure digital (SD) card and transferred to network-attached storage for subsequent analysis.

04 October 2023 14:05:49

4. Physical and audiometric evaluation

Each subject underwent a general physical and otologic examination. 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 sound pressure level (SPL)] using an audiometer (Madsen, OB40; Otometrics, Copenhagen, Denmark) calibrated according to the Chinese national standard (GB4854–84). Audiograms were measured at least 16 h after the subjects' last occupational noise exposure.

B. Experimental design

The purpose of this study was to create an accurate prediction model that can predict hearing trauma not only caused by G noise but also non-G complex noises. A machine learning automatic classifier (the SVM classifier) was developed to predict hearing trauma using the human database ($N = 2{,}110$) collected in China. The NIOSH definition of material hearing impairment (MHI) was used (NIOSH, 1998). The SVM model was evaluated by comparing its predictive performance with that of the ISO-1999 model. Model development and evaluation included: (1) the SVM classifier design; (2) the ISO-1999 prediction approach; (3) the categorization of the database; and (4) the performance metrics for both the SVM classifier and the ISO-1999 model.

1. The SVM model design

The SVM model design includes: (1) selection of the SVM model structure, (2) determination of target variables, (3) feature extraction, and (4) feature selection.

a. SVM classifier structure. The SVM is a large margin classifier, which means that the purpose of the SVM algorithm is to find an optimal separating hyperplane (OSH) that generates a maximum margin between two categories of data. To construct an OSH, the SVM maps data onto a higher-dimensional feature space by using a kernel function. For the training dataset, $\{(\boldsymbol{x}_i, y_i), 1 \le i \le N\}$, where each input vector $\boldsymbol{x}_i \in \boldsymbol{R}^d$, belongs to either of two classes identified by the label y_i. The OSH is the solution to the optimization problem,

$$\min \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}\delta_i,$$
$$\text{subject to} \quad y_i\big(\langle \varnothing(\boldsymbol{x_i}), \boldsymbol{w}\rangle + b\big) \ge 1 - \delta_i, \tag{1}$$

where $\varnothing(\boldsymbol{x_i})$ represents a nonlinear transformation that maps the training data to a higher-dimensional feature space, $\delta_i \ge 0,\ i = 1, \ldots, N$ represent losses, and C is a regularization parameter that controls the trade-off between the margin and the losses. Equation (1) can be rewritten into its dual form by using Lagrange multipliers (Bishop, 2006).

$$\max_{\alpha_i} \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i y_i \alpha_j y_j K(\boldsymbol{x}_i, \boldsymbol{x}_j),$$
$$\text{subject to} \quad 0 \le \alpha_i \le C, \quad \sum_{i=1}^{N}\alpha_i y_i = 0, \tag{2}$$

where α_i represents Lagrange multipliers and $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle \varnothing(\boldsymbol{x_i}), \varnothing(\boldsymbol{x_j})\rangle$ is the kernel function. In this study, a radial basis function (RBF) was chosen as the kernel function, which can be expressed by

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = e^{-\gamma\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}. \tag{3}$$

Detailed information about the SVM is available in the literature (Bishop, 2006; Vapnik, 2013).

To make the utmost use of the training database, a k-fold cross-validation was used, with $k = 10$ to modify the classifier. The training database was separated into k equal groups of samples. At each iteration, the kth group of samples was used for testing, and the other k–1 groups of samples were used for training. The performance of the classifier was calculated using the average performance of each fold.

b. Determination of target variable. In this study, the NIOSH definition of MHI was used, i.e., the average HTLs at 1, 2, 3, and 4 kHz for both ears exceeding 25 dB HL. Hearing impairment was used as the target variable and defined as a binary classification.

c. Feature extraction. As discussed above, energy alone is not enough to evaluate hearing trauma caused by complex noise exposures. Noise characteristics in both the time and frequency domains are also needed. A total of 42 candidate features were extracted from the noise records for use as inputs to the SVM model. Of the 42 candidate features (see Table II), 6 of them are features from the time domain, 33 of them are features from the frequency domain, and the remaining 3 are personal features.

a) Rational for extracting features from the time domain: The current damage risk criteria for noise exposure is based entirely on an energy metric. The L_{eq} has long been used as an efficient metric for predicting NIHL. Since non-G complex noise is common in industrial settings, the temporal characteristics of noise need to be factored in to characterize non-G noise for the evaluation of the auditory effects of noise exposure. A variety of animal experiments and epidemiological studies has shown that, in addition to an energy metric, the statistical metric kurtosis provides a characterization of complex noise (Qiu *et al.*, 2013; Zhao *et al.*, 2010). The kurtosis [$\beta(t)$] can be computed on the amplitude distribution of the temporal waveform of the noise presented to the subject or by filtering the waveform; a frequency-specific kurtosis [$\beta(f)$] can be computed from the resulting time-domain signal. The kurtosis of the recorded noise signal is computed in 40-s time windows consecutively without overlap over the full shift recordings with a sampling rate of 48 kHz. Because the kurtosis is dependent on the length of the window over which the calculation is made, and its calculation is limited by the computer's processing capabilities, a compromise was made to use a 40-s time window, which, based on previous animal data (Hamernik *et al.*, 2003), was found to be sufficient to establish an acceptable measure of the kurtosis. Figure 1(A) shows a sample of the 40-s amplitude-time waveform and the corresponding spectrum of a

04 October 2023 14:05:49

TABLE II. Extracted features for the SVM model.

Category	Variable name	N
Features in time-domain:	L_{eq8hr}	1
	Max_Peak_SPL	1
	Median_L_{eq40s}	1
	Time-domain kurtosis (mean kurtosis, median kurtosis, geometric mean_kurtosis)	3
Features in frequency domain	L_{Aeq8hr}	1
	L_{Ceq8hr}	1
	L_{Feq8hr}	1
	Median_L_{Aeq40s}	1
	Median_L_{Ceq40s}	1
	Median_L_{Feq40s}	1
	L_{eq8hr_cf}[a]	9
	Median_L_{eq40s_cf}[a]	9
	Median octave-band frequency-domain kurtosis: median_FK_cf[a]	9
Personal features	Age	1
	Sex	1
	Noise exposure duration	1

[a]cf = 63, 125, 250, 500, 1000, 2000, 4000, 8000, and 16 000 Hz, respectively.

non-Gaussian noise exposure. Figure 1(B) illustrates an example of the kurtosis and L_{eq40s} calculation across a shift-long noise recording. From Fig. 1(B), it can be noted that hundreds of kurtosis values were obtained from the digital recordings of complete shift-long temporal noise waveforms. The distribution of the kurtosis values was studied. The most common distribution is a log normal distribution followed by, but less often, a normal distribution and a uniform distribution. Therefore, the mean, median, or geometric mean of the kurtosis values were calculated and selected as the candidate kurtosis metrics. The maximum peak SPL of the 40-s calculation window was also selected as one of the temporal features of the noise exposure.

The procedure to calculate the mean kurtosis, median kurtosis, geometric mean kurtosis, and the max value of the peak SPL can be briefly summarized as follows. Each shift-long noise recording signal was divided into consecutive 40-s time segments. For each 40-s time segment, the kurtosis β_{T_i} was calculated as

$$\beta_{T_i} = \frac{\frac{1}{N_i}\sum_{n=1}^{N_i}(x_n - \bar{x})^4}{\left(\frac{1}{N_i}\sum_{n=1}^{N_i}(x_n - \bar{x})^2\right)^2}. \tag{4}$$

The peak SPL L_{pk_i} was calculated as

$$L_{pk_i} = 20\lg\left(\max_{n\in[1,N_i]}(|x_n|)\right) - Lx + 94, \tag{5}$$

where x_n refers to the data point, $\bar{x}$ is the sample mean, the index T_i refers to each 40-s time segment and includes N_i data points, and Lx denotes the microphone's sensitivity level. Then, the mean kurtosis, the median kurtosis, the geometric mean kurtosis, and the max value of the peak SPL for these 40-s windows were calculated for the entire shift.

b) Rational for extracting features in the frequency domain: In noise measurement standards an A-weighted equivalent SPL (L_{Aeq}) has been widely used to assess the auditory risk of occupational noise exposures. In addition, C-weighting has also been used in standards to detect the peak SPL of noise. Moreover, Sun *et al.* (2015) developed an F-weighting, a kurtosis-related metric, which may be more suitable for measuring complex noise. Two time scales, i.e., 8-h and 40-s, were used to normalize the L_{eq} in this study. Thus, L_{Aeq8hr}, L_{Ceq8hr}, and L_{Feq8hr}, the median

04 October 2023 14:05:49

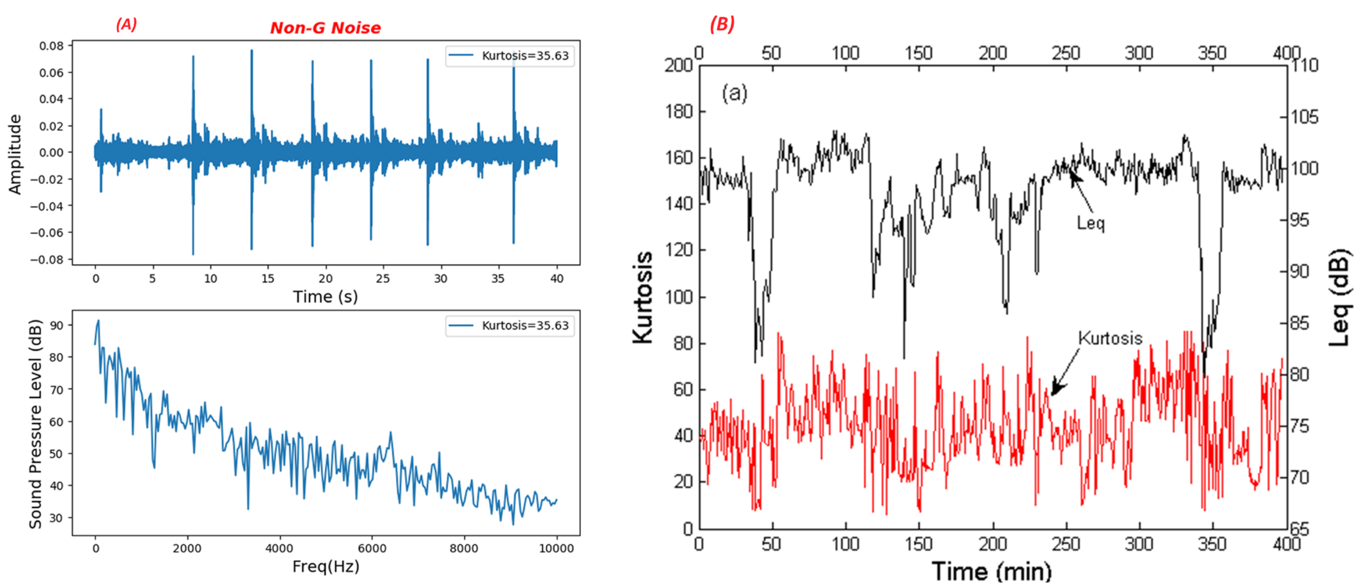

FIG. 1. (Color online) (A) The 40-s sample of the amplitude-time waveform of a non-Gaussian exposure and the spectrum obtained from the 40-s samples of the digitized noise waveform. (B) An example of kurtosis and L_{eq40s} calculation cross shift-long noise records. The kurtosis and L_{eq} of the recorded noise signal were computed for consecutive 40-s time windows without overlap over the full shift using MATLAB software.

L_{Aeq40s}, the median $\mathrm{L}_{\mathrm{Ceq40s}}$, and the median L_{Feq40s} were extracted as candidate spectral features. Sun *et al.* (2015) define the F-weighting as

$$X_{T_i}(f) = \alpha_{A,T_i}(W_A(f) * X_{T_i}(t)) + \alpha_{c,T_i}(W_c(f) * X_{T_i}(t)), \tag{6}$$

where $X_{T_i}(f)$ refers to the filtered noise signal and $X_{T_i}(t)$ refers to the unfiltered noise signal; $W_A(f)$ and $W_C(f)$ refer to the A-weighted and C-weighted filters, respectively (Havelock *et al.*, 2008); the symbol "*" refers to the convolution calculation; and α_{A,T_i} and α_{C,T_i} are the weighting coefficients for A-weighting and C-weighting, respectively. The determination of α_{A,T_i} and α_{C,T_i} is related to two statistical metrics, i.e., the kurtosis and oscillation coefficient of the noise exposure. The kurtosis is used to measure the impulsiveness of a complex noise, while the oscillation coefficient is relevant to the local transition level and frequency of a complex noise signal. Both α_{A,T_i} and α_{C,T_i} reflect the energy distribution of the steady-state and impulsive components of a complex noise signal (Sun *et al.*, 2015).

The spectrum of noise is acknowledged to be an important feature of hearing trauma assessment. Thus, octave-band equivalent SPLs at different center frequencies, i.e., $L_{\mathrm{eq8hr\_cf}}$, were therefore extracted as candidate spectral features, where center frequency (cf) = 63, 125, 250, 500, 1000, 2000, 4000, 8000, and 16 000 Hz, respectively. For similar reasons, the median $L_{\mathrm{eq40s\_cf}}$ was also extracted as a spectral feature.

The time domain kurtosis, β(t), is not sensitive to the frequency components of a non-G noise. From animal experiments, Lei *et al.* (1994) and Hamernik *et al.* (2003) found that the kurtosis computed on the filtered octave band temporal noise signal, β(*f*), was useful in the prediction of location-specific outer hair cell loss and permanent threshold shift. Therefore, the median octave-band frequency-domain kurtosis at different center frequencies (median_FK_cf) was extracted as spectral features in order to estimate hearing trauma in this study.

c) Rational for choosing personal features: The age, sex, and exposure duration of individual workers are key features for NIPTS prediction in the ISO-1999 model.

d. Feature selection. As shown in Table II, there were 42 extracted features for the SVM classifier. To optimize the classification performance, the extracted features were subjected to a feature selection phase for a classifier optimization process. The random forest (RF) algorithm (Louppe *et al.*, 2013) was used to rank the importance of features based on the mean decrease impurity, defined as the total decrease in the node impurity and averaged across all the trees of the RF. The mean decrease in the Gini index (Gini impurity) was used in this study (Louppe *et al.*, 2013). Tree training computes how much each feature decreases the weighted impurity of a tree. For the RF, the impurity decrease of each feature is averaged, and the features are ranked according to this measure. The information from the ranked features was used to determine the relationship between feature properties and hearing impairment for workers exposed to complex industrial noise. The feature selection method is the best way to determine the optimal subspace of ranked features within the original feature space.

2. The ISO-1999 prediction approach

The hearing impairment can be predicted by the ISO-1999 predictive model. The calculation procedures of the ISO-1999 (2013) model can be briefly summarized as follows:

(1) The L_{Aeq} is normalized to a nominal 8-h working day $L_{Ex,8h}$ level in decibels.

(2) The HTLs at 1, 2, 3, and 4 KHz, in decibels, associated with the age and noise, H', of a noise-exposed worker is calculated using the following formulas:

$$H' = H + \mathrm{N} - \frac{H \times \mathrm{N}}{120}, \tag{7}$$

where H is the HTL expressed in decibels and associated with age; N is the actual or potential NIPTS expressed in decibels.

$$H = a(Y - 18)^2, \tag{8}$$

$$N = \begin{cases} \left[u + v lg\left(\dfrac{t}{t_0}\right)\right](L_{Ex,8h} - L_0)^2, & 10 \le t \le 40 \\ \dfrac{lg(t+1)}{lg(11)}\left[u + v lg\left(\dfrac{10}{t_0}\right)\right](L_{Ex,8h} - L_0)^2, & 10 < t, \end{cases} \tag{9}$$

where $t_0 = 1$, L_0 is the SPL defined as a function of the frequency in Table I in the ISO-1999 (2013) document, u and v are coefficients given as a function of frequency in Table I of ISO-1999 (2013), Y is age, and the values for coefficient α are presented in Table A.1 of ISO-1999 (2013).

(3) The average binaural HTL at 1–4 kHz for each subject is calculated according to the NIOSH definition of MHI at 1–4 kHz.

3. Categorization of the database

The ISO-1999 (2013) standard is based on data acquired primarily from G or quasi-G noise exposures (Burns and Robinson, 1970; Passchier-Vermeer, 1974). Considering that many industrial noise environments are non-G and energy metrics (e.g., L_{eq}) are suitable for G noise, there is a need to evaluate alternative metrics or a combination of metrics that have been found to be appropriate for assessing non-G noise environments. This limitation to the ISO-1999 is supported by demographic and experimental animal data showing that non-G noise is more hazardous to hearing than G noise of the same spectrum and L_{eq} (Taylor *et al.*, 1984; Thiery and Meyer-Bisch, 1988; Zhao *et al.*, 2010; Hamernik *et al.*, 2003; Qiu *et al.*, 2013). Results from our animal experiments have shown that: (1) The kurtosis of the amplitude distribution (ratio of the fourth-order central moment to the squared second-order moment of the amplitude distribution) may be

04 October 2023 14:05:49

a reasonable candidate to assess the risk of hearing loss from complex noise (Davis *et al.*, 2012; Qiu *et al.*, 2006; Qiu *et al.*, 2007; Hamernik *et al.*, 2003). (2) The kurtosis ordered the extent of hearing and sensory cell loss from a variety of complex noise exposures, i.e., for a fixed energy level, the noise-induced trauma increased as the kurtosis increased. For human subjects, the question remains as to how the ISO-1999 (2013) standard, developed from the results of steady-state exposures and quantified by weighted energy alone, predicts NIHL from complex noise environments.

In comparing the accuracy of NIPTS by both G and non-G noises using the ISO-1999 (2013) predictive model, Davis and his colleagues (2012) divided the human subjects into two groups based on the kurtosis level of the industrial noise exposures, i.e., $\beta(t) \leq 10$ and $\beta(t) > 10$. They found that the median NIPTS increased by an average of 8 dB as the mean kurtosis level increased from mean $\beta(t) \leq 10$ to $\beta(t) > 10$ across the test frequencies (2–6 kHz) of the noise exposed groups. The ISO-1999 approach significantly underestimated the median NIPTS for the higher kurtosis [$\beta(t) > 10$] group. The number of subjects ($N = 240$) in the Davis' study limited the statistical power of any further detailed analysis. Thus, dividing the subjects into smaller groups based on the value of kurtosis of the noise was not feasible. The large database ($N = 2{,}110$) used in this study makes it possible to further evaluate the accuracy of the ISO-1999 predictions in different kurtosis-specified groups by categorizing the subjects into the following four subgroups based on the kurtosis levels of the noise exposures:

(a) G/quasi-G group [mean $\beta(t) \leq 10$];
(b) low kurtosis group [$10 <$ mean $\beta(t) \leq 25$];
(c) middle kurtosis group [$25 <$ mean $\beta(t) \leq 75$];
(d) high kurtosis group [mean $\beta(t) > 75$].

Table III shows the number of workers and the kurtosis distribution for two hearing classes (i.e., normal hearing and hearing impairment) over the four kurtosis groups. A fifth group with all subjects ($N = 2{,}110$) was also used to evaluate the predictive accuracy of the ISO-1999 for the median NIPTS of workers exposed to both G and non-G complex industrial noises.

4. Performance metrics

The SVM classifier and the ISO-1999 predictive models were used to predict hearing impairment of each individual worker exposed to various industrial noises. There are only two classes for the classifier, i.e., positive or negative. The positive class represents the presence of hearing impairment, and the negative class represents the absence of hearing impairment. Therefore, four possible outcomes can result from the classification model:

(a) True positive (TP)—when the actual class is positive and the predicted class is also positive.
(b) True negative (TN)—when the actual class is negative and the predicted class is also negative.
(c) False positive (FP)—when the actual class is negative, but the predicted class is positive. In the case of hearing impairment prediction, the FP refers to the fact that the actual class indicates a worker did not suffer from hearing impairment, whereas the predicted class says the worker did. The FP indicates that a predictive model overestimates hearing trauma due to noise exposure.
(d) False negative (FN) – when the actual class is positive, but predicted class is negative. The FN refers to the fact that the actual class indicates a worker suffered from hearing impairment, but the predicted class says the worker did not. The FN indicates that a predictive model underestimates the hearing trauma due to noise exposure.

Although both false predictions need to be avoided, it is worth noting that the cost of a FN is much higher than that of a FP in the case of hearing impairment prediction. A FN can result in the hearing of workers not being properly protected in time.

In this study, four performance metrics (i.e., accuracy, precision, recall, and $F1$ score) were used to assess the performance of a classifier. The definitions and interpretation of these four performance metrics are presented as follows (Vihinen, 2012):

- *Accuracy*: Accuracy (A) is the most intuitive performance measure and is the ratio of correctly predicted observations over the total observations.

$$A = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN}}. \tag{10}$$

Accuracy is a good measure only when values of FP and FN are almost identical.

- *Precision:* Precision (P) is the ratio of correctly predicted positive observations over the total predicted positive observations.

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}. \tag{11}$$

TABLE III. The worker and kurtosis distribution of two hearing classes (i.e., normal hearing and hearing impairment) over four kurtosis groups. (n, number of subjects in each group; ±, plus/minus 1 standard deviation; –, minimum to maximum.)

	G/quasi-G group	Low-kurtosis group	Middle-kurtosis group	High-kurtosis group
Mean kurtosis (range)	6.7 ± 1.9	16.5 ± 4.2	41.8 ± 13.1	214.4 ± 149.5
	(3.0–10.0)	(10.1–25.0)	(25.1–75.0)	(75.1–926.0)
Hearing impairment (n)	149	202	177	145
Hearing normal (n)	324	509	382	214
All (n)	473	711	559	359

04 October 2023 14:05:49

Based on Eq. (11), a high precision ratio relates to a low FP rate.

- *Recall:* Recall (R) is the ratio of correctly predicted positive observations over all the observations in the actual positive class.

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{12}$$

Recall is the preferred metric used to select the best model when there is a high cost associated with FN. Thus, the recall ratio is an important metric to evaluate the prediction performance of this study.

- *F1 score*: $F1$ score is the weighted average of precision and recall. This metric takes both FPs and FNs into account.

$$F_1 = \frac{2PR}{P + R} \tag{13}$$

$F1$ is useful when the class distribution is uneven.

C. Statistical analysis

The predictive performance of the SVM and ISO-1999 models were compared among five groups (i.e., four kurtosis groups and one all-data group) using bootstrapping (100 replications) and Wilcoxon signed ranks test (Demsar, 2006). A statistical significance level of $p < 0.05$ was adopted for all analyses. The 95% confidence intervals were obtained through bootstrapping (100 replications).

D. Experimental environment

The experimental environment is a Linux server (Ubuntu 16.04.4 LTS, Xenial, London) with a centeral processing unit (CPU) Intel (R) Xeon (R) E5–2620 (Santa Clara, CA). The frequency of the CPU is 2.10 GHz. The GPUs includes a NVIDIA Tesla K40M. The programming language used is Python (Python 3.6, Holland), and the debugging environment is Spyder (Anaconda, Austin, TX).

III. RESULTS

A. Feature extraction and selection for the SVM classifier

Table IV shows the outline of the demographic and noise exposure data for all subjects, the normal hearing group and the hearing impaired group. From Table IV, it is clear that the hearing impaired group is largely related to age, exposure duration, and L_{eq}. Table V displays the average kurtosis-related metrics for all subjects, the normal hearing group and the hearing impaired group. All the kurtosis-related metrics, either from the time or frequency domain, are generally associated with hearing impairment. The higher the kurtosis value, the more probable the hearing impairment is. Among all the kurtosis-related metrics, the mean kurtosis shows the largest difference between the normal hearing group and the hearing impaired group. This implies that the mean-kurtosis is a more sensitive metric for predicting hearing impairment.

TABLE IV. Outline of demographic and noise exposure data for all subjects, normal hearing and hearing impairment groups. (±, plus/minus 1 standard deviation.)

	All subjects	Hearing impairment	Normal hearing
Number of subjects	2110	673	1437
Male (%)	1530 (72.5)	498 (74.0)	1032 (71.8)
Age (yr)	35.8 ± 10.1	42.3 ± 9.4	32.7 ± 8.8
Exposure duration (yr)	7.0 ± 6.5	9.5 ± 7.3	5.8 ± 5.8
L_{Aeq} (dBA)	87.7 ± 6.5	89.5 ± 5.6	86.9 ± 6.7
L_{Feq} (dB)	91.3 ± 8.4	93.3 ± 8.2	90.3 ± 8.3

TABLE V. The average kurtosis-related metrics for all subjects, the normal hearing group and the hearing impairment group. (±, plus/minus 1 standard deviation.)

	All subjects	Hearing impairment	Hearing normal
Median kurtosis	19.8 ± 42.1	24.6 ± 50.0	17.5 ± 37.7
Mean kurtosis	54.8 ± 96.2	65.5 ± 109.1	49.8 ± 89.1
Geomean kurtosis	19.3 ± 28.6	22.9 ± 32.9	17.6 ± 26.2
median_FK_63	8.5 ± 12.7	8.6 ± 11.9	8.5 ± 13.1
median_FK_125	12.2 ± 19.4	12.2 ± 17.8	12.3 ± 20.1
median_FK_250	17.6 ± 26.6	18.7 ± 28.0	17.0 ± 25.9
median_FK_500	25.4 ± 35.0	28.8 ± 42.4	23.7 ± 30.8
median_FK_1000	38.0 ± 56.1	42.9 ± 64.0	35.7 ± 51.9
median_FK_2000	51.3 ± 100.7	56.7 ± 101.6	48.7 ± 100.2
median_FK_4000	65.4 ± 132.9	74.0 ± 142.0	61.4 ± 128
median_FK_8000	74.2 ± 125.1	77.4 ± 143.6	72.7 ± 115.4
median_FK_16000	108.4 ± 224.3	114.8 ± 255.8	105.3 ± 207.9

Figure 2 displays the distribution of all SPL-related features extracted from the recorded noise waveforms. The line in the center of the box indicates the median of the data. The upper and lower edges of the box display the 25th and 75th percentiles of the data, respectively. The "whiskers" that extend from the top and bottom of the boxes is a distance of 1.5 times the interquartile range (IQR). The space between the top and bottom of the boxes shows the range of the data distribution. Data points beyond the whiskers are considered outliers, which are indicated with the "+" symbol. It can be seen from Fig. 2 that there are many outliers for almost all the features. In order to obtain comparable means and variances, a z-score normalization (Patro and Sahu, 2015) was performed for each feature before feature selection.

The ranking sequence of the 42 extracted features obtained by the RF method is shown in Fig. 3(A). The optimal number of ranked features was determined by comparing the accuracy of the different features using the RF selection method. All the extracted features were evaluated using the SVM classifier for tenfold cross-validation. In other words, the SVM classifier was trained using the tenfold cross-validation to correspond with the first ranked feature to compute the mean accuracy. Then, the second-ranked feature was added to the first-ranked feature, and the tenfold cross-validation was performed in the same way. This process continued until all 42 of the extracted features were considered. Figure 3(B) shows the score curve using different numbers of features of the SVM classifier in terms of the mean accuracy from the tenfold cross-validation. Based on Fig. 3(B), the optimal number of ranked features was nine.

04 October 2023 14:05:49

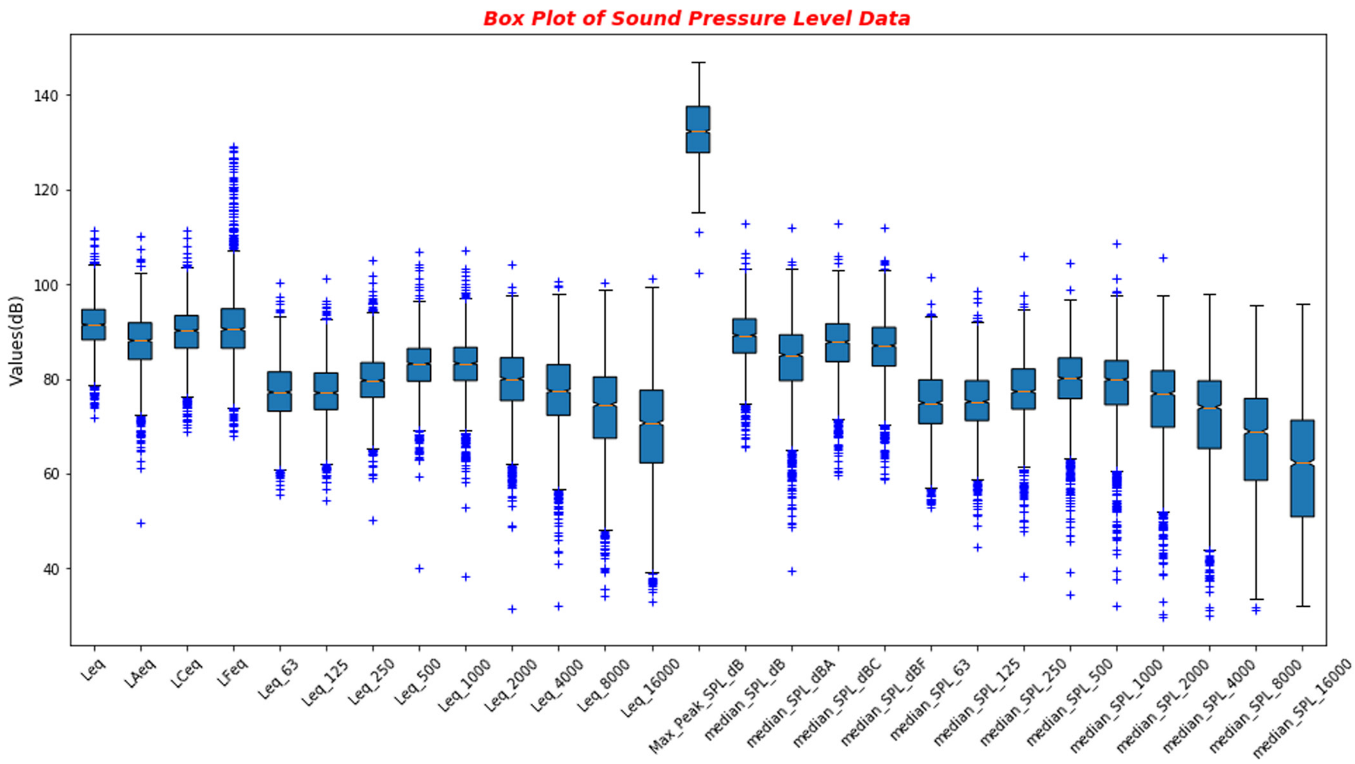

FIG. 2. (Color online) Distributions of all SPL-related features.

04 October 2023 14:05:49

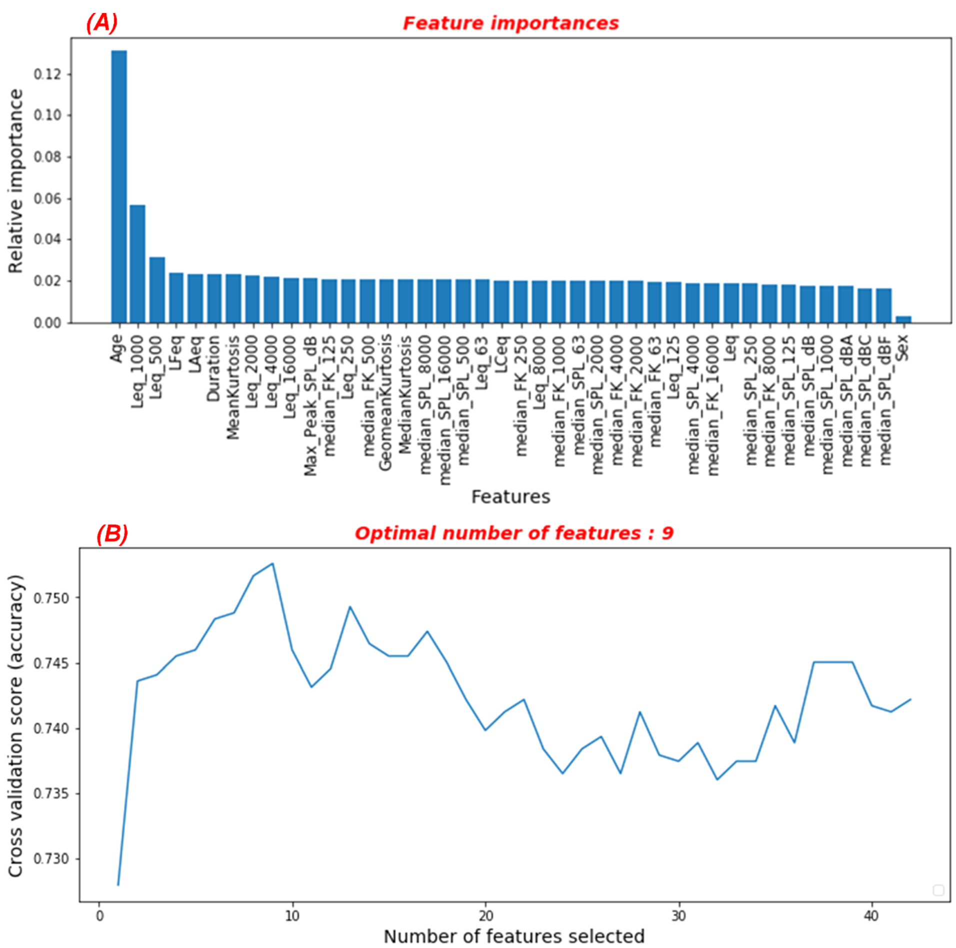

FIG. 3. (Color online) (A) The ranking result of the 42 extracted features obtained by the random forest method. (B) The cross-validation score as a function of the number of different features of the SVM classifier (optimal number of features was nine).

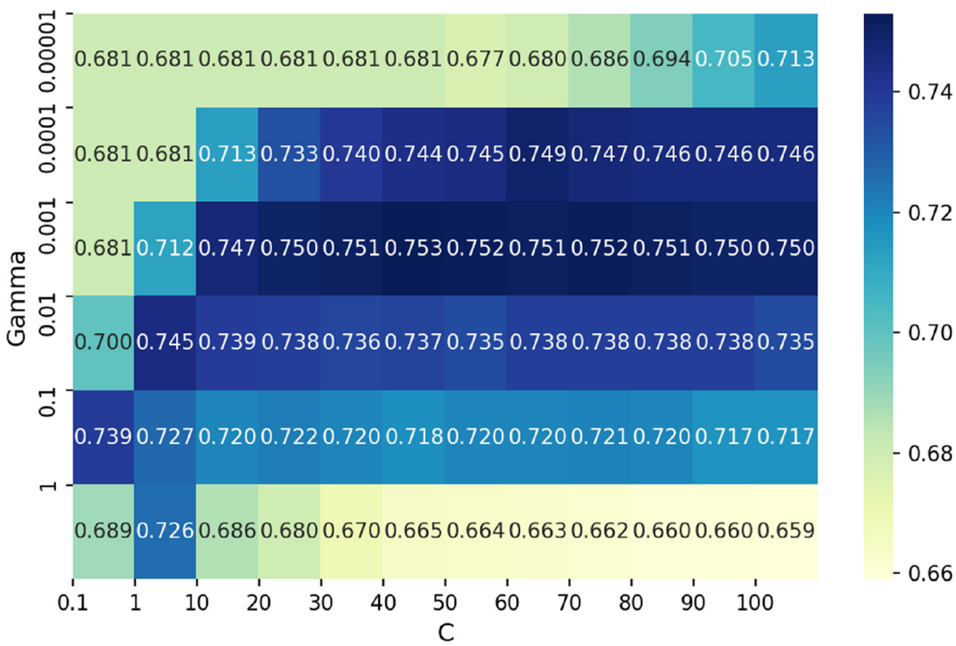

FIG. 4. (Color online) The mean accuracy of the SVM classifier for tenfold cross-validation. Gamma is the kernel parameter, and C is the regularization parameter.

Hence, the final RF optimum set consists of the nine top-ranked features, specifically age, L_{eq_1000}, L_{eq_500}, L_{Feq}, L_{Aeq}, exposure duration, mean kurtosis, L_{eq_2000}, and L_{eq_4000}.

B. Classification

1. Model selection and training

In this study, a Gaussian RBF kernel was used in the SVM model. Two parameters, the gamma and the regularization parameter C, determine the best separation of the data according to the SVM model. The grid search algorithm (Woodford and Phillips, 2011) was used to determine the optimal structure of the machine learning classifier. Figure 4 displays the corresponding mean accuracy vs two parameters, i.e., gamma and C, for the tenfold cross-validation. The parameters C and gamma were tested over [0.1,100] and $[10^{-4},1]$, respectively. The model achieved its optimum performance when gamma was set as 10^{-2} and C as 40.

2. Comparison of classification performance of the SVM and the ISO-1999 methods

Table VI shows the classification performance of the SVM and the ISO-1999 models over four kurtosis groups. The performance of each model was evaluated based upon the four model measures: the accuracy, precision, recall, and $F1$ score. For all performance measures, the SVM model outperforms the ISO-1999 over five groups, with the exception of the precision performance for G/quasi G noise group, which is equivalent between the SVM and the ISO-1999 models.

Since accuracy is the most commonly used and intuitive model measure, accuracy was first used to compare the performance of the two predictive models. The accuracy of the SVM model in the all-data group is significantly better than that of the ISO-1999 model (75.3% vs 68.6%, $p < 0.0001$). The accuracy of the ISO-1999 model in G/quasi-G, low, and middle kurtosis groups is around 70%, but significantly decreases by 10% in the high kurtosis group. In contrast, the accuracy of the SVM model remains above 72% for all four kurtosis groups. The accuracy of all four kurtosis groups is significantly better using the SVM model than using the ISO-1999 model ($p < 0.0001$).

In addition to the accuracy measure, the performance of the SVM classifier and the ISO-1999 model were analyzed from the angles of FP and FN by using the other three model measures, i.e., precision, recall, and $F1$ score. The results are shown in Table VI and summarized as follows:

- Precision: The precision of the SVM model in the all-data group is significantly better than that of the ISO-1999

TABLE VI. Classification performance of the SVM and the ISO-1999 models. (CI, confidence interval; $^{\#}p = 0.58$; $^{**}p < 0.0005$; $^{***}p < 0.0001$.)

Group	Model	Accuracy (%)(95% CI)	Precision (%) (95% CI)	Recall (%) (95% CI)	$F1$ score (%) (95% CI)
All-data	ISO-1999	68.6 (66.3–70.2)	71.0 (65.9–75.0)	51.0 (50.5–51.3)	59.4 (57.3–60.8)
	SVM	**75.3** (75.0–77.3)***	**74.3** (73.7–76.5)**	**68.9** (66.4–71.5)***	**71.5** (70.0–73.8)***
G/quasi-G	ISO-1999	69.6 (66.2–72.3)	78.9 (77.7–80.1)	51.7 (50.6- 52.6)	62.5 (61.3–63.5)
	SVM	**77.2** (76.9–81.2)***	**76.7** (75.8–81.1)$^{\#}$	**72.1** (70.0–77.3)***	**74.3** (73.4–79.2)***
Low kurtosis	ISO-1999	71.7 (70.4–73.6)	68.5 (55.8–75.7)	50.5 (49.9–51.2)	58.2 (53.0–61.1)
	SVM	**74.7** (74.0–9.4)***	**73.0** (72.7–78.4)**	**64.9** (59.8–69.6)***	**68.7** (65.3–73.3)***
Middle kurtosis	ISO-1999	68.9 (66.2–72.1)	69.7 (59.0–78.4)	51.2 (50.5–52.2)	59.0 (55.9–62.1)
	SVM	**75.3** (74.3–80.2)***	**74.7** (73.2–80.0)**	**70.1** (66.3–76.4)***	**72.3** (68.9–78.4)***
High kurtosis	ISO-1999	60.4 (55.9-64.7)	60.4 (57.1–64.8)	51.0 (50.3–51.5)	55.3(53.8–57.2)
	SVM	**72.7** (68.0–75.8)***	**72.4** (67.5–75.6)***	**70.0** (64.8–72.9)***	**71.2** (66.2–74.2)***

04 October 2023 14:05:49

model (74.3% vs 71.0%, $p < 0.0005$). The precision of the ISO-1999 model in the G/quasi-G group is slightly higher than that of the SVM model but the difference is not significant (78.9% vs 76.7%, $p = 0.58$). The precision of the ISO-1999 model in both low and middle kurtosis groups is around 69%, but significantly decreases by 9% in the high kurtosis group. In contrast, the precision of the SVM model remains above 72% in low, middle, and high kurtosis groups. The precision of low, middle, and high kurtosis groups is significantly better using the SVM model than using the ISO-1999 model ($p < 0.0005$ for low and middle kurtosis groups, and $p < 0.0001$ for the high kurtosis group).

- Recall: The recall of the SVM model in the all-data group is significantly better than that of the ISO-1999 model (68.9% vs 51.0%, $p < 0.0001$). The recall of the ISO-1999 model over the four kurtosis groups is around 51%, which is only slightly over 50%, while the recall of the SVM model is 64.9% in the low kurtosis group and over 70% in the other three kurtosis groups. The recall of all four kurtosis groups is significantly better using the SVM model than using the ISO-1999 model ($p < 0.0001$).
- $F1$ score: The $F1$ score of the SVM model in the all-data group is significantly better than that of the ISO-1999 model (71.5% vs 59.4%, $p < 0.0001$). The highest $F1$ score of the ISO-1999 model is 62.5% in the G/quasi-G group and the $F1$ scores are less than 60% in the other three groups. In contrast, the $F1$ score of the SVM model is 68.7% in the low kurtosis group and remains above 71% in the other three kurtosis groups. The $F1$ score of all four kurtosis groups is significantly better using the SVM model than using the ISO-1999 model ($p < 0.0001$).

IV. DISCUSSION

A. The feasibility of the ISO-1999 predictive model

The ISO-1999 (2013) standard presents, in statistical terms, the relation between noise exposure and NIHL in people of various ages and having various exposure durations. The basic approach to this standard was formulated in the 1970s and incorporates a demographic database acquired even earlier. In this widely used document, all noise exposures are quantified by a single metric, that is, a time-integrated pressure-squared or energy metric incorporating a spectral weighting. The energy approach to the standard has been questioned by Bies and Hansen (1990), who showed that an alternate formulation (non-energy-based) is also consistent with the demographic database, and by Kraak (1981), whose approach is consistent with that of Bies and Hansen (1990). Similarly, there are many examples in the literature that demonstrate the limitations of the energy metric based on experimental evidence (e.g., Lei *et al.*, 1994). The ISO-1999 approach allowing age correction has also been questioned on methodological grounds by Bies and Hansen (1990) and Humes and Jesteadt (1991), and questioned on experimental grounds by Mills *et al.* (1997). It is therefore noted that, based on the above, the ISO energy-based approach is not unique, and other formulations, using the same database, will produce divergent estimates of NIHL. In addition, the enormous variability seen in the hearing loss data that were incorporated in the ISO-1999 standard (Mills *et al.*, 1996) severely limits the predictive value of the approach embodied in the standard.

The results obtained from this study using the recently collected human data ($N = 2{,}110$) in China also showed some uncertainties with the use of the ISO-1999 model to predict noise-induced hearing trauma. If accuracy is used as a model metric to assess the performance of the ISO-1999 approach, the model was around 70% accurate for G/quasi-G, low, and middle-kurtosis groups, but for the high kurtosis group, the accuracy decreased to 60.4%. The results showed that the ISO-1999 model may reasonably predict hearing impairment in workers exposed to G or low kurtosis non-G noises, but for noise with high kurtosis, its prediction accuracy decreased. The accuracy metric is a comprehensive model measure as it integrates all four predicted outputs as shown in Eq. (10). The accuracy metric can only assess the predictive model accurately when the values of FP and FN are almost the same and their costs similar. As mentioned above, the cost of FN is much higher than the cost of FP in the case of predicting hearing impairment because FN can result in the improper protection of workers' hearing. Therefore, the performance of the model needs to be further evaluated given the imbalanced costs of FP and FN.

As shown in Eq. (11), a high precision metric relates to a low FP value, and in Eq. (12), a low recall metric corresponds to a high FN value. Precision and recall metrics are compared over four kurtosis groups in the ISO-1999 model. The difference between the precision and recall metrics is significant for the G/quasi-G group (27%), the low kurtosis group (18%), and the middle kurtosis group (18%). A big difference between the precision and recall metrics suggests that the different values between the FP and FN are large, implying that the class distribution is uneven. Therefore, the ability to evaluate the accuracy of the ISO model is affected by class imbalance.

It is noteworthy that the recall metric is the preferred model metric when there is a high cost associated with the FN. The $F1$ score metric can also be useful as it is the weighted average of the precision and recall ratios. In this study, the values of the recall for all four kurtosis groups were about 51% (only slightly over 50%) in the ISO-1999 model. A low recall value in all four kurtosis groups indicates that the ISO-1999 not only underestimates hearing trauma from the high kurtosis group, but also underestimates hearing trauma from the middle kurtosis group, low kurtosis group, and even G noise group. For the $F1$ score metric, the ISO-1999 model's prediction accuracy is 62.5% for G/quasi-G group and less than 60% accurate for the other three groups. Therefore, caution must be exercised in the use of the ISO-1999 model to predict hearing impairment from noise exposure.

In this circumstance, a better predictive strategy utilizing more sophisticated statistical techniques and incorporating more essential exposure variables is needed. In this study, a machine learning model with the SVM algorithm

04 October 2023 14:05:49

was developed and a better prediction performance was obtained compared to the ISO-1999 approach.

B. The selection of features for the SVM model

In this study, a total of 42 candidate features were extracted as inputs to the SVM model. Of the 42 candidate features, 39 of them are features associated with noise waveforms, and the remaining three are demographic features, as shown in Table II. It has been demonstrated in both animal experiments and epidemiological studies that the characteristics of a noise exposure should be considered as important factors in evaluating NIHL (Qiu *et al.*, 2013; Zhao *et al.*, 2010; Xie *et al.*, 2016). The features of a noise exposure can be extracted from both the time and frequency domains. Initially, 39 features were extracted from the noise exposure as listed in Table II. These features can be classified as energy-related metrics and kurtosis-related metrics. Eventually, nine features were selected using the RF algorithm to achieve an optimal performance of the SVM model. Age was ranked as the number one feature for hearing impairment evaluation because the NIOSH defines hearing impairment as a non-age correcting criterion. Noise exposure energy and duration are undoubtedly important features for evaluating noise-induced hearing trauma. Therefore, it is not surprising that L_{Aeq}, $L_{\mathrm{eq\_500}}$, $L_{\mathrm{eq\_1000}}$, $L_{\mathrm{eq\_2000}}$, and $L_{\mathrm{eq\_4000}}$ are five of the top nine selected features given the NIOSH definition of hearing impairment. The other two selected features, i.e., L_{Feq} and the mean kurtosis, are discussed as follows.

L_{Feq}. This feature was first introduced by Sun *et al.* (2015). The basic design concept of this feature is to blend the A-weighting and the C-weighting to ensure accurate measurement of the non-G complex noise. Equation (6) gives the definition of F-weighting. An F-weighting encompasses the advantages of both the A-weighting and the C-weighting, and can adaptively evaluate complex noise based on the kurtosis and the oscillation coefficient of the noise segments. Therefore, L_{Feq} is a kurtosis-related feature. The importance of L_{Feq} in evaluating industrial noise-induced hearing impairment is confirmed in this study.

Mean kurtosis. Based on the large sample size and unique data set ($N = 2{,}110$) consisting of work shift-long noise recordings, it was found that the kurtosis metric is not fixed over the duration of the work shift but varies when a 40-s calculation window is applied. The most common distributions are log normal and normal. The question is how to extract a single number from the distribution of kurtosis that would best correlate with noise trauma. Three kurtosis-related metrics, i.e., the mean-, the median-, and the geometric mean kurtosis, were selected as candidate features. The mean kurtosis was eventually selected as one of the top nine features by the RF feature selection algorithm. The mean kurtosis was also used to categorize the noise for the collected noise database. For the ISO-1999 predictive model, the values of three model measures (i.e., accuracy, precision, and $F1$ score) decreased as the mean kurtosis of noise increased. This result indicates that (1) the ISO-1999 model is more suitable for evaluating hearing trauma due to G or quasi-G noise than for non-G complex noise with a high kurtosis value, and (2) kurtosis should be used with energy in the prediction of hearing trauma from non-G noise exposures. The experimental data used in this study strongly suggest that a kurtosis measure should be a necessary part of noise measurement and in the practice of noise evaluation.

C. The feasibility of the SVM predictive model

Industrial workers are often exposed to non-G complex noise environments. Evidence from noise studies using both animal models and human demographic data has questioned the validity of the ISO-1999 database, which was constructed using equivalent continuous sound levels (e.g., Dunn *et al.*, 1991; Lei *et al.*, 1994; Hamernik *et al.*, 2003; Qiu *et al.*, 2006; Qiu *et al.*, 2007; Taylor *et al.*, 1984; Thiery and Meyer-Bisch, 1988; Zhao *et al.*, 2010; Davis *et al.*, 2012; Xie *et al.*, 2016). The results of this study indicated that the ISO-1999 standard underestimates hearing hazard due to complex noise exposure.

In the present study, a novel classification model that combines noise and personal features with the SVM algorithm to predict hearing impairment is presented. The 9 features with the strongest correlation to hearing impairment were selected from 42 candidate features using a RF algorithm for the SVM classifier. Based on the outcomes shown in Table VI, the SVM model significantly outperforms the ISO-1999 model on all four model measures, especially on the recall and the F1-score metrics, which are crucial for assessing the good generalization performance of a predictive model. The experimental results demonstrated that the SVM model is capable of predicting hearing trauma caused by various noises due to its powerful learning ability and superior generalization capabilities. The performance of the SVM model can be further improved by collecting more data from a larger number of workers with well-documented noise exposures and personal biological information, such as blood pressure, smoking habits, and genetic disposition as it relates to the susceptibility of hearing loss.

The present study is designed to demonstrate that a statistical learning model (e.g., the SVM model) can be built that will predict, within an acceptable margin of error, the consequences of noise exposure for an individual. As a corollary, the model can also identify metrics or combinations of metrics that are necessary for accurately predicting NIHL. Such a model will offer new and more effective approaches to creating a DRC for both steady-state and complex industrial noise environments.

V. CONCLUSION

In this paper, a predictive model using the SVM algorithm was developed to evaluate hearing trauma caused by diverse industrial noises. The prediction performance of the SVM and the conventional ISO-1999 models were assessed using a large human database ($N = 2{,}110$). The results suggest that: (1) The SVM model can accurately and robustly predict hearing impairment due to various industrial noise

04 October 2023 14:05:49

exposures. (2) The ISO-1999 model could not properly predict the hearing loss caused by noise with a high kurtosis value. Moreover, the ISO-1999 model is likely to underestimate hearing impairment caused by both G and non-G complex noises. Caution must therefore be exercised in the use of the ISO-1999 model for the evaluation of noise-induced hearing impairment. (3) The kurtosis could be used as an adjunct metric to energy in the prediction of hearing trauma from a variety of noise exposures.

ACKNOWLEDGMENTS

This work was supported by Grant Nos. 200-2015-M-63857, 200-2016-M-91922 from the National Institute for Occupational Safety and Health, USA; Grant No. N00014-17-1-2198 from the Office of Naval Research, USA; Grant No. 2015C03039 from Key Research and Development Program of Zhejiang Province, China; Grant No. 81771936 from National Natural Science Foundation, China; Grant No. 2018DG0ZX01 from Major Scientific Project of Zhejiang Lab; and Grant No. 2018YFC0116901 from National Key Research and Development Program of China.

Ahroon, W. A., Hamernik, R. P., and Davis, R. I. (**1993**). "Complex noise exposures—An energy analysis," J. Acoust. Soc. Am. **93**(2), 997–1006.

Bies, D. A., and Hansen, C. H. (**1990**). "An alternative mathematical description of the relationship between noise exposure and hearing loss," J. Acoust. Soc. Am. **88**, 2743–2754.

Bishop, C. M. (**2006**). *Pattern Recognition and Machine Learning* (Springer, New York), pp. 291–320.

Burns, W., and Robinson, D. W. (**1970**). "Hearing and noise in industry," (HMSO, London).

Committee on Hearing, Bioacoustics, and Biomechanics (CHABA). (**1992**). *Hazardous Exposure to Impulse Noise* (National Research Council, Washington, DC).

Davis, R. I., Qiu, W., Heyer, N. J., Zhao, Y. M., Yang, Q. L., Li, N., Tao, L. Y., Zhu, L. L., Zeng, L., and Yao, D. H. (**2012**). "The use of the kurtosis metric in the evaluation of occupational hearing loss in workers in China: Implications for hearing risk assessment," Noise Health **14**(61), 330–342.

Demsar, J. (**2006**). "Statistical comparisons of classifiers over multiple data sets," J. Mach. Learn. Res. **7**, 1–30.

Dunn, D. E., Davis, R. R., Merry, C. J., and Franks, J. R. (**1991**). "Hearing-loss in the chinchilla from impact and continuous noise exposure," J. Acoust. Soc. Am. **90**(4), 1979–1985.

Hamernik, R. P., and Qiu, W. (**2001**). "Energy-independent factors influencing noise-induced hearing loss in the chinchilla model," J. Acoust. Soc. Am. **110**(6), 3163–3168.

Hamernik, R. P., Qiu, W., and 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**(1), 386–395.

Havelock, D., Kuwano, S., and, Vorländer, M. (**2008**). *Handbook of Signal Processing in Acoustics* (Springer, New York), pp. 107–121.

Humes, L. E., and Jesteadt, W. (**1991**). "Modeling the interactions between noise exposure and other variables," J. Acoust. Soc. Am. **90**, 182–188.

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

Johnson D. L. (**1994**). "The development of the data base for ISO-1999, an evaluation of its strengths and weaknesses," J. Acoust. Soc. Am. **95**(5), 2861.

Kardous, C. A., Franks, J. R., and Davis, R. R. (**2005**). "NIOSH/NHCA best-practices workshop on impulsive noise," Noise Control Eng. J. **53**(2), 53–60.

Kraak, W. (**1981**). "Investigations on criteria for the risk of hearing loss due to noise," in *Hearing Research and Theory*, edited by J. V. Tobias and E. D. Schubert (Academic, New York), Vol. I, pp. 187–303.

Lei, S. F., Ahroon, W. A., and Hamernik, R. P. (**1994**). "The application of frequency and time-domain kurtosis to the assessment of hazardous noise exposures," J. Acoust. Soc. Am. **96**(3), 1435–1444.

Louppe, G., Wehenkel, L., Sutera, A., and Geurts, P. (**2013**). "Understanding variable importances in forests of randomized trees," in *Advances in Neural Information Processing Systems*, pp. 431–439.

Martin, A. (**1976**). "The equal energy concept applied to impulse noise," in *Effects of Noise on Hearing*, edited by D. Henderson, R. P. Hamernik, D. S. Dosanjh, and J. H. Mills (Raven, New York), pp. 421–453.

Mills, J. H., Boettcher, F. A., and Dubno, J. R. (**1997**). "Interaction of noise-induced permanent threshold shift and age-related threshold shift," J. Acoust. Soc. Am. **101**, 1681–1686.

Mills, J. H., Lee, F.-S., Dubno, J. R., and Boettcher, F. A. (**1996**). *Interactions between Age-Related and Noise-Induced Hearing Loss* (Thieme Medical Publishers, Inc., New York), Chap. 17, pp. 193–212.

NIOSH (**1998**). "Criteria for a recommended standard: Occupational exposure to noise—Revised criteria 1998," available at https://www.cdc.gov/niosh/docs/98-126/ (Last viewed June 6, 2014).

Passchier-Vermeer, W. (**1974**). "Hearing loss due to continuous exposure to steady-state broad-band noise," J. Acoust. Soc. Am. **56**, 1585–1593.

Passchier-Vermeer, W. (**1977**). "Hearing levels of non-noise exposed subjects and of subjects exposed to constant noise during working hours," Report B367 (Research Institute for Environmental Hygiene, Netherlands).

Patro, S., and Sahu, K. K. (**2015**). "Normalization: A preprocessing stage," arXiv:1503.06462.

Patterson, J. H., Hamernik, R. P., Hargett, C. E., and Ahroon, W. A. (**1993**). "An isohazard function for impulse noise," J. Acoust. Soc. Am. **93**(5), 2860–2869.

Qiu, W., Davis, B., and 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**(3), 1613–1620.

Qiu, W., Hamernik, R. P., and 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**(6), 3901–3906.

Qiu, W., Hamernik, R. P., and 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**(5), 2856–2866.

Sun, P. F., Qin, J., and Qiu, W. (**2015**). "Development and validation of a new adaptive weighting for auditory risk assessment of complex noise," Appl. Acoust. **103**, 30–36.

Taylor, W., Lempert, B., Pelmear, P., Hemstock, I., and Kershaw, J. (**1984**). "Noise levels and hearing thresholds in the drop forging industry," J. Acoust. Soc. Am. **76**, 807–819.

Thiery, L., and Meyer-Bisch, C. (**1988**). "Hearing-loss due to partly impulsive industrial noise exposure at levels between 87-Db(a) and 90-Db(a)," J. Acoust. Soc. Am. **84**(2), 651–659.

Vapnik, V. (**2013**). *The Nature of Statistical Learning Theory* (Springer, New York), pp. 123–265.

Vihinen, M. (**2012**). "How to evaluate performance of prediction methods? Measures and their interpretation in variation effect analysis," BMC Genomics **13**(Suppl 4), S2.

Woodford, C., and Phillips, C. (**2011**). *Numerical Methods with Worked Examples* (Chapman and Hall, London, UK).

Xie, H., Qiu, W., Heyer, N. J., Zhang, M., Zhang, P., Zhao, Y., and Hamernik RP. (**2016**). "The use of the kurtosis-adjusted cumulative noise exposure metric in evaluating the hearing loss risk for complex noise," Ear Hear. **37**(3), 312–323.

Zhao, Y. M., Qiu, W., Zeng, L., Chen, S. S., Cheng, X. R., Davis, R. I., and Hamernik, R. P. (**2010**). "Application of the kurtosis dtatistic to the rvaluation of the tisk of hearing loss in workers exposed to high-level complex noise," Ear Hear. **31**(4), 527–532.

04 October 2023 14:05:49