

Evaluation of the Characteristics and Machine Learning-Based Identification Accuracy of the International Electrotechnical Commission (IEC) 60601-1-8 Medical Equipment Auditory Alarms

Review began 04/07/2025

Review ended 04/17/2025

Published 04/18/2025

© Copyright 2025

Ishida et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.7759/cureus.82488

Kai Ishida¹, Kiyotaka Fujii²

1. Faculty of Engineering, Shonan Institute of Technology, Fujisawa, JPN 2. School of Allied Health Sciences, Kitasato University, Minato, JPN

Corresponding author: Kai Ishida, gishikai310@gmail.com

Abstract

Introduction

Auditory alarms in clinical settings signal sudden changes in a patient's condition and failures in medical equipment. However, distinguishing between simultaneously sounding alarms, particularly when superimposed with various ambient sounds, remains challenging. This study aimed to develop a machine learning (ML) model for identifying auditory alarms issued by medical equipment.

Methods

We targeted old and new auditory alarms for medical equipment as specified in the International Electrotechnical Commission (IEC) 60601-1-8 standard. First, we evaluated the characteristics of both normal and degraded auditory alarms using cosine similarity among old and new alarms. Next, we evaluated the accuracy of ML-based identification of deteriorated alarm sound sources in both the old and new alarm groups.

Results

The cosine similarity among old alarms was over 0.99, while new alarms ranged from 0.886 to 0.985, and exhibited more distinct characteristics. When noise was superimposed, the similarity among old alarms increased further, making differentiation more difficult. In contrast, for most new alarms, cosine similarity values exceeded 0.99 but retained slight acoustic differences even after noise-induced degradation, demonstrating improved distinguishability. The accuracy for identifying a single degraded alarm sound was 71.9% for the support vector machine. The models exhibited a high number of misclassifications when identifying the old alarms. Conversely, the models achieved higher accuracy when classifying new alarms, with recall exceeding 80%, precision above 70%, and F-measure greater than 80% for all new alarms. The identification accuracies for two simultaneous alarms were under 20% and approximately 50% for old and new alarms, respectively. The accuracy declined when estimating two simultaneous new alarms; however, when at least one of the two alarms was correctly classified, the accuracy exceeded 90%.

Conclusions

This study evaluated the characteristics of old and new auditory alarms issued by medical equipment as specified in IEC 60601-1-8 and constructed ML models for identifying the type of alarms.

Categories: Quality Improvement, Healthcare Technology

Keywords: auditory alarm, cosine similarity, machine learning, medical equipment, random forest, support vector machine

Introduction

Medical equipment and systems are equipped with auditory alarm systems to alert medical personnel about abrupt changes in patient status, operational abnormalities, or malfunctions. However, questionnaire survey data on the efficiency of medical equipment auditory alarms indicate instances of missed alarms and confusion with other auditory signals [1]. For instance, a 2017 survey revealed that approximately 40% of medical incidents and near-misses caused by problems related to patient monitor and equipment alarms occurred despite the triggering of alarm signals [2]. According to previous research, ambient noise sources in intensive care units include staff conversations, door opening and closing, footsteps, operating equipment, and medical equipment alarms [3]. In medical environments, particularly in intensive care and critical care units, auditory alarms sound frequently. Although lacking a standardized definition, alarm fatigue, wherein staff become desensitized owing to excessive alarm exposure, is considered a major concern [4]. Consequently, many hospitals are implementing alarm reduction strategies [5-8]. For instance, technical

How to cite this article

Ishida K, Fujii K (April 18, 2025) Evaluation of the Characteristics and Machine Learning-Based Identification Accuracy of the International Electrotechnical Commission (IEC) 60601-1-8 Medical Equipment Auditory Alarms. Cureus 17(4): e82488. DOI 10.7759/cureus.82488

alarms indicating mechanical issues, such as electrode or probe disconnections and signal loss, can be minimized through thorough daily equipment checks [8]. However, reducing the frequency of high-priority alarms indicating physiological abnormalities or sudden changes in patient vital signs remains challenging, emphasizing the importance of reliable alarm recognition and response measures.

The International Electrotechnical Commission (IEC) 60601-1-8 standard specifies general requirements for medical equipment alarm systems [9]. Revisions to this standard in 2020 introduced changes to the alarm categorization criteria outlined in the 2012 version. Additionally, the tonal characteristics assigned to each alarm category were significantly modified to enhance ergonomic effectiveness. This revision was prompted by findings indicating that the pulse alarm tone specified in the previous edition was difficult to memorize and prone to misinterpretation [10]. The auditory alarm tones defined in the revised standard have been reported to enable improved auditory discrimination and detectability compared to previous versions [11]. However, humans are continuously exposed to environmental sounds, which impose inherent limitations on the accurate recognition of simultaneous alarms.

In addition to auditory alarms, supplementary visual alert systems, such as flashing or illuminated indicators, are also used in medical settings. However, these systems present inherent limitations. An alternative approach involves networking medical devices within a facility and integrating them with the hospital information system (HIS). For instance, patient monitors, infusion pumps, and ventilators can be connected to the HIS through a wireless local area network, enabling continuous monitoring of their operational status. In this scenario, system malfunctions trigger appropriate alarms [12,13]. However, the implementation of such systems encounters several challenges: communication protocol differences between medical equipment and HIS manufacturers complicate system construction, many medical equipment systems lack external output terminals, system construction costs remain high, and collaboration between stakeholders remains limited. Alternatively, recent research has explored the application of sound recognition and machine learning (ML) using edge devices [14]. ML has the potential to process complex acoustic patterns, improve detection in noisy environments, and reduce misclassification. For instance, if an edge device can detect auditory alarms and extract alarm information, it may enable seamless HIS integration for equipment lacking external output terminals. However, no studies have yet examined the identification accuracy of both the old and new alarm tones specified in IEC 60601-1-8 using ML approaches. Moreover, actual clinical environments present additional challenges. For instance, environmental noise in these settings degrades alarm sounds, and multiple alarms are often issued simultaneously. The current evaluation of these factors remains insufficient, warranting further investigation.

This study aimed to develop an ML model for identifying auditory alarms issued by medical equipment. The following steps were undertaken. First, we evaluated the characteristics of old and new medical equipment alarms, as defined by IEC 60601-1-8. Next, we assessed the characteristics of auditory alarms when combined with various environmental sounds. Finally, based on the obtained findings, an ML model was constructed to identify degraded alarms and those occurring simultaneously and frequently.

Materials And Methods

Sound characteristics of old and new auditory alarms

This study evaluated the frequency characteristics and similarity of auditory alarms. The analysis focused on the sample sounds of old and new medical equipment alarms outlined in IEC 60601-1-8 [9]. Notably, the IEC 60601-1-8 standard mandates that specific tones must alert personnel of equipment malfunctions, process abnormalities, and changes in patient pathology, such as power supply failure, completion of infusion, and the occurrence of arrhythmias. For example, the cardiovascular icon defined by the old standard uses trumpet call, call to arms, and major chord. However, its new standard uses "lun-dup" heartbeat sound. The previous edition of the standard stipulated that the following eight categories of alarms must be structured as five-pulse tone sequences with distinct patterns: cardiovascular, drug delivery, general, oxygen, perfusion, power failure, temperature, and ventilation (hereafter referred to as CARD, DRUG, GEN, OXY, PERF, POWER, TEMP, and VENT, respectively). The 2020 revision introduced auditory icon sounds tailored to each of the following abnormalities: cardiovascular, drug delivery, failure, oxygen, perfusion, temperature, and ventilation (hereafter referred to as CARD, DRUG, FAIL, OXY, PERF, TEMP, and VENT, respectively). Furthermore, both editions specify different alarm tones based on the priority level (medium or high). This study exclusively focused on high-priority auditory alarms. Alarm sound samples were obtained from audio files published on the web [15,16]. A schematic of the evaluation process is presented in Figure 1. To analyze the time-domain waveforms of both old and new alarms, a fast Fourier transform (FFT) was applied to generate spectrograms using Python's Librosa library [17]. The two-dimensional coordinate space data of the generated spectrograms were subsequently transformed into one-dimensional data, from which the cosine similarity between alarms was calculated. Cosine similarity is a measure of similarity between two non-zero vectors defined in an inner product space and applied in various fields such as pattern recognition, machine learning, decision making, and image processing [18].

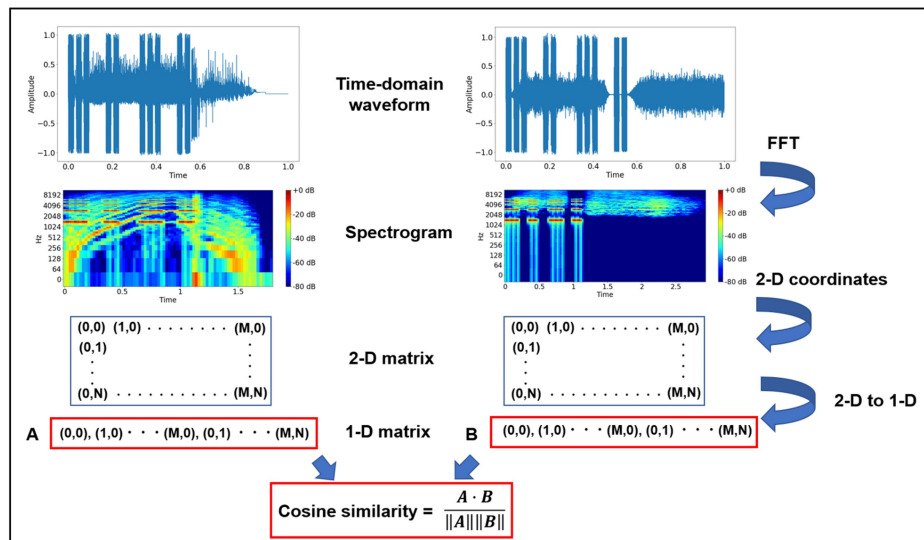


FIGURE 1: Schematic of the cosine similarity calculation for sample auditory alarms.

1-D: one dimension; 2-D, two dimensions; FFT: fast Fourier transform.

Evaluation of degraded auditory alarms

To simulate real-world conditions, we assumed that alarms were issued in actual environments and evaluated the characteristics of alarm sounds with various superimposed noise sources. A schematic of the evaluation process is depicted in Figure 2. This analysis focused on the eight old alarm types and seven new alarm types described in the previous section. A total of 10 noise sources were used, including helicopter and aircraft engine sounds, construction noise, conversation noise, as well as white noise and pink noise. The pydub library in Python was used to synthesize sound sources and generate training data [19]. The volume of each original alarm was incrementally reduced in 0.1 dB steps, ranging from 0.1 to 2.0 dB, before being combined with each noise source. Thus, each alarm was modified across 20 volume levels and paired with one of 10 noise types, generating 200 data samples. Z-score normalization was applied to the dataset. Next, Python's Librosa library was used to obtain key acoustic features, including the chromagram, zero-crossing rate, spectral centroid, spectral bandwidth, spectral roll-off frequency, and the first 20 mel-frequency cepstral coefficients (MFCCs) [17,20]. To evaluate the similarity between different alarm sounds under degraded conditions, we computed the average value of each index across the 200 data samples per alarm. Cosine similarity was then calculated based on these averaged feature values.

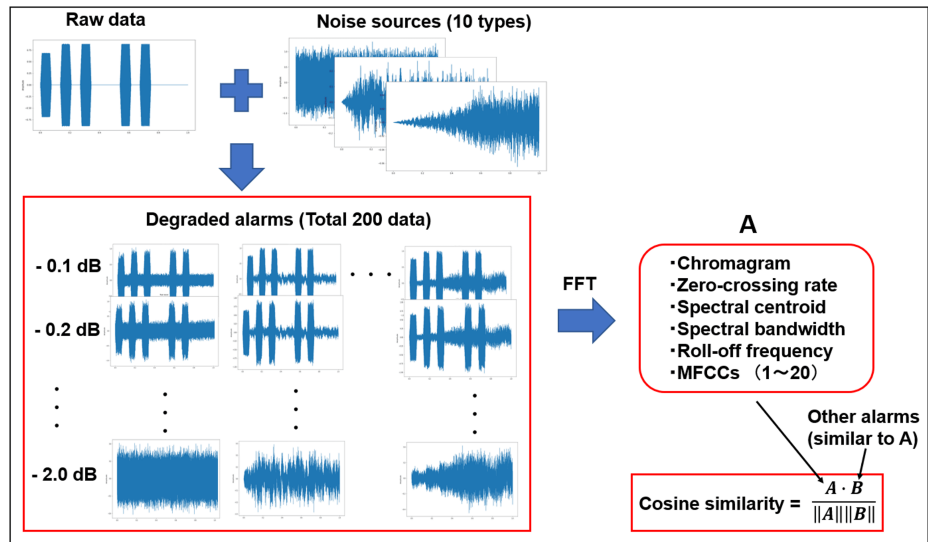


FIGURE 2: Schematic of the cosine similarity calculation for degraded alarms.

FFT: fast Fourier transform; dB: decibel; MFCCs: mel-frequency cepstral coefficients.

ML-based identification of auditory alarms

We evaluated the accuracy of ML-based identification of deteriorated alarm sound sources in both the old and new alarm groups. This study comprised two tasks: single-alarm identification, which involved identifying one type of degraded alarm, and multi-alarm identification, which involved identifying two overlapping alarm sound sources. The analysis targeted the eight types of old alarms and seven types of new alarms, as specified in IEC 60601-1-8. For the single-alarm identification task, we used the same dataset as that specified in the previous section, where each alarm was modified across 20 volume levels and combined with one of 10 noise types. ML models were then trained and validated using this dataset. In the multi-alarm identification task, two different alarms were first synthesized at the same volume. Figure 3 illustrates the dataset construction process. Next, a sound source containing these two alarms was combined with one noise type from the 10 noise categories. During this process, the volume of the alarm sound source was varied in three steps: 5 dB, 15 dB, and 25 dB. The final dataset was constructed for training and validation.

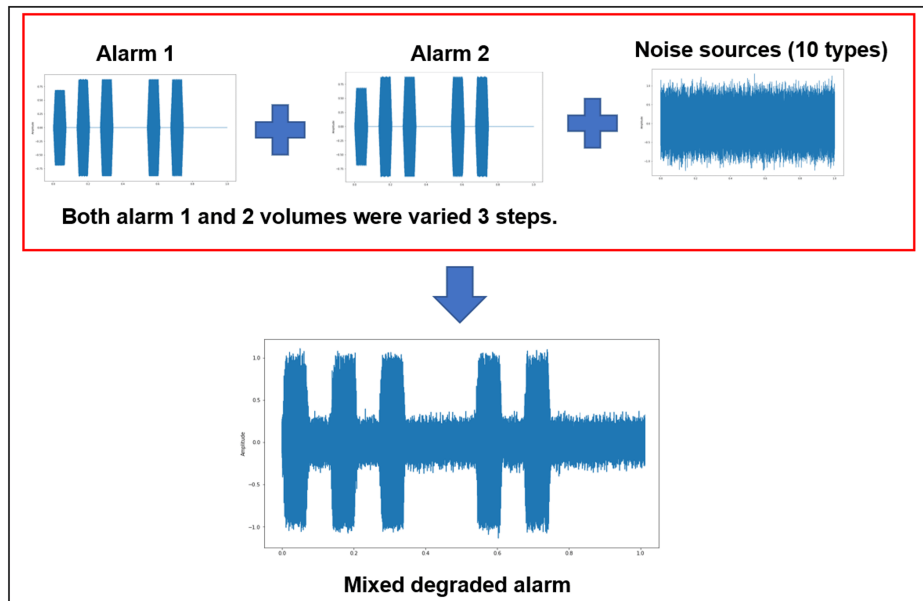


FIGURE 3: Schematic of the dataset construction process for identifying two simultaneously sounding alarms.

In this study, we extracted 31-dimensional features from the measured data using signal processing techniques to make the various characteristics of alarm sounds into ML [21]. Figure 4 illustrates the signal processing framework. First, simple moving averages with an interval of 500 points were applied to the time-domain waveforms to detect envelope waves and extract low-frequency features. The mean, standard deviation, minimum (Min), maximum (Max), and gradient were then computed from envelope waves extracted from the middle one-third of the total time. Additionally, the number of peaks in the envelope waves was counted within the time range spanning from one-eighth to three-fourths of the total duration. For high-frequency feature extraction, a Butterworth filter with a passband edge frequency of 10 Hz and a stopband edge frequency of 40 Hz was applied to the time-domain waveforms. To eliminate the influence of data on different scales and make model learning smoother, the filtered data underwent Z-score normalization before further processing. Next, an FFT was applied to compute the frequency spectra. Finally, spectral features, including the chromagram, zero-crossing rate, spectral centroid, spectral bandwidth, roll-off frequency, and the first 20 MFCCs, were extracted using the Librosa library.

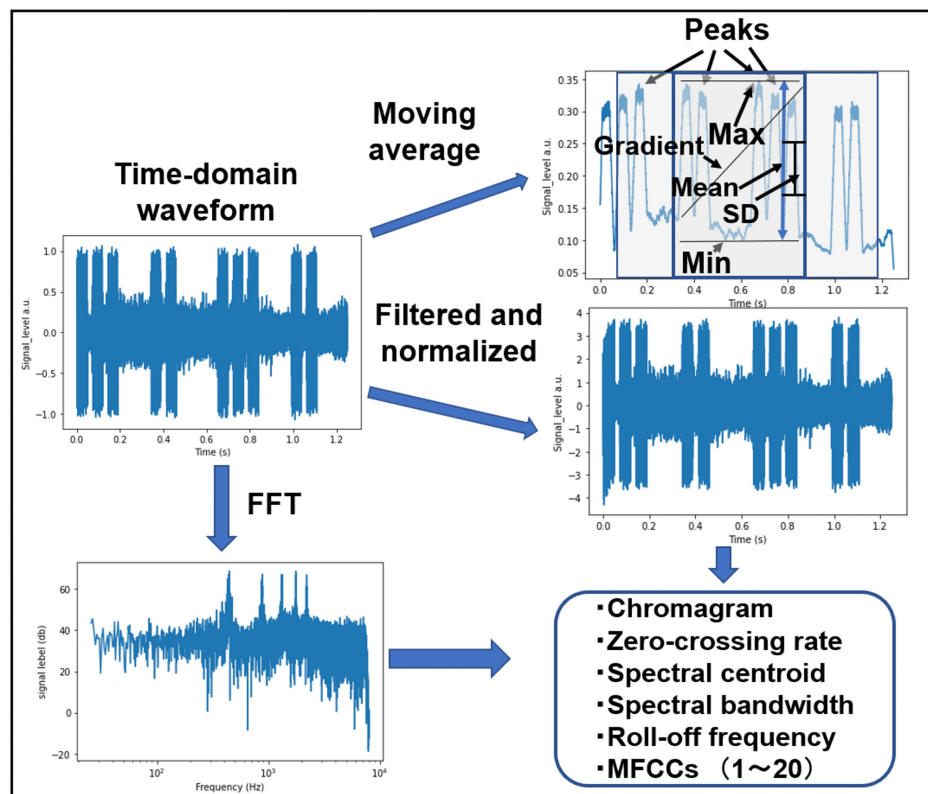


FIGURE 4: Schematic of the signal processing steps used to extract features.

FFT: fast Fourier transform; Max: maximum; Min: minimum; MFCCs: mel-frequency cepstral coefficients; SD: standard deviation.

Candidates for the regressors and classifiers included the random forest (RF) and support vector machine (SVM) algorithms. Hyperparameters were tuned for each of these classifiers, including the number of trees and maximum depth for RF, as well as the radial basis function kernel parameter (gamma) and regularization parameter (C) for SVM. Optimal hyperparameters were selected after tuning. To prevent overfitting, we employed a stratified group K-fold cross-validation method with K = 5. The model's performance was evaluated using overall accuracy. Additionally, recall, precision, and F-measure were calculated for each alarm. To assess these evaluation metrics, the data of each subject were divided into 90% training data for supervised learning and 10% test data for evaluation.

Results

Characteristics of old and new auditory alarms

Figures 5, 6 depict the spectrograms of the original high-priority alarm sounds for the eight types of old alarms and the seven types of new alarms examined in this study. Notably, the old alarms exhibit fundamental frequencies between approximately 500 and 1800 Hz, with accompanying overtone frequencies. The old high-priority alarm sound includes a beep within the 500-1000 Hz range, varying in pitch over a five-step scale. While each frequency component differs slightly, human auditory perception

may struggle to distinguish between them. New alarms also display a common signal around 1500 Hz with harmonic frequencies, representing a beep tone used across all high-priority alarms. However, beyond this shared signal, the spectrograms of the new alarms exhibit significant differences from those of the old alarms.

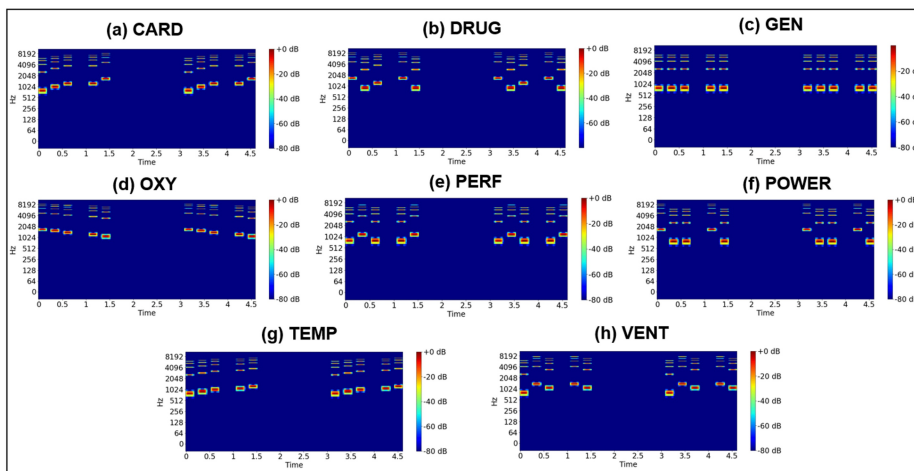


FIGURE 5: Spectrograms of auditory alarms based on the old standard.

CARD: cardiovascular; DRUG: drug delivery; GEN: general; OXY: oxygen; PERF: perfusion; POWER: power failure; TEMP: temperature; VENT: ventilation.

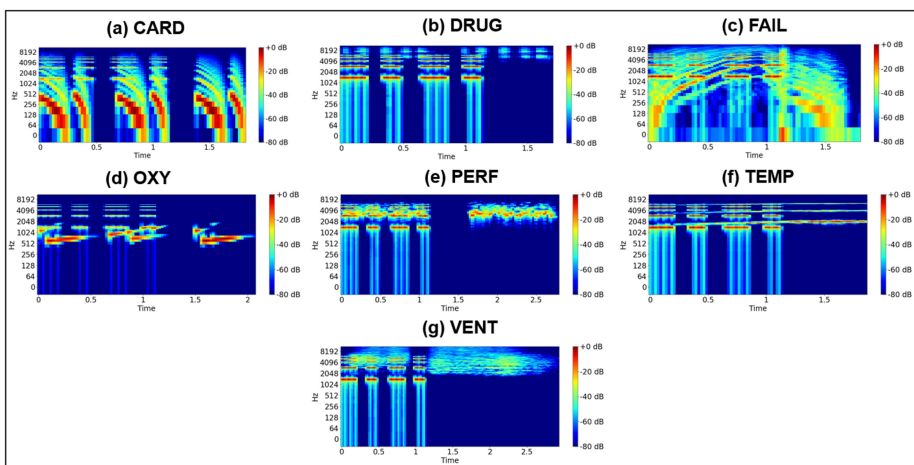


FIGURE 6: Spectrograms of auditory alarms based on the new standard.

CARD: cardiovascular; DRUG: drug delivery; FAIL: failure; OXY: oxygen; PERF: perfusion; TEMP: temperature; VENT: ventilation.

Figure 7 presents the cosine similarity values computed from the spectrogram components of the old and new auditory alarms. The similarity values for all combinations of old alarms were 0.99 or higher. In contrast, the similarity values among new alarms ranged from 0.886 to 0.985. These results suggest that the new alarms are more distinguishable than the old alarms.

		OLD								NEW							
		CARD	DRUG	GEN	OXY	PERF	POWER	TEMP	VENT	CARD	DRUG	FAIL	OXY	PERF	TEMP	VENT	
OLD	CARD	1.000	0.992	0.992	0.992	0.992	0.990	0.992	0.992	0.956	0.944	0.940	0.983	0.954	0.969	0.931	
	DRUG		1.000	0.991	0.993	0.991	0.994	0.992	0.990	0.956	0.943	0.939	0.982	0.953	0.968	0.929	
	GEN			1.000	0.991	0.997	0.995	0.993	0.993	0.957	0.944	0.940	0.984	0.955	0.971	0.931	
	OXY				1.000	0.991	0.992	0.991	0.992	0.956	0.943	0.939	0.982	0.953	0.968	0.930	
	PERF					1.000	0.992	0.994	0.993	0.957	0.944	0.940	0.984	0.954	0.970	0.931	
	POWER						1.000	0.991	0.991	0.956	0.943	0.939	0.982	0.953	0.969	0.929	
	TEMP							1.000	0.993	0.957	0.944	0.940	0.984	0.955	0.970	0.931	
	VENT								1.000	0.957	0.944	0.940	0.983	0.954	0.970	0.931	
NEW	CARD									1.000	0.937	0.937	0.972	0.955	0.967	0.908	
	DRUG										1.000	0.952	0.949	0.922	0.940	0.902	
	FAIL											1.000	0.945	0.920	0.941	0.886	
	OXY												1.000	0.973	0.985	0.934	
	PERF													1.000	0.973	0.923	
	TEMP														1.000	0.932	
	VENT															1.000	

FIGURE 7: Cosine similarity values derived from the spectrogram components of old and new alarms.

CARD: cardiovascular; DRUG: drug delivery; FAIL: failure; GEN: general; OXY: oxygen; PERF: perfusion; POWER: power failure; TEMP: temperature; VENT: ventilation.

Evaluation of degraded auditory alarms

Figure 8 presents the cosine similarity values for degraded alarm sounds. For old alarms, all similarity values were one. In contrast, for new alarms, except for FAIL and OXY, all cosine similarity values exceeded 0.99 but remained below one. This suggests that new alarms exhibit slight spectral differences, potentially improving their distinguishability.

		CARD								NEW							
		CARD	DRUG	GEN	OXY	PERF	POWER	TEMP	VENT	CARD	DRUG	FAIL	OXY	PERF	TEMP	VENT	
OLD	CARD	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.998	0.995	0.995	0.995	0.998	0.995	
	DRUG		1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.998	0.995	0.995	0.995	0.998	0.995	
	GEN			1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.997	0.995	0.995	0.995	0.998	0.994	
	OXY				1.000	1.000	1.000	1.000	1.000	0.998	0.998	0.996	0.996	0.995	0.998	0.995	
	PERF					1.000	1.000	1.000	1.000	0.998	0.998	0.995	0.995	0.995	0.998	0.995	
	POWER						1.000	1.000	1.000	0.998	0.998	0.995	0.995	0.995	0.998	0.995	
	TEMP							1.000	1.000	0.998	0.998	0.995	0.995	0.995	0.998	0.995	
	VENT								1.000	0.998	0.998	0.995	0.995	0.995	0.998	0.994	
NEW	CARD									1.000	0.998	0.998	0.998	0.997	0.999	0.996	
	DRUG										1.000	0.999	0.999	0.998	0.999	0.999	
	FAIL											1.000	1.000	0.998	0.999	0.999	
	OXY												1.000	0.998	0.999	0.999	
	PERF													1.000	0.999	0.998	
	TEMP														1.000	0.998	
	VENT															1.000	

FIGURE 8: Cosine similarity values derived from the spectrogram components of degraded old and new alarms.

CARD: cardiovascular; DRUG: drug delivery; FAIL: failure; GEN: general; OXY: oxygen; PERF: perfusion; POWER: power failure; TEMP: temperature; VENT: ventilation.

Estimation results of auditory alarms by ML

The estimation accuracies for identifying a single degraded alarm sound were 64.1% and 71.9% for RF and SVM, respectively. Figure 9 presents the confusion matrix obtained by SVM for both old and new degraded alarms. The models exhibited a high number of misclassifications when identifying the old alarms, resulting in recall values ranging from 40% to 66%, precision values between 34% and 70%, and F-measure values between 37% and 64%. In contrast, the models achieved higher accuracy when classifying new alarms, with recall exceeding 80%, precision above 70%, and F-measure greater than 80% for all new alarms. Notably, VENT was identified with 100% accuracy.

		Predicted										Recall	Precision	F-measure						
		OLD					NEW													
Actual	OLD	CARD	DRUG	GEN	OXY	PERF	POWER	TEMP	VENT	CARD	DRUG	FAIL	OXY	PERF	TEMP	VENT				
		CARD	80	0	10	35	20	30	20	0	5	0	0	0	0	0	0	40.0%	34.3%	37.0%
		DRUG	24	113	5	20	0	0	0	35	3	0	0	0	0	0	0	56.5%	62.8%	59.5%
		GEN	8	4	129	0	24	14	0	4	10	7	0	0	0	0	0	64.5%	60.3%	62.3%
		OXY	23	16	1	132	0	4	15	6	3	0	0	0	0	0	0	66.0%	62.6%	64.2%
	PERF	27	0	36	4	81	37	3	1	10	1	0	0	0	0	0	40.5%	44.5%	42.4%	
	POWER	13	2	10	0	41	99	27	0	8	0	0	0	0	0	0	49.5%	47.6%	48.5%	
	TEMP	32	0	3	20	12	24	102	0	7	0	0	0	0	0	0	51.0%	61.1%	55.6%	
	VENT	22	44	20	0	4	0	0	104	6	0	0	0	0	0	0	52.0%	69.3%	59.4%	
	NEW	CARD	0	1	0	0	0	0	0	0	196	3	0	0	0	0	98.0%	74.0%	84.3%	
DRUG		4	0	0	0	0	0	0	0	7	185	0	0	1	3	92.5%	93.4%	93.0%		
FAIL		0	0	0	0	0	0	0	0	0	187	13	0	0	0	93.5%	94.4%	94.0%		
OXY		0	0	0	0	0	0	0	0	0	33	167	0	0	0	83.5%	92.8%	87.9%		
PERF		0	0	0	0	0	0	0	0	0	0	0	196	0	4	98.0%	98.5%	98.2%		
TEMP	0	0	0	0	0	0	0	0	10	2	0	0	2	186	0	93.0%	98.4%	95.6%		
VENT	0	0	0	0	0	0	0	0	0	0	0	0	0	200	100.0%	98.0%	99.0%			

FIGURE 9: Confusion matrix for degraded old and new alarms classified using support vector machine.

CARD: cardiovascular; DRUG: drug delivery; FAIL: failure; GEN: general; OXY: oxygen; PERF: perfusion; POWER: power failure; TEMP: temperature; VENT: ventilation.

The identification accuracies for two simultaneous alarms were 15.3% and 18.6% for RF and SVM, respectively. For new alarms, RF achieved an identification accuracy of 51.2%, while SVM achieved an accuracy of 46.2%. Figure 10 presents the confusion matrix corresponding to the identification of two types of old alarms using SVM, while Figure 11 presents the confusion matrix corresponding to the identification of two types of new alarms using RF. The red numbers indicate cases where the identified alarm pair was entirely different from the correct combination. The recall rate for old alarms was low, ranging from a few percent to 30%. Although certain alarm combinations had a precision rate exceeding 70%, the F-measure remained below 40% across all cases. For new alarms, recall ranged from 30% to 70%, precision varied between 50% and 90%, and the F-measure ranged from 40% to 80%, indicating higher accuracy than that for old alarms. However, these results were still poor compared to single-alarm estimation. For instance, when only one alarm was identified, accuracy improved to 68.1% for old alarms and 94.4% for new alarms.

		Predicted																				Recall	Precision	F-measure									
		CARD					DRUG					GEN					OXY																
Actual	CARD	DRUG	GEN	OXY	PERF	POWER	TEMP	VENT	GEN	OXY	PERF	POWER	TEMP	VENT	OXY	PERF	POWER	TEMP	VENT	POWER	TEMP	VENT											
		DRUG	3	5	3	0	2	1	0	0	0	0	0	0	12	13	1	4	0	0	4	1	10	3	1	0	0	0	3	0	4.3%	13.6%	6.5%
		GEN	0	14	0	2	0	1	0	0	0	4	0	0	5	1	7	7	2	6	7	0	5	2	0	0	3	0	0	4	23.2%	10.7%	14.2%
		OXY	1	2	6	2	2	4	1	0	5	0	0	8	4	1	2	0	0	3	2	14	3	6	0	0	0	4	0	8.6%	27.3%	13.9%	
		PERF	0	1	0	17	0	7	3	0	0	0	0	8	0	4	8	0	0	5	5	6	2	0	0	1	2	0	0	24.3%	16.3%	19.9%	
	DRUG	POWER	1	4	1	3	7	0	0	0	6	0	0	3	4	3	5	0	3	0	19	2	0	2	0	0	1	1	0	5.3%	31.8%	8.5%	
		TEMP	0	1	2	5	0	13	3	0	0	0	0	7	1	0	5	0	0	5	4	3	2	3	0	0	4	0	0	8	21.0%	17.3%	19.0%
		VENT	0	0	0	4	0	5	14	0	0	0	0	4	0	0	3	0	0	5	1	4	6	9	0	0	1	0	0	13	24.6%	21.9%	23.0%
		GEN	1	6	0	0	0	0	0	0	14	0	2	2	5	5	0	6	8	8	2	1	6	2	0	0	1	0	0	20.0%	60.9%	30.1%	
		PERF	0	3	6	0	0	1	0	0	18	0	1	12	5	1	2	0	0	3	0	14	0	3	0	0	0	3	0	25.7%	40.9%	31.6%	
GEN	POWER	0	10	0	2	0	0	0	0	0	0	6	0	8	7	9	6	0	0	4	3	8	2	0	0	3	0	0	8.6%	23.1%	12.9%		
	TEMP	3	2	0	0	0	0	0	8	6	0	12	11	9	2	4	2	1	2	0	10	0	0	0	0	2	0	0	17.1%	70.6%	27.6%		
	VENT	0	6	0	0	0	5	1	0	0	0	12	11	1	6	0	0	0	0	13	7	0	0	0	1	0	2	0	17.1%	5.9%	8.7%		
	OXY	0	5	0	0	0	0	0	0	0	10	18	0	3	0	0	0	5	0	7	5	10	0	0	1	0	1	0	22.9%	12.9%	16.9%		
	PERF	0	8	0	0	0	0	0	0	2	0	7	2	16	4	0	0	5	4	8	2	0	0	0	0	1	2	0	22.9%	20.0%	21.3%		
OXY	POWER	0	8	0	1	0	1	0	0	3	0	5	1	2	9	6	16	4	0	5	2	0	0	4	0	0	0	0	3	22.9%	33.3%	27.1%	
	TEMP	0	4	0	5	0	1	2	0	0	0	0	0	5	0	6	3	0	2	24	0	5	3	0	0	3	0	2	5	34.3%	17.2%	22.9%	
	VENT	0	3	3	9	0	2	6	0	0	0	0	7	5	1	4	0	0	4	10	15	2	0	0	0	0	3	0	14.3%	9.1%	11.1%		
	GEN	7	3	3	0	4	0	0	10	0	12	2	5	1	0	0	3	0	17	0	0	0	0	0	0	0	3	0	24.3%	13.0%	16.9%		
	PERF	0	2	2	5	0	7	9	0	0	6	7	0	1	0	0	5	1	9	10	2	0	0	3	0	1	1	0	14.3%	35.8%	20.4%		
POWER	TEMP	0	1	1	3	0	4	4	0	0	0	0	5	7	0	1	0	0	5	0	4	3	99	0	0	0	0	0	41.4%	35.8%	38.4%		
	VENT	0	10	0	4	2	0	0	0	1	0	8	4	4	7	5	0	4	5	1	6	10	0	0	1	0	0	0	14.3%	52.6%	22.5%		
	GEN	0	1	0	9	0	0	2	0	0	0	5	0	4	9	0	0	7	4	0	2	0	0	0	4	14	1	0	3	20.0%	35.9%	25.7%	
	TEMP	0	0	0	13	0	0	7	0	0	0	0	0	0	0	3	0	0	8	4	0	4	4	0	3	6	0	3	8.6%	19.4%	11.9%		
	VENT	0	12	0	1	0	0	0	0	2	0	10	6	5	6	3	1	3	7	0	3	0	1	3	8	4	0	0	5.7%	16.0%	8.4%		
PERF	TEMP	1	4	1	5	1	0	2	0	0	0	7	9	4	1	0	0	7	4	0	3	1	0	0	0	13	5	0	7.1%	13.5%	9.3%		
	VENT	0	1	0	4	0	10	8	0	0	0	4	0	0	2	0	0	7	3	0	4	7	0	0	2	0	4	14	20.0%	21.9%	20.9%		

FIGURE 10: Confusion matrix for two simultaneously sounding degraded old alarms classified using support vector machine.

CARD: cardiovascular; DRUG: drug delivery; FAIL: failure; GEN: general; OXY: oxygen; PERF: perfusion; POWER: power failure; TEMP: temperature; VENT: ventilation.

		Predicted																					
		CARD						DRUG						FAIL									
		DRUG	FAIL	OXY	PERF	TEMP	VENT	FAIL	OXY	PERF	TEMP	VENT	OXY	PERF	TEMP	VENT	PERF	TEMP	VENT	Recall	Precision	F-measure	
Actual	CARD	DRUG	29	12	10	3	11	0	4	1	0	0	0	0	0	0	0	0	0	0	70.7%	52.7%	60.4%
		FAIL	8	36	5	4	17	0	0	0	0	0	0	0	0	0	0	0	0	0	51.4%	53.7%	52.6%
		OXY	8	3	34	5	9	0	3	0	0	0	0	8	0	0	0	0	0	0	48.6%	42.0%	45.0%
	DRUG	PERF	0	0	0	42	0	5	0	0	1	0	0	1	2	0	0	9	0	1	60.0%	38.5%	46.9%
		TEMP	2	10	3	5	47	0	0	0	0	0	0	0	0	3	0	0	0	0	67.1%	30.7%	42.2%
		VENT	0	0	0	2	0	50	0	0	0	0	0	2	0	0	0	5	0	0	71.4%	41.3%	52.4%
	FAIL	OXY	2	1	5	3	9	0	39	8	0	0	0	0	0	0	0	0	0	3	55.7%	44.3%	49.4%
		PERF	0	0	0	4	0	4	0	0	33	0	0	1	0	0	0	12	0	0	47.1%	84.6%	60.6%
		TEMP	0	1	2	3	15	0	9	2	0	27	0	0	0	5	0	6	0	0	38.6%	75.0%	50.9%
	OXY	VENT	0	0	0	1	0	12	0	0	0	0	23	0	0	0	13	0	0	13	32.9%	69.7%	44.7%
		PERF	3	1	10	4	10	0	10	3	0	0	0	27	0	0	0	0	2	0	38.6%	52.9%	44.6%
		TEMP	0	0	0	5	0	4	0	0	0	0	0	1	32	0	0	14	0	0	45.7%	82.1%	58.7%
	PERF	VENT	0	0	0	1	0	14	0	0	0	0	0	0	0	0	37	0	0	12	0	4	2
		TEMP	0	1	2	4	18	0	7	0	0	4	0	1	0	25	0	8	0	0	35.7%	67.6%	46.7%
		VENT	0	0	0	8	0	2	0	0	2	0	0	1	3	0	0	42	0	0	60.0%	45.2%	51.5%
OXY	PERF	1	1	3	2	8	0	4	0	0	5	0	3	0	4	0	5	31	3	0	0	0	
	TEMP	0	0	0	2	0	14	0	0	0	0	7	0	0	0	0	0	43	0	3	1	1	
	VENT	0	0	0	2	0	4	0	0	0	0	0	1	2	0	0	11	0	1	52.9%	97.4%	68.5%	
PERF	TEMP	0	0	0	5	0	7	0	0	3	0	1	0	0	0	0	3	0	0	72.9%	38.3%	50.2%	
	VENT	0	0	0	1	0	5	0	0	0	0	0	0	0	0	9	0	0	11	0	3	41	
	TEMP	0	0	0	1	0	5	0	0	0	0	0	0	0	0	9	0	0	11	0	3	41	

FIGURE 11: Confusion matrix for two simultaneously sounding degraded new alarms classified using random forest.

CARD: cardiovascular; DRUG: drug delivery; FAIL: failure; GEN: general; OXY: oxygen; PERF: perfusion; POWER: power failure; TEMP: temperature; VENT: ventilation.

Discussion

Many studies have evaluated the audibility of medical equipment alarms [11,22-24]. However, only a few have examined the audibility of alarms that are triggered simultaneously in the presence of environmental noise or attempted to identify them using ML. While some studies have explored signal processing algorithms for alarm identification, they have focused only on alarms based on the old standard, without considering techniques for distinguishing simultaneously sounding alarms [25]. Other studies have used deep learning approaches to predict simultaneously sounding alarms; however, these studies were limited to four alarm types and exclusively used old standard alarms [26]. In contrast, this study innovatively evaluates the characteristics of both old and new auditory alarms while also assessing their identification accuracy using ML.

The higher discrimination accuracy of the models for the new alarms compared to that for the old alarms is likely due to the greater tonal differences between alarms in the revised standard, as well as their distinct extracted features. This is supported by our evaluations using cosine similarity, revealing that the new alarms exhibit lower similarity between different alarm types. Additionally, even when alarm sounds are degraded by superimposed environmental sounds, the frequency components in the new alarms are designed to allow detection, unlike the pulse patterns of the old alarms. As illustrated in Figure 6, the cosine similarity of degraded alarm sounds exceeded 0.99 for new alarms owing to the influence of noise frequency components. This suggests that human auditory perception may struggle to distinguish between such degraded alarm sounds. However, ML techniques enable the detection of subtle alarm characteristics, improving identification accuracy.

Distinguishing two simultaneously sounding alarms was difficult even under the new standard. This difficulty likely arises from the following issue: as the number of simultaneously sounding alarms increases, the combined signal approaches white noise according to the central limit theorem, reducing distinctive features. Previous research has demonstrated that the identification accuracy for simultaneous alarms decreases with increasing alarm counts [26]. However, consistent with current findings, the probability of complete misidentification remains low.

In addition to the study's findings, several limitations and future challenges also warrant discussion. This study focused on sample sounds of auditory alarms as specified in IEC 60601-1-8 rather than actual alarm sounds from medical equipment. Additionally, the evaluation used mechanically synthesized degraded sounds rather than actual speaker outputs. The study assessed the alarm discrimination accuracy of ML approaches but did not evaluate healthcare personnel's auditory discrimination. Human auditory perception is significantly influenced by the type and volume of surrounding environmental sounds, necessitating a separate study to accurately evaluate these effects. Moreover, the same medical equipment is used multiple times in clinical settings, such as the intensive care unit. At the moment, we cannot distinguish which medical equipment is dispatching an alarm. In the future, it may be possible to identify sounds by using directional sound pickup devices and utilizing the intensity information. Another limitation is that medical equipment compliant with the new standard still includes alarms based on the old standard, meaning that both alarm types may coexist in medical settings. This study did not account for such scenarios, as it separately evaluated the distinguishability of old and new alarms. A more realistic assessment would involve evaluating a mixed model incorporating both alarm types. However, combining both old and new alarm types (15 alarms) would result in 105 possible two-alarm combinations, making it unlikely that current ML models would achieve high accuracy across all cases. In this study, we used a classical learning approach with relatively low computational cost. Future research could explore deep learning methods to improve identification accuracy.

The application of ML for identifying alarms based on the new IEC 60601-1-8 standard demonstrates that alarm sounds can be automatically detected with high accuracy. As a potential future application, detected alarm information could be integrated into the HIS and transmitted via wireless communication to mobile devices used by medical personnel. In addition to detecting alarm sounds emitted by medical equipment, this approach would enable automatic recognition of alarm signals, allowing information related to clinical condition changes or equipment malfunctions to be directly relayed to medical staff. This capability could facilitate communication limited to specific alarm types, reducing unnecessary notifications. This study focused exclusively on high-priority alarms. However, if the system could reliably distinguish high-priority alarms from moderate-level alarms, it would allow for the selective transmission of only critical alerts to medical personnel. Such improvements could help mitigate alarm fatigue. In future work, we will verify the results in an actual environment and evaluate their usefulness.

Conclusions

This study evaluated the characteristics of old and new auditory alarms issued by medical equipment as specified in IEC 60601-1-8. The cosine similarity between old alarms was extremely high, while new alarms exhibited greater variation. When noise was superimposed, the similarity among old alarms further increased, making differentiation more challenging. In contrast, even after degradation due to noise, slight acoustic differences remained detectable in the new alarms, demonstrating their improved distinguishability. Using ML, a single-alarm under the new standard could be identified with high accuracy, even in noisy conditions. When identifying two simultaneous alarms, the overall accuracy was low; however, when at least one of the two alarms was correctly classified, the accuracy exceeded 90%.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Kai Ishida

Acquisition, analysis, or interpretation of data: Kai Ishida, Kiyotaka Fujii

Drafting of the manuscript: Kai Ishida

Critical review of the manuscript for important intellectual content: Kai Ishida, Kiyotaka Fujii

Supervision: Kai Ishida

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue.

Animal subjects: All authors have confirmed that this study did not involve animal subjects or tissue.

Conflicts of interest: In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Watanabe S: The purpose of Health and Labour Sciences Research Grants Study in the present status of alarms of medical equipment and alarm-related problems. (Article in Japanese). *Iryou kikigaku*. 2002, 72:436-8. [10.4286/ikakikaigaku.72.9_436](https://doi.org/10.4286/ikakikaigaku.72.9_436)
2. Ishikawa M, Saito N: Strategies to prevent recurrence of incidents and accidents related to medical device alarm systems. (Article in Japanese). *Iryou kikigaku*. 2017, 87:285-91. [10.4286/jjmi.87.285](https://doi.org/10.4286/jjmi.87.285)
3. Pal J, Taywade M, Pal R, Sethi D: Noise pollution in intensive care unit: a hidden enemy affecting the physical and mental health of patients and caregivers. *Noise Health*. 2022, 24:130-6. [10.4103/nah.nah_79_21](https://doi.org/10.4103/nah.nah_79_21)
4. Graham KC, Cvach M: Monitor alarm fatigue: standardizing use of physiological monitors and decreasing nuisance alarms. *Am J Crit Care*. 2010, 19:28-34. [10.4037/ajcc2010651](https://doi.org/10.4037/ajcc2010651)
5. Cvach MM, Frank RJ, Doyle P, Stevens ZK: Use of pagers with an alarm escalation system to reduce cardiac monitor alarm signals. *J Nurs Care Qual*. 2014, 29:9-18. [10.1097/NCQ.0b013e3182a61887](https://doi.org/10.1097/NCQ.0b013e3182a61887)
6. Allan SH, Doyle PA, Sapirstein A, Cvach M: Data-driven implementation of alarm reduction interventions in a cardiovascular surgical ICU. *Jt Comm J Qual Patient Saf*. 2017, 43:62-70. [10.1016/j.jcjq.2016.11.004](https://doi.org/10.1016/j.jcjq.2016.11.004)
7. Ruppel H, De Vaux L, Cooper D, Kunz S, Duller B, Funk M: Testing physiologic monitor alarm customization software to reduce alarm rates and improve nurses' experience of alarms in a medical intensive care unit. *PLoS One*. 2018, 13:e0205901. [10.1371/journal.pone.0205901](https://doi.org/10.1371/journal.pone.0205901)
8. Haga H, Osaki C: Improving medical safety by reducing false alarms associated with electrocardiogram monitors. (Article in Japanese). *J Showa Univ Soc*. 2019, 79:676-82. [10.14930/jshowaunivsoc.79.676](https://doi.org/10.14930/jshowaunivsoc.79.676)

9. International Electrotechnical Commission. IEC 60601-1-8. IEC 60601-1-8:2006/AMD2:2020 Amendment 2 - Medical electrical equipment - Part 1-8: General requirements for basic safety and essential performance - Collateral standard: General requirements, tests and guidance for alarm systems in medical electrical equipment and medical electrical systems. (2020). <https://webstore.iec.ch/en/publication/59648>.
10. Wee AN, Sanderson PM: Are melodic medical equipment alarms easily learned?. *Anesth Analg*. 2008, 106:501-8. [10.1213/01.ane.0000286148.58823.6c](https://doi.org/10.1213/01.ane.0000286148.58823.6c)
11. Edworthy J, Reid S, McDougall S, et al.: The recognizability and localizability of auditory alarms: setting global medical device standards. *Hum Factors*. 2017, 59:1108-27. [10.1177/0018720817712004](https://doi.org/10.1177/0018720817712004)
12. Sanuki M: Inclusion of infusion pump in automatic control system. (Article in Japanese) . *ICU CCU*. 2020, 44:117-22.
13. Yoshioka J, Ishikawa S, Saitoh D, et al.: Development and testing of a ventilator remote adapter via communication between ventilator and receiving smart device utilizing accompanying app. (Article in Japanese). *Iryou kigaku*. 2018, 88:449-57. [10.4286/jjmi.88.449](https://doi.org/10.4286/jjmi.88.449)
14. Cerutti G, Andri R, Cavigelli I, Farella E, Magno M, Benini L: Sound event detection with binary neural networks on tightly power-constrained IoT devices. *Proceedings of the ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED '20)*. ACM, New York, NY; 2020. 19-24. [10.1145/3370748.3406588](https://doi.org/10.1145/3370748.3406588)
15. ISO/IEC 60601-1-8, Patterson and other alarms in medical equipment. Sample alarm sounds - sirens, buzzers and other sounds. (2000). <https://th.id.au/alarms/>.
16. International Standards Organization. IEC 60601-1-8 auditory alarm signal files . (2019). <https://www.iso.org/committee/52012.html?t=NACZ4V-7LtVKTQYH2VA6GHkkLwj4tuD3Zu0ITKcgW5sTPKPMwpeqzYZV0ZNMmRkg&view=docu....>
17. McFee B, Raffel C, Liang D, Ellis DPW, McVicar M, Battenberg E, Nieto O: Librosa: audio and music signal analysis in Python. *Proceedings of the 14th Python in Science Conference*. Curvenote, Calgary, Canada; 2015. 18-25. [10.25080/Majora-7b98e5ed-003](https://doi.org/10.25080/Majora-7b98e5ed-003)
18. Ye J: Cosine similarity measures for intuitionistic fuzzy sets and their applications . *Math Comput Model*. 2011, 53:91-7. [10.1016/j.mcm.2010.07.022](https://doi.org/10.1016/j.mcm.2010.07.022)
19. Pydub. (2024). Accessed: April 7, 2025: <https://github.com/jiaaro/pydub>.
20. Sahidullah M, Saha G: Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition. *Speech Commun*. 2012, 54:543-65. [10.1016/j.specom.2011.11.004](https://doi.org/10.1016/j.specom.2011.11.004)
21. Kai I: Novel estimation technique for the carrier-to-noise ratio of wireless medical telemetry using software-defined radio with machine-learning. *Sci Rep*. 2023, 13:4162. [10.1038/s41598-023-31225-3](https://doi.org/10.1038/s41598-023-31225-3)
22. Momtahan K, Héту R, Tansley B: Audibility and identification of auditory alarms in the operating room and intensive care unit. *Ergonomics*. 1993, 36:1159-76. [10.1080/00140139308967986](https://doi.org/10.1080/00140139308967986)
23. Takeuchi A, Hirose M, Shinbo T, Imai M, Mamorita N, Ikeda N: Development of an alarm sound database and simulator. *J Clin Monit Comput*. 2006, 20:317-27. [10.1007/s10877-006-9037-9](https://doi.org/10.1007/s10877-006-9037-9)
24. Atyeo J, Sanderson PM: Comparison of the identification and ease of use of two alarm sound sets by critical and acute care nurses with little or no music training: a laboratory study. *Anaesthesia*. 2015, 70:818-27. [10.1111/anae.13020](https://doi.org/10.1111/anae.13020)
25. Potnuru P, Epstein RH, McNeer R, Bennett C: Development and validation of an algorithm for the identification of audible medical alarms. *Cureus*. 2020, 12:e11549. [10.7759/cureus.11549](https://doi.org/10.7759/cureus.11549)
26. Kishimoto M, Takemura T, Sugiyama O, et al.: Prediction of polyphonic alarm sound by deep neural networks. (Article in Japanese). *Trans Jpn Soc Med Bio Eng*. 2022, 60:8-15. [10.11239/jsmbe.60.8](https://doi.org/10.11239/jsmbe.60.8)