Tech Science Press

Using Audiometric Data to Weigh and Prioritize Factors that Affect Workers' Hearing Loss through Support Vector Machine (SVM) Algorithm

Hossein ElahiShirvan[1], MohammadReza Ghotbi-Ravandi[2], Sajad Zare[3,*] and Mostafa Ghazizadeh Ahsaee[4]

[1]Student Research Committee, Kerman University of Medical Sciences, Kerman, Iran
[2]Associate Professor, Department of Occupational Health, School of Public Health, Kerman University of Medical Sciences, Kerman, Iran
[3]Assistant Professor, Department of Occupational Health, School of Public Health, Kerman University of Medical Sciences, Kerman, Iran
[4]Assistant Professor, Department of Computer Engineering, Shahid Bahonar University, Kerman, Iran
*Corresponding Author: Sajad Zare. Email: ss_zare87@yahoo.com
Received: 15 October 2019; Accepted: 10 December 2019

Abstract: Workers' exposure to excessive noise is a big universal work-related challenges. One of the major consequences of exposure to noise is permanent or transient hearing loss. The current study sought to utilize audiometric data to weigh and prioritize the factors affecting workers' hearing loss based using the Support Vector Machine (SVM) algorithm. This cross sectional-descriptive study was conducted in 2017 in a mining industry in southeast Iran. The participating workers (n = 150) were divided into three groups of 50 based on the sound pressure level to which they were exposed (two experimental groups and one control group). Audiometric tests were carried out for all members of each group. The study generally entailed the following steps: (1) selecting predicting variables to weigh and prioritize factors affecting hearing loss; (2) conducting audiometric tests and assessing permanent hearing loss in each ear and then evaluating total hearing loss; (3) categorizing different types of hearing loss; (4) weighing and prioritizing factors that affect hearing loss based on the SVM algorithm; and (5) assessing the error rate and accuracy of the models. The collected data were fed into SPSS 18, followed by conducting linear regression and paired samples t-test. It was revealed that, in the first model (SPL < 70 dBA), the frequency of 8 KHz had the greatest impact (with a weight of 33%), while noise had the smallest influence (with a weight of 5%). The accuracy of this model was 100%. In the second model (70 < SPL < 80 dBA), the frequency of 4 KHz had the most profound effect (with a weight of 21%), whereas the frequency of 250 Hz had the lowest impact (with a weight of 6%). The accuracy of this model was 100% too. In the third model (SPL > 85 dBA), the frequency of 4 KHz had the highest impact (with a weight of 22%), while the frequency of 250 Hz had the smallest influence (with a weight of 3%). The accuracy of this model was 100% too. In the fourth model, the frequency of 4 KHz had the greatest effect (with a weight of 24%), while the frequency of 500 Hz had the smallest effect (with a weight of 4%). The accuracy of this model was found to be 94%. According to the modeling conducted using the

SVM algorithm, the frequency of 4 KHz has the most profound effect on predicting changes in hearing loss. Given the high accuracy of the obtained model, this algorithm is an appropriate and powerful tool to predict and model hearing loss.

Keywords: Noise; modeling; hearing loss; data mining; support vector machine algorithm

1 Introduction

Noise exposure is a common consequence of industrialization that there is no significant relationship between its pressure, frequency, and wavelength ranges; These types of noises (that have no significant relationship between its characteristics (pressure, frequency, and wavelength ranges)), are produced and emitted considerably in different industries [1]. In fact, exposure to excessive noise has become a widely-blamed harmful agent and a major risk factor among workers in various industries [2]. Numerous studies have shown that harmful factors in the workplace such as noise, vibration, and shift work and so on have detrimental effects on workers' health [3–6]. The effects of environmental noise pollution have become even more profound because of the speedy growth of economy and urbanization [7].

In the US, over 30 million workers are exposed to dangerous noises (85 dB) (A-weighted noise exposure level normalized to an 8 hour working day) and 7.4–10 million industrial workers are at the risk of hearing loss caused by occupational noise [8]. In Michigan only, around 86000 people are suffering from noise-induced hearing loss. Although this hazard can be prevented, hearing loss is a common occupational disease in the US [9]. In 1990, $200 million were paid in compensation for hearing loss caused by noise exposure [10]. On the other hand, over the past 10 years, the proportion of the European population at the risk of exposure to noises above 65 dB has increased from 15% to 26% [11].

Exposure to excessive noise will lead to physiological (high blood pressure, adrenaline secretion, more likelihood of heart attack, change in respiratory rate and the amount of consumed oxygen, deterioration of the auditory system, and increase in stomach and intestinal activities) as well as psychological, social, and economic complications. Furthermore, work efficiency tends to decline in people who are exposed to excessive noise. Such individuals are also more likely to face problems in their communicative abilities and understanding of warning/safety sings [12,13]. Many cases of absenteeism in industrial environments and workers' report of chronic fatigue and suffering in many indicators of physical health can further be attributed to excessive noise exposure. Research has also revealed that people who are in contact with harmful noise demonstrate impaired psychological responses, especially in stressful situations, and report reduction in the quality and quantity of their sleep [14]. One of the major consequences of exposure to excessive noise is permanent or transient hearing loss [8].

Noise-induced hearing loss commonly occurs in the first 10–15 years of work if workers are in contact with high frequency noises (over 4000 Hz). The degree of the influence of excessive noise exposure depends on various personal and environmental factors. Noise-induced hearing loss classically affects frequencies less than 3, 4, and 6 KHz first. If exposure to noise prolongs for longer times, individuals' ability in hearing higher frequency noises will be impaired too. Thus, audiometry should be conducted for the definitive diagnosis of hearing loss [9]. Age and work experience are believed to influence both permanent and transient noise-induced hearing loss [10].

Audiometry is a method for assessing people's hearing sensitivity, hence shedding light on the nature, degree, and probable causes of hearing impairment [15,16]. Using this technique, auditory stimuli with different intensity levels are presented to the person and his/her responses are recorded. The hearing threshold is determined in the light of the minimum intensity level stimulus to which the person responds consistently. The output of this diagnostic test, known as audiogram, can be used for adopting appropriate treatments and hearing aids [16,17].

Data mining (DM), which is a subfield of computer science and is located at the intersection of artificial intelligence, machine learning, statistics, and database systems, is a process for detecting patterns in large data sets. DM is intended to extract information from a set of data and transform it to an easily understandable structure. The extracted information may include categorization and estimation of outcomes after an intervention, detection of associations among different variables, or prediction of deviations [18]. In the current study, DM was adopted for two reasons: first, DM is different from applying algorithms blindly, hence the usefulness of the extracted information. Thus, successful application of DM needs appropriate knowledge within the studied domain and clearly stated objectives [19]. Second, the use of DM in new areas usually casts light on interesting issues that are unknown to commonly used techniques. This presents opportunities for useful algorithmic development and extensions [20].

Over the past years, since support vector machines (SVM) has demonstrated appropriate generalization performance in a wide array of learning problems (e.g., handwritten digit recognition, classification of web pages, and face detection), researchers in the community of machine learning have become interested in its application. Problems like multilocal minima, curse of dimensionality, and neural network overfitting are rarely observed in SVM. Further, originating from statistical learning theory, SVM has a rigid theoretical basis. Nonetheless, training SVM is still a challenge, particularly for a large-scale learning problem. As a result, it is necessary to devise a fast training SVM algorithm to be applied to various engineering problems in our field [21].

Since hearing loss is a widespread disorder caused by excessive noise, it imposes huge costs on industrial workers or individuals who are in contact with noise. However, few studies have tried to weigh and prioritize factors affecting workers' hearing loss based on audiometric findings by the use of SVM algorithm. Therefore, the present study aimed to:

1- Determine workers' equivalent sound level and other predictor factors.
2- Assess the hearing loss of both ears.
3- Weigh and prioritize factors that may affect hearing loss based on SVM algorithm.
4- Estimate the error rate and accuracy of the models emerging from the SVM algorithm.

2 Materials and Methods

2.1 Sampling Procedure

The study was conducted in a mining industry in southeast Iran. In the light of individuals' equivalent sound level, the findings of previous studies, and the type of algorithm used in modeling hearing loss, three groups (a control group in the office that was exposed to a low equivalent sound level and two case groups that were selected from the workshop and were exposed to high equivalent sound level) were generally involved in the study. Based on the sound equivalent level of the three groups, fifty participants were included in each group, hence having 150 participants in total [22].

2.2 Research Design

This study adopted a cross-sectional, descriptive, analytical, prospective design. It generally entailed the following steps: (1) selecting predicting variables to weigh and prioritize factors affecting hearing loss; (2) conducting audiometric tests and assessing permanent hearing loss in each ear and then evaluating total hearing loss; (3) categorizing different types of hearing loss; (4) weighing and prioritizing factors that affect hearing loss based on the SVM algorithm; and (5) assessing the error rate and accuracy of the models [23].

2.3 Selecting the Variables Influencing Hearing Loss (PREDICTORS)

Age, work experience, equivalent sound level, and frequency were taken into consideration for weighing and prioritizing factors that may influence hearing loss among individuals [23,24]. All the participants were adults with three age ranges, namely 20–35, 35–50, and over 50 years old [23]. With regard to their work experience, the participants were divided into three groups too: less than 10 years of experience, between 10 and 20 years of experience, and over 20 years of experience [22]. Following ISO 9612, equivalent sound level was measured using a TES-1345 dosimeter manufactured in Taiwan. Before using the machine, a CEL 110/2 calibrator (made in the UK) was utilized to calibrate the dosimeter [25,26]. Pure-toe hearing thresholds were recorded at 250, 500, 1000, 2000, 4000, 6000 and 8000 Hz using a CA 120 clinical audiometer manufactured in Denmark [27,28]. Upon eliminating the effect of age, hearing loss estimates in four frequencies (500, 1000, 2000, and 4000 Hz) were entered into Eq. (1), hence computing noise-induced hearing loss (NIHL). Then, Eq. (2) was exploited to calculate permanent hearing loss [28,29].

$$NIHL = \frac{(TL_{500\ Hz}) + (TL_{1000\ Hz}) + (TL_{2000\ Hz}) + (TL_{4000\ Hz})}{4} \tag{1}$$

TL: the hearing loss of each ear in a particular frequency (dB)

NIHL: noise-induced hearing loss (dB)

$$NIHL_t = \frac{(NIHL_b \times 5) + (NIHL_p)}{6} \tag{2}$$

$NIHL_t$: the permanent hearing loss in both ears (dB)

$NIHL_b$: the permanent hearing loss of the better ear (dB)

$NIHL_p$: the permanent hearing loss of the poor ear (dB)

According to the classification proposed by the World Health Organization (WHO), the hearing ability in the range of 0–25 dBA is regarded as normal hearing, while that in the range of 40–26 dBA is known as mild injury. In addition, people with hearing thresholds of 41–60 dBA, 61–80 dBA, and over 80 dBA are believed to suffer from moderate, severe, and profound hearing loss respectively [30].

2.4 Weighing and Prioritizing Factors Affecting Hearing Loss Using SVM

Finally, modeling the hearing loss of workers was done by entering all information about age, work experience, equivalent sound level and hearing loss frequencies in IBM SPSS Modeler 18.0. The algorithm's function is described as follows.

2.4.1 Support Vector Machine Algorithm

A major way through which people perceive the world and receive the required knowledge is learning from the data. Let $\{x_i, y_i\}$, $i = 1, \ldots, l$, $y_i \in \{-1, 1\}$ and $x_i \in IR^n$ be the training samples where x_i is the training vector and y_i is its corresponding target value. To have consistent notations, lowercase bold letters are used to refer to column vectors (for example, x_i), while uppercase bold letters are used to show matrices. The notation $(A)_{ij}$ is used to refer to the (i th) rowand the (j th) column element of matrix (A) [31]. According to Boser et al., the training Support Vector Machine for a pattern recognition problem can be formulated as the quadratic optimization problem [32]. Its function is showed in Eq. (3):

$$maximize : \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\alpha^T Q\alpha \tag{3}$$

$$subject : 0 \leq \alpha_i \leq C,$$
$$\sum_{i=1}^{l} y_i \alpha_i = 0, \qquad\qquad i = 1, ..., l$$

where α is a vector of length l and its component α_i corresponds to a training sample $\{x_i, y_i\}$, Q is a $l \times l$ semidefinite kernel matrix, and C is a parameter chosen by the user. A larger C assigns a higher penalty to the training errors. The training vector x_i whose corresponding α_i is nonzero is known as support vector. Support vector machine maps training vector x_i into a high-dimensional feature space by the function $\Phi(x)$ such that $(Q)_{ij} = y_i y_j (K)_{ij} = y_i y_j K(x_i, x_j)$ and $K(x_i, x_j) = \Phi^T(x_i)\Phi(x_j)$. When the above optimization problem is solved, we can obtain an optimal hyperplane in a high-dimensional feature space to make a distinction between the two-class samples. The decision function is given by Eqs. (4) and (5):

$$f(x) = \text{sgn}\left(\sum_{i=1}^{l} y_i \alpha_i K(x_i, x) - b \right), \tag{4}$$

where

$$\text{sgn}(u) = \begin{pmatrix} 1, & & u > 0, \\ -1, & for & u < 0. \end{pmatrix} \tag{5}$$

In the latter algorithm, a technique presented by Keerthi et al. is used to choose the two variables for optimization and determine the stopping conditions. First, training patterns are split into five sets [33,34]:

1. $I_0 = \{i : 0 < \alpha_i < C\}$,
2. $I_1 = \{i : y_i = 1, \alpha_i = 0\}$,
3. $I_2 = \{i : y_i = -1, \alpha_i = C\}$,
4. $I_3 = \{i : y_i = 1, \alpha_i = C\}$,
5. $I_4 = \{i : y_i = -1, \alpha_i = 0\}$.

Then, we define them in Eq. (6):

$$b_{up} = F_{i-up} = \min\{F_i : i \in I_0 \cup I_1 \cup I_2\},$$
$$b_{low} = F_{i-low} = \max\{F_i : i \in I_0 \cup I_3 \cup I_4\}, \tag{6}$$

where $F_i = \sum_{j=1}^{l} y_j \alpha_j K(x_j, x_i) - y_i$. As demonstrated by Keerthi et al., conditions for an optimal solution to (Eq. (3)) hold at some α if and only if $b_{low} \leq b_{up}$. Additionally, the worst violating pair of patterns $(i - up, i - low)$ are chosen for optimization, which may result in a measurable rise in the objective function [33].

2.4.2 Assessing the Accuracy of the Model Generated by SVM Algorithm

In categorization algorithms, which are used for classifying discrete output variables, criteria such as accuracy, confusion matrix, sensitivity, features, etc. are used for conducting assessment. In the current study, accuracy and confusion matrix were utilized. Confusion matrix is a square matrix whose dimensions are equal to the number of output categories. In this matrix, the main diameter indicates the percentage of accurate predictions. According to Eq. (7), accuracy of the model is obtained by dividing the percentage of correct predictions by all predictions [35].

$$Accuracy = \frac{\textbf{True Postive cases} + \textit{True Negative cases}}{\textbf{All cases}} \tag{7}$$

2.5 Ethical Considerations

Ethical approval was sought from the Ethics Committee of Kerman University of Medical Sciences (ID: IR.KMU.REC. 1396.2458). In addition, all the participants signed a written consent form prior to the study.

2.6 Data Analysis

The collected data were fed into the Statistical Package for Social Sciences (SPSS) version 18 (SPSS, Inc., Chicago, Illinois, USA).The obtained mean scores, standard deviations, and correlation coefficients were analyzed using linear regression and paired samples t-test. Also, IBM SPSS Modeler 18.0 was used to model changes in hearing loss.

3 Results

3.1 Demographic Information

Tab. 1 demonstrates the participants' demographic information.

Table 1: Demographic information of the study sample (n = 150)

	Variables	Mean	SD*
The first group (n = 50) (SPL < 70 dBA)	Age	37.66	9.91
	Work Experience	9.1	4.9
The second group (n = 50) (SPL 70–80 dBA)	Age	35.56	11.45
	Work Experience	8.48	5.38
The Third group (n = 50) (SPL > 85 dBA)	Age	41.76	10.93
	Work Experience	11.34	5.32

* SD = Standard deviation

3.2 Measuring Equal Sound Level

The participants in the first group were exposed to equal sound level of less than 70 dB, while those in the second and third groups were exposed to equal sound levels of 70–80 dB and over 85 dB respectively. The mean scores and standard deviations of equivalent sound level for the three groups respective were 70 ± 3 dB, 77.62 ± 4.43 dB, and 89.7 ± 3.03 dB.

3.3 The Results of Hearing Loss

Tab. 2 displays the results related to hearing loss of workers' both ears. The results are divided into 5 categories in the light of the degree of hearing loss. The results of paired samples t-test showed no significant difference in the hearing loss of right and left ears of participants in the three groups in similar frequencies ($P > 0.05$).

3.4 Modeling Hearing Loss Changes

In this study, four different models of hearing loss changes were calculated. The first model represents the audiometric data of the first group (SPL < 70 dBA), the second model entails the audiometric data of the second group (70 < SPL < 80 dBA), the third model demonstrates the audiometric data of workers in the thrid group (SPL > 85 dBA), and the fourth model is based on the audiometric data of all participating workers.

Table 2: Classification the participants' hearing loss (n = 150)

	Normal (0–25 dB)	Mild (26–40 dBA)	Moderate (41–60 dBA)	Severe (61–80 dBA)	Profound (80 dBA <)
The first group (n = 50) (SPL < 70 dBA)	40 participants (80%)	10 participants (20%)	–	–	–
The second group (n = 50) (SPL 70–80 dBA)	37 participants (74%)	10 participants (20%)	3 participants (6%)	–	–
The Third group (n = 50) (SPL > 85 dBA)	29 Participant (58%)	15 Participant (30%)	4 Participant (8%)	2 Participant (4%)	–

3.4.1 The First Model: Modeling Hearing Loss Changes Based on Audiometric Data of Workers in the First Group (SPL < 70 dBA)

Fig. 1 demonstrates the output model presented by IBM SPSS Modeler 18.0 for the hearing loss changes of workers in the first group. It is observed that the frequency of 8 KHz with a weight of 33% had the greatest influence, while noise (with a weight of 5%) had the lowest impact.
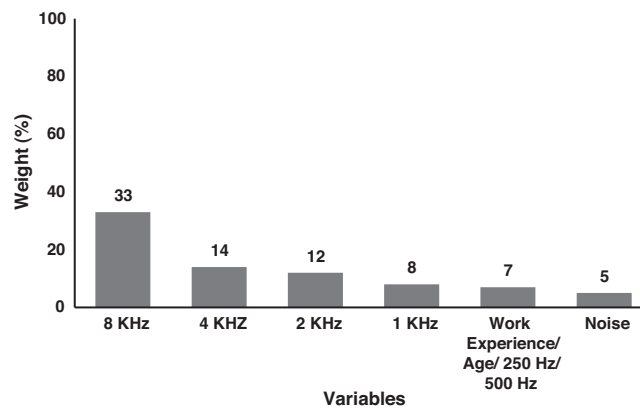
Figure 1: The first model: the weight (%) of hearing loss predicting variables for participants in the first group (SPL < 70 dBA)

Tab. 3 illustrates the results of the corresponding confusion matrix. The model presented an accurate prediction for all workers whose hearing loss was normal or mild. The accuracy of SVM algorithm in this modeling was equal to 100%.
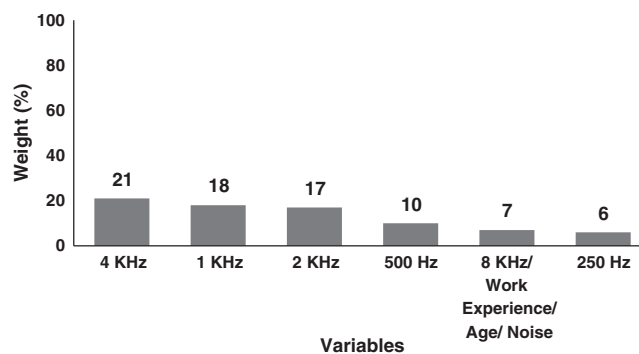
3.4.2 The Second Model: Modeling Hearing Loss Changes Based on Audiometric Data of Workers in the Second Group (70 < SPL < 80 dBA)

Fig. 2 demonstrates the output model presented for the hearing loss changes of workers in the second group. As observed, the frequencies of 4 KHz (with a weight of 21%) and 1 KHz (with a weight of 18%) had the highest impact. Frequencies of 2 KHz and 500 Hz came third and fourth in that order. Finally, the frequency of 250 KHz (with a weight of 6%) had the lowest influence.

This model accurately predicted the hearing loss of workers in the normal, mild, and moderate categories. The accuracy of SVM algorithm in this modeling was found to be 100%. The results of the confusion matrix corresponding to the algorithm of the second model are presented in Tab. 4.

Table 3: The confusion matrix of SVM algorithm for the first model

Observed	Predicted Severity of NIHL	
	Normal	Mild
Normal	100%	0.0%
Mild	0.0%	100%

Figure 2: The second model: the weight (%) of hearing loss predicting variables for participants in the second group (70 < SPL < 80 dBA)

Table 4: The confusion matrix of SVM algorithm for the second model

Observed	Predicted Severity of NIHL		
	Normal	Mild	Moderate
Normal	100%	0.0%	0.0%
Mild	0.0%	100%	0.0%
Moderate	0.0%	0.0%	100%

3.4.3 The Third Model: Modeling Hearing Loss Changes Based on Audiometric Data of Workers in the Third Group (SPL > 85 dBA)

The results of modeling hearing loss changes among the members of the third group (SPL > 85 dBA) are demonstrated in Fig. 3. The frequency of 4 KHz had the biggest impact with a weight of 22%. Age, work experience, and the frequency of 2 KHz equally occupied the second position (with weights of 13% each). Finally, the frequency of 250 KHz had the smallest impact with a weight of 3%.

The results of the corresponding confusion matrix are presented in Tab. 5. Accordingly, the model accurately predicts hearing loss among workers with normal, mild, moderate, and severe hearing loss. The accuracy of SVM algorithm in this modeling is 100%.

3.4.4 The Fourth Model: Modeling Hearing Loss Changes Based on Audiometric Data of Workers in all Groups

The results of the fourth model (which included the data of all the 150 participants) are presented in Fig. 4. It is observed that the frequency of 4 KHz (with a weight of 24%) had the greatest impact on
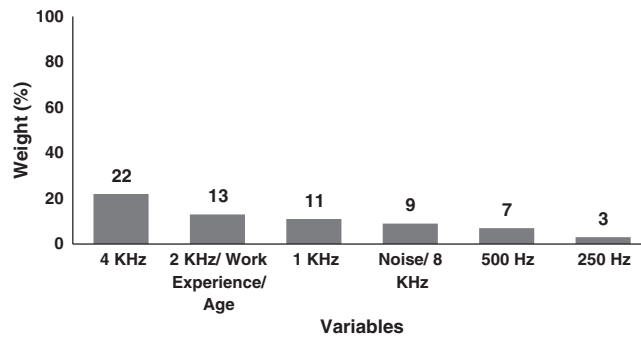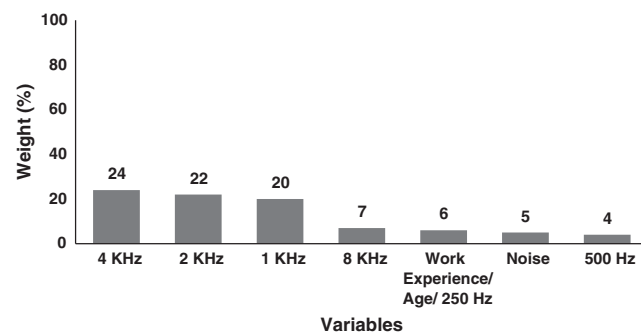
Figure 3: The third model: the weight (%) of hearing loss predicting variables for participants in the third group (SPL > 85 dBA)

Table 5: The confusion matrix of SVM algorithm for the third model

Observed	Predicted Severity of NIHL			
	Normal	Mild	Moderate	Severe
Normal	100%	0.0%	0.0%	0.0%
Mild	0.0%	100%	0.0%	0.0%
Moderate	0.0%	0.0%	100%	0.0%
Severe	0.0%	0.0%	0.0%	100%

Figure 4: The fourth model: the weight (%) of hearing loss predicting variables for participants in the three groups

hearing loss, followed by the frequencies of 2 KHz, 1 KHz, and 8 KHz in that order. Conversely, the frequency of 500 Hz had the smallest influence (with a weight of 4%).

Tab. 6 contains the results of the corresponding confusion matrix. Accordingly, the model accurately predicts hearing loss among workers with normal, moderate, and severe hearing loss. However, with regard to workers with mild hearing loss, the model predicted that 25.71% of them had normal in terms of their hearing ability and the rest (74.28%) had mild hearing loss. The algorithm accuracy in this regard is 94%, with an error rate of 6%.

Table 6: The confusion matrix of SVM algorithm for the fourth model

Observed	Predicted Severity of NIHL			
	Normal	Mild	Moderate	Severe
Normal	100%	0.0%	0.0%	0.0%
Mild	25.71%	74.28%	0.0%	0.0%
Moderate	0.0%	0.0%	100%	0.0%
Severe	0.0%	0.0%	0.0%	100%

4 Discussion

This study used SVM algorithm to weigh and prioritize the factors that affect workers' hearing loss based on audiometric findings. The factors such as age, work experience, A-weighted equivalent sound pressure levels and frequencies were considered predictor factors. There were severity hearing loss (target factor) in the second and third groups. The results showed that the average exposure to equal level sound for the first, second, and third groups were 70 ± 3 dBA, 77.62 ± 4.43 dBA, and 89.7 ± 3.03 dBA respectively. It was also discovered that, in the first two groups, age and work experience had significant correlations with hearing loss. More specifically, the correlation coefficients for the relationship between age and hearing loss were r = 0.385 ($P = 0.008$) for the first group and r = 0.394 ($P = 0.008$) for the second one. Additionally, the correlation coefficients for the relationship between work experience and hearing loss were r = 0.362 ($P = 0.014$) for the first group and r = 0.32 ($P = 0.038$) for the second group. However, no significant association was detected between age/experience and hearing loss in the third group. That is, the correlation coefficient of the relationship between age and hearing loss was r = 0.189 ($P = 0.277$), while that of the association between experience and hearing loss was r = 0.28 ($P = 0.076$). The results of Pearson correlation and linear regression further revealed a statistically considerable relationship between noise exposure and hearing loss for all the participants (n = 150) (r = 0.414, $P = 0.0001$).

In a descriptive, analytical, cross-sectional study, Halvani et al. [9] examined the correlation between noise exposure and hearing loss among the workers of a textile factor, Taban-e-Yazd. They investigated hearing loss in the left and right ears of 100 workers who worked in spinning, knitting, and mechanical sectors of the factory. In this study, the workers were exposed to noises with frequencies of 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz, and 8000 Hz and their hearing ability was measured using audiometric tests. The results showed that, keeping the work experience constant, a rise of one unit in the sound pressure level would lead to 18% noise-induced hearing loss. Also, a rise of one unit in the work experience would result in 37% noise-induced hearing loss [9]. In line with these findings, the results of the current study also indicated that the likelihood of hearing loss goes up as a result of increasing work experience. In another study, Tajic et al. (2008) examined the effect of noise pollution on the auditory system of workers in a metal factory in Arak. The participants worked in the sheet metal manufacturing, assembly, and welding sectors. The results demonstrated that the highest degree of hearing loss was recorded among workers in the age range of 41–50 years with 21–30 years of work experience. They also discovered significant relationships between hearing loss, on the one hand, and sound pressure level, age, and work experience, on the other [36]. The results of the present study are in agreement with the findings reported by Tajic et al. [36], meaning that the rise of age, work experience, and sound pressure level would lead to higher possibility of hearing loss.

According to the models generated by the SVM algorithm, in the first model (SPL < 70 dBA), the frequency of 8 KHz had the highest weight (33%), whereas noise had the smallest weight (5%) (Fig. 1). This model accurately predicted the hearing loss of people with normal hearing ability or mild hearing

loss, with the accuracy of the SVM model being 100% (Tab. 3). (It means 0.33 8 KHz + 0.14 4 KHz + 0.12 2 KHz + 0.08 1 KHz + 0.07 work experience & age & 250 Hz & 500 Hz + 0.05 noise = 100% normal hearing loss). In the second model (70 < SPL < 80 dBA), the frequency of 4 KHz had the most profound effect on hearing loss (with a weight of 21%), while the frequency of 250 Hz exerted the smallest amount of influence (with a weight of 6%) (Fig. 2). Also, the accuracy of the SVM algorithm was 100% (Tab. 4). (It means 0.21 4 KHz + 0.18 1 KHz + 0.17 2 KHz + 0.1 500 Hz + 0.07 work experience & age & noise & 8 KHz + 0.06 250 Hz = 100% normal hearing loss). In the third model (SPL > 85 dBA), the biggest influence on hearing loss belonged to the frequency of 4 KHz (with a weight of 22%), while the smallest one was observed in the case of work experience (with a weight of 6%) (Fig. 3). This model accurately predicted hearing loss among people with normal hearing ability as well as individuals with mild, moderate, and severe hearing loss. The accuracy of the corresponding algorithm was 100% (Tab. 5). (It means 0.22 4 KHz + 0.13 2 KHz & work experience & age + 0.11 1 KHz + 0.09 8 KHz & noise + 0.07 500 Hz + 0.03 250 Hz = 100% normal or mild hearing loss). In the fourth model, the frequency of 4 KHz exerted the greatest influence on hearing loss (with a weight of 24%), while the frequency of 500 Hz had the smallest impact (with a weight of 4%) (Fig. 4). This model accurately predicted the degree of hearing loss among people with normal hearing ability as well as individuals with moderate or severe hearing loss. With regard to workers with mild hearing loss, the model predicted that 25.71% of them enjoyed normal hearing ability and the rest (74.28%) had mild hearing loss. The accuracy of the corresponding algorithm was 94% and the error rate was 6% (Tab. 6). (It means 0.24 4 KHz + 0.22 2 KHz + 0.2 1 KHz + 0.07 8 KHz + 0.06 250 Hz & work experience & age + 0.05 noise + 0.04 500 Hz = 94% normal or mild hearing loss).

Dubno et al. [37] aimed to categorize audiometric phenotypes caused by aging through using animal models. They gathered audiograms from 338 samples classified into four phenotypes (older normal, metabolic, sensory, metabolic + sensory). Accuracy of the QDA analysis, SVM algorithm, and RF were found to be 93.2%, 89.9%, and 89.3% respectively. The accuracy indices obtained in their study is relatively low compared to the ones obtained in the current research, in which the accuracy of the first three models was 100% and the accuracy of the fourth one was 94%. Acir et al. [38] conducted a study entitled "automatic classification of auditory brainstem responses using SVM-based feature selection algorithm for threshold detection". In their study, the findings were classified according to the SVM algorithm, with the reported accuracy being 96.2%. The accuracy obtained in the current study is as high as the one reported by Acir et al. [38].

Wang et al. [39] carried out another study entitled "wavelet entropy and directed acyclic graph support vector machine for detection of patients with unilateral hearing loss in MRI scanning". They scanned 49 patients using MRI and subsequently classified them into three groups; the first group included 14 people with right ear hearing loss (RHL), the second one included 15 members with left ear hearing loss (LHL), and the third group, which comprised 20 individuals, functioned as the control group (HC). They selected wavelet entropy (WE) from the MRI of each participant and subsequently submitted it to a directed acyclic graph support vector machine method (DAG-SVM). The accuracy of DAG-SVM method for the RHL, LHL and HC groups were found to be 97.14%, 96.33%, and 96.73% respectively [39]. The accuracy indices obtained in the current study are high (close to 100%), like the ones obtained by Wang et al. [39] Taylor et al. [40] conducted another study entitled "Audio Gene: predicting hearing loss genotypes from phenotypes to guide genetic screening". They intended to describe the algorithm underlying Audio Gene by employing a software system which utilized machine learning techniques. These techniques used phenotypic information derived from audiograms to estimate the genetic cause of hearing loss in people segregating ADNSHL. The results indicated that Audio Gene had an accuracy of 68% in predicting the causative gene within its top three predictions. In contrast, the accuracy reported for a majority classifier was found to be 44%. The accuracy indices reported in their study was relatively low, while the ones obtained in the current research were high.

Majumder et al. [22] used both unsupervised (i.e., Expectation Maximization (EM), K-means, Linear Vector Quantization (LVQ), and Self Organization Map (SOM)) and supervised (i.e., Naïve Bayes, Instance-based (IB), Back Propagation Network (BPN), and Radial Basis Function (RBF)) data mining procedures to model hearing loss changes using audiometric data among professional drivers. They concluded that, save for RBF, all data mining algorithms demonstrated relatively high adaptation and performance for the right ear. Likewise, in the current study, the models generated by the SVM algorithm showed high accuracy. Noma et al. [41] intended to predict hearing loss symptom using audiometric data and FP-Growth technique. The findings indicated that, in five different models, the accuracy rates were 100%, 99.5%, 98.25%, and 94.6% with frequencies higher than 10. In this study, save for the last algorithm, all the other FP-Growth algorithms had high accuracy. Likewise, in the current study, the accuracy of the first three models was 100%, while that of the last one was 94%.

In this study, some variables that had not been investigated in previous research were weighed and prioritized. They included various sound pressure levels, frequencies (250 Hz, 500 Hz, 1 KHz, 2 KHz, 4 KHz, and 8 KHz), age, and work experience. Most of the studies investigating hearing loss have only reported the error rate and accuracy of the model, without mentioning anything about the weight and the effect size of different variables. One of the limitations of the current study was the problems the researchers encountered to convince the stakeholders in the industry and participants to cooperate with the research team.

5 Conclusion

This study focused on weighing factors that can cause hearing loss. According to the obtained results, in the first model, out of all the predictors (age, work experience, equal sound level, and frequency), the frequency of 8 KHz had the greatest impact (with a weight of 33%), while noise registered the smallest effect (with a weight of 5%). The accuracy of this model in predicting hearing loss was 100%. In the second model, the frequency of 4 KHz had the biggest impact (with a weight of 21%), whereas the frequency of 250 Hz had the smallest influence (with a weight of 6%). The accuracy of this model was 100% too. In the third model, the frequency of 4 KHz had the highest impact (with a weight of 22%), while the frequency of 250 Hz had the lowest effect (with a weight of 3%), with the accuracy of this model being 100% as well. In the fourth model, the frequency of 4 KHz had the greatest impact (with a weight of 24%), while the frequency of 500 Hz exerted the smallest effect (with a weight of 4%). The accuracy of this model was found to be 94%. Based on the accuracy indices obtained from these models, SVM algorithm can be regarded as an appropriate and powerful instrument to predict and model hearing loss.

References

1. Yu, P., Jiao, J., Chen, G., Zhou, W., Zhang, H. et al. (2018). Effect of GRM7 polymorphisms on the development of noise-induced hearing loss in Chinese Han workers: a nested case-control study. *BMC Medical Genetics, 19,* 4. DOI 10.1186/s12881-017-0515-3.
2. Thepaksorn, P., Koizumi, A., Harada, K., Siriwong, W., Neitzel, R. L. (2019). Occupational noise exposure and hearing defects among sawmill workers in the south of Thailand. *International Journal of Occupational Safety and Ergonomics, 25(3),* 458–466. DOI 10.1080/10803548.2017.1394710.

3. Zamanian, Z., Mohammadi, H., Rezaeeyani, M. T., Dehghany, M. (2012). An investigation of shift work disorders in security personnel of 3 hospitals of Shiraz University of Medical Sciences, 2009. *Iran Occupational Health, 9,* 52–57.

4. Zamanian, Z., Dehghani, M., Hashemi, H. (2013). Outline of changes in cortisol and melatonin circadian rhythms in the security guards of shiraz university of medical sciences. *International Journal of Preventive Medicine, 4,* 825.

5. Zamanian, Z., Nikravesh, A., Monazzam, M. R., Hassanzadeh, J., Fararouei, M. (2014). Short-term exposure with vibration and its effect on attention. *Journal of Environmental Health Science and Engineering, 12,* 135.

6. Zamanian, Z., Monazzam, M. R., Satyarvand, M., Dehghan, S. F. (2012). Presentation of a model to identify dominant noise source in agricultural sector of sugarcane industry. *Advances in Environmental Biology, 6,* 3002–3007.

7. Lin, J., Wang, H., Yan, F., Tang, K., Zhu, H. et al. (2018). Effects of occupational exposure to noise and dust on blood pressure in Chinese industrial workers. *Clinical and Experimental Hypertension, 40,* 257–261. DOI 10.1080/10641963.2017.1368534.

8. Nassiri, P., Zare, S., Monazzam, M. R., Pourbakht, A., Azam, K. et al. (2017). Evaluation of the effects of various sound pressure levels on the level of serum aldosterone concentration in rats. *Noise and Health, 19(89),* 200. DOI 10.4103/nah.NAH_64_16.

9. Halvani, G. H., Zare, M., Barkhordari, A. (2008). Noise induced hearing loss among textile workers of Taban factories in Yazd. *Journal of Birjand University of Medical Sciences, 15,* 69–74.

10. Ballenger, J. J. (1991). *Diseases of the nose, throat, ear, head, and neck.* Philadelphia: Lea & Febiger.

11. Tavakoli, M., Talebi, H., Shomeil Shushtari, S., Mazahery Tehrani, N., Faghihzadeh, S. (2014). Audiometric results and cervical vestibular evoked myogenic potentials in patients with type I and II diabetes mellitus. *Bimonthly Audiology-Tehran University of Medical Sciences, 23,* 40–48.

12. Nassiri, P., Zare, S., Monazzam, M. R., Pourbakht, A., Azam, K. et al. (2016). Modeling signal-to-noise ratio of otoacoustic emissions in workers exposed to different industrial noise levels. *Noise & Health, 18,* 391.

13. Safari Variani, A., Ahmadi, S., Zare, S., Ghorbanideh, M. (2018). Water pump noise control using designed acoustic curtains in a residential building of Qazvin city. *Iran Occupational Health, 15,* 126–135.

14. Aluclu, I., Dalgic, A., Toprak, Z. F. (2008). A fuzzy logic-based model for noise control at industrial workplaces. *Applied Ergonomics, 39(3),* 368–378. DOI 10.1016/j.apergo.2007.08.005.

15. Biswas, A. (1995). *Clinical audio-vestibulometry for otologists and neurologists.* Bombay, India: Bhalani Publishing House.

16. Aras, V. P. (2003). Audiometry techniques, circuits, and systems. M. Tech. Credit Semin. Report, Electron. Syst. Group, EE Dep. *Bombay, Citeseer.*

17. Nalamwar, M. K. (2004). Advanced audiometer: a novel signal generator technique. M. Tech. Credit Semin. Report, Electron. *Syst. Gr.*

18. Guerra-Jiménez, G., De Miguel, Á. R., González, J. C. F., Barreiro, S. A. B., Plasencia, D. P. et al. (2016). Cochlear implant evaluation: prognosis estimation by data mining system. *Journal of International Advanced Otology, 12 (1),* 1–7. DOI 10.5152/iao.2016.510.

19. Montgomery, A. (1998). Data mining-business hunching, not just data crunching. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 98,* 39–48.

20. won Lee, J., Giraud-Carrier, C. (2013). Results on mining NHANES data: a case study in evidence-based medicine. *Computers in Biology and Medicine, 43(5),* 493–503. DOI 10.1016/j.compbiomed.2013.02.018.

21. Dong, J. X., Suen, C. Y., Krzyżak, A. (2003). A fast SVM training algorithm. *International Journal of Pattern Recognition and Artificial Intelligence, 17,* 367–384. DOI 10.1142/S0218001403002423.

22. Majumder, J., Sharma, L. K. (2014). Application of data mining techniques to audiometric data among professionals in India. *Journal of Scientific Research and Reports, 3,* 2860–2971.

23. Ramos-Miguel, A., Perez-Zaballos, T., Perez, D., Falconb, J. C., Ramosb, A. (2015). Use of data mining to predict significant factors and benefits of bilateral cochlear implantation. *European Archives of Oto-Rhino-Laryngology, 272(11),* 3157–3162. DOI 10.1007/s00405-014-3337-3.

24. Nawi, N. M., Rehman, M. Z., Ghazali, M. I. (2011). Noise-induced hearing loss prediction in Malaysian industrial workers using gradient descent with adaptive momentum algorithm. *International Review on Computers and Software, 6,* 740–748.

25. ISO 9612. 2009. (2009). Acoustics-determination of occupational noise exposure-engineering method. *Geneva, Switz Int Organ Stand*.

26. Golmohammadi, R., Aliabadi, M. (1999). *Noise and vibration engineering*. Hamedan: Daneshju.

27. Gubbels, S. P., Gartrell, B. C., Ploch, J. L., Hanson, K. D. (2017). Can routine office-based audiometry predict cochlear implant evaluation results? *Laryngoscope, 127(1),* 216–222. DOI 10.1002/lary.26066.

28. Schlauch, R. S., Nelson, P. (2009). Puretone evaluation. *Handbook of Clinical Audiology, 6,* 30–49.

29. ISO 1999. 2013. (2013). Acoustics—estimation of noise-induced hearing loss.

30. WHO. 1991. (1991). Report of the informal working group on prevention of deafness and hearing impairment programme planning. *Geneva*, 18–21.

31. Dong, J., Krzyzak, A., Suen, C. Y. (2005). Fast SVM training algorithm with decomposition on very large data sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(4),* 603–618. DOI 10.1109/TPAMI.2005.77.

32. Boser, B. E., Guyon, I. M., Vapnik, V. N. (1992). *A training algorithm for optimal margin classifiers*. New York, NY, USA: ACM.

33. Keerthi, S. S., Shevade, S. K., Bhattacharyya, C., Murthy, K. R. K. (2001). Improvements to Platt's SMO algorithm for SVM classifier design. *Neural Computation, 13(3),* 637–649. DOI 10.1162/089976601300014493.

34. Keerthi, S. S., Gilbert, E. G. (2002). Convergence of a generalized SMO algorithm for SVM classifier design. *Machine Learning, 46(1/3),* 351–360. DOI 10.1023/A:1012431217818.

35. Marcoulides, G. A. (2005). Review of discovering knowledge in data: an introduction to data mining. *Journal of the American Statistical Associations, 100(472),* 1465.

36. Tajic, R., Ghadami, A., Ghamari, F. (2008). The effects of noise pollution and hearing of metal workers in Arak. *Zahedan Journal of Research in Medical Sciences, 10,* 293–301.

37. Dubno, J. R., Eckert, M. A., Lee, F. S., Matthews, L. J., Schmiedt, R. A. (2013). Classifying human audiometric phenotypes of age-related hearing loss from animal models. *Journal of the Association for Research in Otolaryngology, 14(5),* 687–701. DOI 10.1007/s10162-013-0396-x.

38. Acır, N., Özdamar, Ö., Güzeliş, C. (2006). Automatic classification of auditory brainstem responses using SVM-based feature selection algorithm for threshold detection. *Engineering Applications of Artificial Intelligence, 19(2),* 209–218. DOI 10.1016/j.engappai.2005.08.004.

39. Wang, S., Yang, M., Du, S., Yang, J., Liu, B. et al. (2016). Wavelet entropy and directed acyclic graph support vector machine for detection of patients with unilateral hearing loss in MRI scanning. *Frontiers in Computational Neuroscience, 10,* 106.

40. Taylor, K. R., DeLuca, A. P., Shearer, A. E., Hildebrand, M. S., Black-Ziegelbein, E. A. et al. (2013). Audio gene: predicting hearing loss genotypes from phenotypes to guide genetic screening. *Human Mutation, 34,* 539–545.

41. Noma, G., Khanapi, N., Ghani, M., Mohamad Khir, A., Noorizan, A. et al. (2013). Predicting hearing loss symptoms from audiometry data using FP-growth algorithm and Bayesian classifier. *Australian Journal of Basic and Applied Sciences, 7(9),* 35–43.