Inverse-Mel scale spectrograms for high-frequency feature extraction and audio anomaly detection in industrial machines

Kader Basha Tajuddin Shaikh[1], Naresh P. Jawarkar[2], Vasif Ahmed[3], Nadir Nizar Ali Charniya[4]

[1]Department of Automation and Robotics Engineering, Vivekanand Education Society's Institute of Technology, Mumbai, India
[2]Department of Electrical and Power Engineering, Government College of Engineering, Amravati, India
[3]Department of Artificial Intelligence and Data Science, Babasaheb Naik College of Engineering, Pusad, India
[4]Department of Electronics and Telecommunication Engineering, Vivekanand Education Society's Institute of Technology, Mumbai, India

Article Info

ABSTRACT

Unlike humans, the energies in industrial machine sounds (IMS) vary across a wide range of frequencies. Mel scales, which are developed for the perception of human audio, fail to capture the complete information present in IMS. To improve performance, we propose using an inverse-Mel scale, along with the concatenation and combination of Mel and inverse-Mel scale based spectrograms, as feature vectors for audio anomaly detection (AAD) in industrial machines. Adaptation in the Librosa Python package and the DCASE 2022 Challenge Task 2 baseline system is pursued for the construction of inverse-Mel scale spectrograms. Experiments are conducted using the malfunctioning industrial machine investigation and inspection for domain generalization (MIMII DG) datasets. Systems based on the inverse-Mel scale achieve a maximum improvement of up to 37% in the bearing machine and an average improvement of up to 9% in the area under the curve (AUC) score across all machines in the MIMII DG datasets. The proposed features also enhance DG, overcoming the effects of environmental and operational domain shifts caused by variations in recording setup, load, background noise, and operational patterns. Challenge official evaluator assessed the proposed system against the evaluation datasets, ranking it three positions higher than the baseline system.

Corresponding Author:

Kader Basha Tajuddin Shaikh
Department of Automation and Robotics Engineering, Vivekanand Education Society's Institute of Technology
Mumbai 400074, India
Email: kader.shaikh@ves.ac.in

1. INTRODUCTION

Industrial machine sounds (IMS) convey considerable information about the status of a machine [1]–[3]. Through astute listening and careful observation, an operator can quickly assess the health of the machine. An experienced operator can easily identify faults that may arise in an otherwise healthy working machine. The operator's expertise enables the anticipation and prevention of potential crises. Audio anomaly detection (AAD) systems for industrial machines mimic the behavior of operators to identify machine health conditions and operational anomalies. AAD for fault diagnosis and prognosis in industrial machines is being widely researched and has been one of several tasks in all editions of the DCASE challenges since 2020 [4]–[7].

Several researchers focused on the analysis of high-frequency regions in IMS. Liu *et al*. [8] explored fault analysis in belt conveyor idlers. The effective distinguishing frequency bands for various fault conditions due to damaged cages, raceway slots, and large pits in the inner/outer races on the rolling element, are found to be concentrated in the medium to high-frequency (6–20 kHz) ranges. Guochao *et al*. [3] examined the audible sounds produced by milling machines and found that the sound signals spanned the full audible range. The authors identified low-frequency sound signals generated by tool holder vibrations, mid-range frequency sounds from metal deformation processes, and high-frequency sounds from friction mechanisms. Liu *et al*. [9] proposed a lightweight fault diagnosis network called MPNet for identifying bearing faults in rotating machinery. Authors outlined the limitations of Mel-frequency cepstral coefficients (MFCC) being sensitive only to low-frequency information and instead used linear spectrograms constructed using short-time Fourier transform as features. Liu *et al*. [10] observed high-frequency components in the audio signals of belt conveyors, specifically in the range of 1 to 5 kHz. The impacts and vibrations from defective rollers contribute to the generation of these high-frequency audio signals. Zhou *et al*. [11] noted acoustic signals generated by bulge conditions in tire endurance tests conducted on a drum testing machine to generate high energy peaks in the high-frequency regions. Zhao *et al*. [12] noted that features extracted from high-frequency regions of vibration signals are more effective in characterizing faults in power end bearings. Ma *et al*. [13] proposed the fusion of MFCCs, inverted Mel-scale frequency cepstrum coefficients (IMFCCs), Gammatone frequency cepstral coefficients (GFCCs), and linear prediction cepstral coefficients (LPCCs) to create a hybrid cepstral feature known as Mel-inverted-Gammatone-linear cepstral coefficients (MIGLCCs). This feature encapsulated the individual advantages of each constituent feature. Their findings indicated that the fusion of MFCCs and IMFCCs yielded the best results among all dual feature combinations tested. All the above research emphasized the importance of focusing on the energy present in higher frequency regions and highlights the benefits achieved through the use of inverse-Mel scale frequency warping technique. However, the application of inverse-Mel scale based spectrograms for AAD in industrial machines was not considered.

Based on the original investigation, this research pursued the construction of an inverse-Mel scale, a combination of Mel and inverse-Mel scale spectrograms, as front-end features for extracting energy distribution across the complete range of frequencies in IMS. The spectrograms are constructed by adapting the Librosa Python library. These constructed spectrograms serve as input for an autoencoder-based AAD system designed to identify anomalous operations in industrial machines. Experiments conducted on the malfunctioning industrial machine investigation and inspection for domain generalization (MIMII DG) dataset [4] demonstrate that AAD systems with inverse-Mel scale spectrograms perform better.

This work is motivated by the DCASE Challenge 2022 Task 2 [4]–[7], which focuses on AAD and domain generalization (DG) techniques in industrial machines. A total of 31 teams submitted 81 entries to the challenge. Most participants used Mel scale based acoustic features such as Mel energies, log-Mel energies, MFCC, Mel spectrograms, and log-Mel spectrograms in their systems [7]. Use of inverse-Mel scale based acoustic features for AAD and DG on the challenge datasets is proposed in this research. This research is the first of its kind to propose the use of the inverse-Mel scale for DCASE Challenge 2022 Task 2. Comparison with the published challenge scores [7] derives a relative position of 21st rank for the results presented in this research. This ranking is three positions higher than the official ranking of the baseline system.

Rest of the paper is organized as follows: section 2 describes the materials and methods employed in this experimentation. It includes the methods for construction of spectrograms, details the MIMII DG dataset, and the experimental setup along with the evaluation metrics for the DCASE Challenge 2022 Task 2. Section 3 presents and discusses the results, including performance scores and improvements observed on both the development and evaluation datasets. Section 4 summarizes the conclusions drawn from this research.

2. MATERIALS AND METHODS

2.1. Sound database of industrial machines

MIMII DG [4] a public database shared as a development and evaluation dataset for Task 2 of the DCASE Challenge 2022 [7] is used in this work. This dataset includes normal and anomalous operating sounds from five different industrial machines. It is designed for the development and evaluation of AAD and DG techniques in industrial machines. The dataset is divided into source and target domain data. The source domain data contains only the normal and anomalous operating sounds of the machine under test, whereas operational and environmental domain shifts commonly encountered in industrial setups are synthetically infused into these

sounds to generate the target domain data. The source domain data is employed for evaluating AAD, while the target domain data is used for evaluating DG.

2.2. Construction of inverse-Mel scale spectrograms
2.2.1. Equations of inverse-Mel scale

The two commonly used implementations for transformation between linear and Mel scale frequencies are hidden Markov toolkit 3 (HTK) [14] and Slaney [15]. Slaney implementations apply a linear formula for frequencies up to 1 kHz and a logarithmic or anti-logarithmic formula for conversions above 1 kHz. HTK implementations follow a logarithmic or anti-logarithmic formula for the entire range of frequencies. HTK implementations are used in this work. The relationship between linear frequency scale (f_{Hz}) and Mel-frequency scale (f_{mel}) is noted in (1) and (2),

$$f_{mel} = 2595 * log_{10}(1 + \frac{f_{Hz}}{700}) \tag{1}$$

$$f_{Hz} = 700 * (10^{(f_{mel}/2595)} - 1) \tag{2}$$

Several researchers in [13], [16]–[22] defined the inverse-Mel scale as the complement of the Mel scale. The authors suggested flipping the original Mel filterbank around its midpoint to derive the inverse-Mel filterbank. Mathematical relationships between the linear frequency scale (f_{Hz}) and the inverse-Mel frequency scale (f_{iMel}) are proposed by Chakroborty [23], [24], Sharma [25], Latha [16], Lalitha [18], and Ma [13]. Latha [16] and Lalitha [18] introduced in (3), Ma [13] proposed in (4), Chakroborty [23], [24] and Sharma [25] presented in (5).

$$f_{iMel} = 2146.1 - 2595 * log_{10}(1 + \frac{4000 - f_{Hz}}{700}) \tag{3}$$

$$f_{iMel} = 2146.1 - 1127 * log_{10}(1 + \frac{4000 - f_{Hz}}{700}) \tag{4}$$

$$f_{iMel} = 2195.286 - 2595 * log_{10}(1 + \frac{4031.25 - f_{Hz}}{700}) \tag{5}$$

Equation (5) is employed in this research. In the works of Chakroborty [23], [24] and Sharma [25], the sampling frequency is 8 kHz, whereas the sampling frequency in the MIMII DG [4] database is 16 kHz. Hence, the constant terms are changed from 2195.286 to 2844.06 and 4031.25 to 8031.25. The modified equations used in this work to convert between the linear frequency scale (f_{Hz}) and the inverse-Mel frequency scale (f_{iMel}) are presented in (6) and (7).

$$f_{iMel} = 2844.06 - 2595 * log_{10}(1 + \frac{8031.25 - f_{Hz}}{700}) \tag{6}$$

$$f_{Hz} = 8031.25 - 700 * (10^{(2844.06 - f_{iMel}/2595)} - 1) \tag{7}$$

Figure 1 shows plot of center frequencies for all filters in Mel scale and inverse-Mel scales. Center frequencies represent the midpoint of frequency bands used in Mel and inverse-Mel transformations. The Mel scale follows a logarithmic scale, whereas inverse-Mel scale functions on an anti-logarithmic scale.
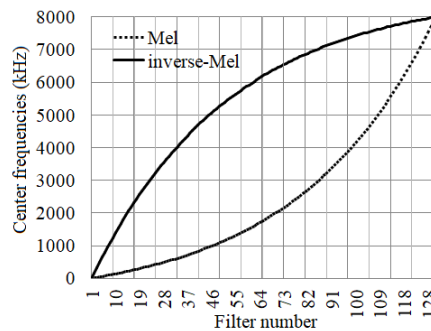

Figure 1. Center frequencies in Mel and inverse-Mel scales

2.2.2. Variants of spectrograms

The following types of spectrograms are constructed in this work.

– Mel scale spectrogram constructed using the standard equations of the Mel scale. Functions for constructing Mel spectrograms, as defined in the Librosa Python package, are employed.

– Inverse-Mel scale spectrogram constructed using the inverse-Mel scale equations described in section 2.2.1. The adaptation made to the Librosa Python package for constructing inverse-Mel scale spectrograms is described in section 2.2.3.

– Concatenated spectrogram constructed by vertically stacking Mel and inverse-Mel spectrograms. The Mel spectrogram captures lower frequencies, ranging from 0 to 4 kHz, while the inverse-Mel spectrogram captures higher frequencies from 4 to 8 kHz.

– Combinational spectrograms constructed by aggregating Mel and inverse-Mel spectrograms across the entire frequency range. The value at a specific frequency is determined by applying maximum, minimum, or average pooling to the Mel and inverse-Mel values. Consequently, this work develops three types of combinational spectrograms: maximum, minimum, and average value spectrograms.

2.2.3. Adaptations in Librosa package and DCASE 2022 baseline system for construction of inverse-Mel scale spectrograms

Adaptations have been made in several source files of the Librosa package [26] for the construction and presentation of inverse-Mel spectrograms. Two additional parameters, "isInverseMel" and "isHTK," are included as arguments in the melspectrogram, mel, and mel_frequencies functions in the 'filters.py' and 'spectral.py' files of the Librosa package. The "isInverseMel" parameter allows for toggling between Mel and inverse-Mel scale formulas, while the "isHTK" parameter enables the selection of either Slaney or HTK implementations. The concatenation and combination of Mel and inverse-Mel spectrograms are performed in the 'common.py' file of the DCASE 2022 baseline system. The adapted source files are available for download under the GNU General Public License at https://github.com/KaderShaikhVESIT/inverse-Mel.

2.3. Experimental set-up and evaluation metrics

Focus of this work is to introduce the inverse-Mel scale and discuss its implications. Hence, this work utilized the baseline system of DCASE 2022 Challenge Task 2 [4], [6], [7] as a detector. The baseline detector is a deep autoencoder. Each 10 seconds of audio is converted into a spectrogram that acts as an input feature vector for the autoencoder. The development and evaluation datasets of DCASE Challenge 2022 Task 2 are used for training and testing the detector. Confusion matrix, precision, recall, F1 score, and area under the curve (AUC) are calculated for both source and target domain data, whereas partial area under the curve (pAUC) is calculated for combined source and target data. Equations for calculation of AUC and pAUC scores are defined in [4], [6]. System evaluation and ranking is done using the official evaluator shared by the organizers [27].

3.   RESULTS AND DISCUSSION

3.1.   Inferences on all spectrograms

All spectrograms of a typical machine sound recording from the Slide rail machine (section_00_source_train_normal_0010_vel_1100.wav) in the MIMII DG dataset are shown in Figure 2. The spectrograms utilize a blue-white-red (BWR) colormap. Where bright red indicates higher amplitude or activity and blue indicates lower amplitude.

Figure 2(a) presents the spectrogram using a linear frequency scale based on short-time Fourier transform (STFT), which exhibits bright red spots in both low and high-frequency regions, suggesting that sound energy is distributed across the entire frequency range. In Figure 2(b), the Mel spectrogram emphasizes lower frequency regions while suppressing the higher frequency regions. Frequencies above 2 kHz are primarily depicted in white-blue color, indicating a repression of high-frequency components. This limitation suggests that the Mel scale spectrogram fails to capture and present the complete information inherent in IMS.

In contrast, the inverse-Mel scale spectrogram shown in Figure 2(c) enhances the high-frequency regions, effectively revealing the energy content that is otherwise suppressed in the Mel scale spectrogram. Energy components above 6 kHz, which are often obscure in Mel spectrograms, are vividly displayed here. The concatenated spectrogram shown in Figure 2(d) merges Mel and inverse-Mel spectrograms at the midpoint

frequency of 4 kHz, capturing prominent characteristics from both. This concatenated spectrogram effectively captures and represents regions of high amplitude and activity present in both types of spectrograms.

Figures 2(e) to 2(g) show pixel-wise combinations of Mel and inverse-Mel spectrograms using average, maximum, and minimum aggregation methods, respectively. These spectrograms successfully capture the shape and vivid colors characteristic of both Mel scale and inverse-Mel scale spectrograms. The intensity of the colors varies depending on the aggregation formula used in their construction.

Thus, the use of the inverse-Mel scale enables complete representation of the information present in IMS. The concatenation and combination spectrograms further support this representation. With these spectrograms, this research is able to delve into unexplored regions of IMS.
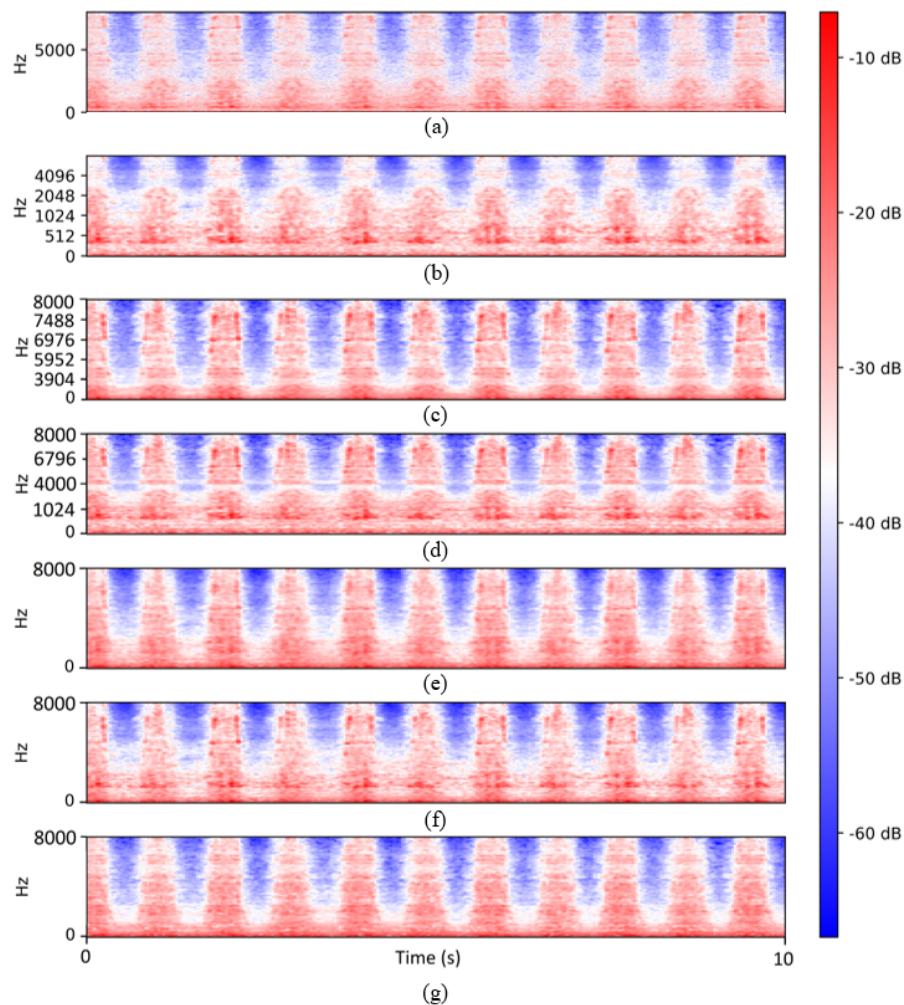

Figure 2. Spectrogram representations of a typical slide rail machine sound from the MIMII DG dataset: (a) linear-frequency spectrogram using STFT, (b) Mel scale spectrogram, (c) inverse-Mel scale spectrogram, (d) concatenated Mel and inverse-Mel spectrograms, (e) combined average spectrogram, (f) combined maximum spectrogram, and (g) combined minimum spectrogram

3.2. Inferences on the experiment results

Evaluations are conducted for all machine types, sections, and domains in the MIMII DG development and evaluation datasets [4]. Tables 1 and 2 present the scores and percentage improvements observed in the development datasets. Tables 3 and 4 present the scores and percentage improvements observed in the evaluation datasets. The source domain AUC, target domain AUC, and pAUC scores for all machine types, sections, and domains in the development datasets are listed in Tables 1(a) and 1(b). Tables 2(a) and 2(b) lists the percentage improvements for these scores relative to the results from Mel scale spectrograms.

Table 1. AUC and pAUC scores of all machines on development dataset (best values are highlighted) with (a) Mel scale, inverse Mel scale, and combination maximum; and (b) concatenated, combination average, and combination minimum

(a)

Section	Mel scale (AUC)			Inverse Mel scale (AUC)			Combination maximum (AUC)		
	source	target	partial	source	target	partial	source	target	partial
					Bearing				
0	0.5504	0.6048	**0.50737**	0.5613	**0.6915**	0.48922	0.5322	0.6103	0.50316
1	0.7176	0.5547	0.54869	0.5293	**0.7608**	0.4979	**0.751**	0.6068	**0.60395**
2	0.4563	0.5581	0.52316	0.415	0.5275	0.52764	0.4972	0.5695	0.55132
Average	0.57477	0.57254	0.52641	0.50187	**0.65994**	0.50492	**0.59347**	0.59554	0.55281
					Fan				
0	0.778	0.343	0.59158	0.7338	0.3745	0.59053	**0.7969**	0.3397	0.59237
1	0.7096	0.4577	0.51843	0.6691	0.4386	0.505	**0.7721**	0.4377	**0.53395**
2	0.7744	**0.6346**	0.62764	0.7712	0.5825	0.56606	**0.8985**	0.6093	**0.64369**
Average	0.754	0.47844	0.57922	0.7247	0.4652	0.55386	**0.8225**	0.46224	**0.59**
					Gearbox				
0	0.6558	0.6555	0.61369	**0.707**	0.7604	**0.63869**	0.6088	0.6981	0.61079
1	0.6605	0.5803	0.535	**0.6866**	0.6241	0.52737	0.6599	0.5707	0.51369
2	0.7744	0.6623	0.61711	0.8108	0.6928	**0.66053**	0.7484	0.6589	0.6079
Average	0.6969	0.6327	0.5886	**0.7348**	**0.69244**	**0.60886**	0.67237	0.64257	0.57746
					Slider				
0	0.8068	0.5681	0.61843	0.751	0.6088	**0.68264**	**0.8469**	0.5944	0.61237
1	0.6841	0.4969	0.53895	**0.7755**	0.5775	0.54632	0.678	0.4657	0.54579
2	0.8709	0.3866	0.53658	0.8809	0.4324	0.56158	**0.8838**	0.3431	0.525
Average	0.78727	0.48387	0.56465	0.80247	0.53957	0.59685	0.8029	0.46774	0.56106
					Valve				
0	0.5408	0.5182	0.52474	**0.5991**	**0.5506**	0.51158	0.5195	0.504	0.51974
1	0.5257	0.5313	0.50106	**0.5808**	**0.5951**	0.49527	0.5388	0.5083	0.49606
2	0.5187	0.4422	0.49395	**0.5891**	0.5008	0.49711	0.5635	0.4461	0.4879
Average	0.5284	0.49724	0.50658	**0.58967**	**0.54884**	0.50132	0.5406	0.48614	0.50123
Average overall	0.66827	0.53296	0.55309	0.6707	**0.5812**	0.55316	**0.68637**	0.53084	0.55651

(b)

Section	Concatenated			Combination average			Combination minimum		
	source	target	partial	source	target	partial	source	target	partial
					Bearing				
0	0.4945	0.6248	0.49369	0.4959	0.6138	0.5	**0.5617**	0.68	0.49158
1	0.6664	0.6121	0.57369	0.7009	0.6333	0.57685	0.5748	0.6511	0.55632
2	**0.5327**	**0.6153**	**0.59474**	0.4817	0.5834	0.49158	0.4791	0.5662	0.48922
Average	0.56454	0.6294	**0.55404**	0.5595	0.61017	0.52281	0.53854	0.63244	0.51237
					Fan				
0	0.7862	0.3522	0.58948	0.7808	0.3805	0.59343	0.6314	**0.4412**	**0.59343**
1	0.6763	**0.4758**	0.51527	0.6775	0.4517	0.5129	0.6876	0.402	0.52183
2	0.7506	0.5375	0.60974	0.7364	0.5882	0.6	0.5619	0.6061	0.59343
Average	0.7377	0.45517	0.5715	0.73157	0.47347	0.56878	0.62697	**0.4831**	0.56957
					Gearbox				
0	0.6615	0.6905	0.58895	0.4991	0.5832	0.49843	0.6484	0.7284	0.57237
1	0.66	0.5972	**0.54106**	0.6162	0.5455	0.52316	0.6709	0.5925	0.52474
2	0.7841	0.6822	0.61685	0.7656	0.678	0.63211	**0.8228**	0.6682	0.625
Average	0.70187	0.65664	0.58229	0.62697	0.60224	0.55123	0.71404	0.66304	0.57404
					Slider				
0	0.8127	0.5875	0.63685	0.7914	0.5787	0.64106	0.7855	**0.6671**	0.65027
1	0.7487	0.5592	0.55869	0.7096	0.5207	0.55316	0.7244	**0.6876**	**0.60922**
2	0.8766	0.4179	0.56711	0.8632	0.4327	**0.57737**	0.8689	0.387	0.54922
Average	**0.81267**	0.52154	0.58755	0.78807	0.5107	0.59053	0.79294	**0.50857**	**0.6029**
					Valve				
0	0.5599	0.5498	**0.52632**	0.564	0.5305	0.52185	0.5346	0.5224	0.5129
1	0.5188	0.5393	0.50053	0.5277	0.5314	**0.50343**	0.5155	0.5315	0.50237
2	0.5286	0.4699	0.49422	0.5494	0.4794	0.49685	0.5659	0.4973	**0.49843**
Average	0.53577	0.51967	0.50702	0.54704	0.51377	**0.50737**	0.53867	0.51707	0.50457
Average overall	0.67051	0.55648	**0.56048**	0.65063	0.54207	0.54815	0.64223	0.57524	0.55269

Table 2. Percentage improvements in scores on development dataset (best mean values are highlighted)
(a) inverse Mel scale, combination maximum, and concatenated; and (b) combination average and
combination minimum

(a)

Section	Inverse-Mel scale (AUC)			Combination maximum (AUC)			Concatenated (AUC)		
	source	target	partial	source	target	partial	source	target	partial
				Bearing					
0	1.99	**14.34**	-3.58	-3.31	0.91	-0.83	-10.16	3.31	-2.7
1	-26.25	**37.16**	-9.26	**4.66**	9.4	**10.08**	-7.14	10.35	4.56
2	-9.06	-5.49	0.86	8.97	2.05	5.39	**16.75**	**16.7**	**13.69**
Average	-12.69	**15.27**	-4.09	**3.26**	4.02	5.02	-1.78	9.94	**5.25**
				Fan					
0	-5.69	9.19	-0.18	**2.43**	-0.97	0.14	1.06	2.69	-0.36
1	-5.71	-4.18	-2.6	**8.81**	-4.37	**3**	-4.7	**3.96**	-0.61
2	-0.42	-8.21	-9.82	**16.03**	-3.99	**2.56**	-3.08	-15.31	-2.86
Average	-3.89	-2.77	-4.38	**9.09**	-3.39	**1.87**	-2.17	-4.87	-1.34
				Gearbox					
0	**7.81**	**16.01**	4.08	-7.17	6.5	-0.48	0.87	5.34	-4.04
1	**3.96**	**7.55**	-1.43	-0.1	-1.66	-3.99	-0.08	2.92	1.14
2	4.71	**4.61**	7.04	-3.36	-0.52	-1.5	1.26	3.01	-0.05
Average	**5.44**	**9.45**	3.45	-3.52	1.56	-1.9	0.72	3.79	-1.08
				Slider					
0	-6.92	7.17	**10.39**	4.98	4.63	-0.98	0.74	3.42	2.98
1	**13.37**	16.23	1.37	-0.9	-6.28	1.27	9.45	12.54	3.67
2	1.15	11.85	4.66	**1.49**	-11.26	2.16	0.66	8.1	5.69
Average	1.94	11.52	5.71	1.99	-3.34	-0.64	**3.23**	7.79	4.06
				Valve					
0	**10.79**	**6.26**	-2.51	-3.94	-2.29	-0.96	3.54	6.1	**0.31**
1	**10.49**	**12.01**	-1.16	2.5	-4.33	-1	-1.32	1.51	-0.11
2	**13.58**	**13.26**	0.64	8.64	-0.89	-1.23	1.91	6.27	0.06
Average	**11.6**	**10.38**	-1.04	2.31	-2.24	-1.06	1.4	4.52	0.09
Average over all machines	0.37	**9.06**	0.02	**2.71**	-0.4	0.62	0.34	4.42	**1.34**

(b)

Section	Combination average (AUC)			Combination minimum (AUC)		
	source	target	partial	source	target	partial
			Bearing			
0	-9.91	1.49	-1.46	**2.06**	12.44	-3.12
1	-2.33	14.17	5.14	-19.9	17.38	1.4
2	5.57	4.54	-6.04	5	1.46	-6.49
Average	-2.66	6.58	-0.69	-6.31	10.47	-2.67
			Fan			
0	0.36	10.94	0.32	-18.85	**28.63**	**0.32**
1	-4.53	-1.32	-1.07	-3.11	-12.17	0.66
2	-4.91	-7.32	-4.41	-27.45	-4.5	-5.46
Average	-2.98	-1.04	-1.81	-16.85	**0.98**	-1.67
			Gearbox			
0	-23.9	-11.03	-18.79	-1.13	11.13	-6.74
1	-6.71	-6	-2.22	1.58	2.11	-1.92
2	-1.14	2.38	2.44	**6.25**	0.9	1.28
Average	-10.04	-4.82	-6.35	2.46	4.8	-2.48
			Slider			
0	-1.91	1.87	3.66	-2.65	**17.43**	5.15
1	3.73	4.79	2.64	5.9	**38.38**	**13.04**
2	-0.89	**11.93**	**7.61**	-0.23	0.11	2.36
Average	0.11	5.55	4.59	0.73	**19.99**	**6.78**
			Valve			
0	4.29	2.38	-0.56	-1.15	0.82	-2.26
1	0.39	0.02	**0.48**	-1.95	0.04	0.27
2	5.92	8.42	0.59	9.1	12.47	**0.91**
Average	3.53	3.33	**0.16**	1.95	3.99	-0.4
Average over all machines	-2.64	1.71	-0.9	-3.9	7.94	-0.08

Table 3. AUC and pAUC scores of all machines on evaluation dataset (best values are highlighted)

		Harmonic mean over all machine types, sections, and domains	Official score
Mel scale	AUC	0.476997654	0.485524897
	pAUC	0.503527942	
Inverse-Mel scale	AUC	0.476953026	0.487278196
	pAUC	0.509330358	
Combination maximum	AUC	**0.490115307**	**0.495681532**
	pAUC	0.507202091	
Concatenated	AUC	0.475036924	0.485275108
	pAUC	0.507135059	
Combination average	AUC	0.46975013	0.481373758
	pAUC	0.506436574	
Combination minimum	AUC	0.475282691	0.486243539
	pAUC	**0.509755228**	

Table 4. Percentage improvements in scores on evaluation dataset (best mean values are highlighted)

		Harmonic mean over all machine types, sections, and domains	Official score
Inverse-Mel scale	AUC	-0.01	0.37
	pAUC	1.16	
Combination Maximum	AUC	**2.76**	**2.1**
	pAUC	0.73	
Concatenated	AUC	-0.42	-0.06
	pAUC	0.72	
Combination Average	AUC	-1.52	-0.86
	pAUC	0.58	
Combination Minimum	AUC	-0.36	0.15
	pAUC	**1.24**	

Use of plain inverse-Mel scale spectrograms has enhanced the target domain AUC in all machines, except for the fan machine. The most significant improvement, approximately 37%, is noted in the target domain AUC for the type 2 domain shift condition of the bearing machine. On average, there is about a 9% increase in the target domain AUC across all machines. Experiments conducted under various domain shift conditions show that the target domain AUC improves within a range of 5-36% for all machines, excluding the fan machine. Commonly occurring domain shifts—such as changes in microphone location (bearing machine section 2), varying loads (gearbox machine section 2), fluctuations in operational voltages (gearbox machine section 1), differences in operational speeds (bearing machine section 1), variations in operational velocity (slide rail machine section 1), changes in operational acceleration (slide rail machine section 2), differing operational patterns (valve machine section 1), and the mixing of various factory noises at different indexes (slide rail machine section 3)—are effectively identified by inverse-Mel scales. These improvements highlight the effectiveness of the inverse-Mel scale in accurately detecting operational and environmental domain shifts commonly encountered in IMS.

Use of plain inverse-Mel scale spectrograms has also improved the source domain AUC and pAUC in gearbox, slide rail, and valve machines. An average improvement of approximately 6% and 3% in source domain AUC and pAUC, respectively, is observed across the above three machines. These improvements prove the supremacy of the inverse-Mel scale in the detection of anomalous behavior from IMS.

Combinational maximum spectrograms are observed to enhance target domain AUC and pAUC scores in both the bearing and fan machines. Use of plain inverse-Mel scale spectrograms resulted in poor performance for these machines. This is due to the fact that bearing and fan machines produce a low level of sound energy, with the emitted energy primarily concentrated in the low-frequency regions. Nevertheless, the use of combinational maximum spectrograms has demonstrated improved detection accuracy. On average, there is an improvement of approximately 6% in source domain AUC and 4% in pAUC across both machines. Additionally, concatenated spectrograms are observed to enhance pAUC scores in bearing, slide rail, and valve machines, yielding an average improvement of around 3% in pAUC across these three machines. These results suggest that plain inverse-Mel scales may not always yield optimal results; however, the use of combinational or concatenated spectrograms could improve the performance of the detection system.

Evaluations are also conducted on the evaluation dataset. The official DACSE 2022 Challenge evaluator [27] is executed with the anomaly scores and decision results generated by the trained models. Harmonic

means of AUC and pAUC scores calculated across all machine types, sections, and domains of evaluation datasets are presented in Table 3. The official scores, as evaluated by the official evaluator, are listed in Table 3. The official scores are utilized to rank the participating systems and teams. Table 4 lists the percentage improvements in all the aforementioned scores in comparison to the results obtained from Mel scale spectrograms.

Among all the proposed methods, the combinational maximum spectrograms have been shown to generate the best values of harmonic mean for both AUC and pAUC scores across all machine types, sections, and domains. Improvements of approximately 3% in AUC scores and 1% in pAUC scores are observed. The official score for the combinational maximum spectrograms indicates an improvement of about 2% compared to the official score of Mel scale spectrograms. This enhanced score results in a rank of 21st in the official ranking released by the DCASE Challenge 2022 Task 2 [7]. This ranking is three positions higher than that of the baseline system.

4. CONCLUSION

In this work, inverse-Mel scales are used to capture the energy present in the high frequencies of IMS. This approach captures the information neglected by standard Mel scales. An autoencoder employing inverse-Mel scales, as well as the concatenation and combination of Mel and inverse-Mel scale spectrograms as front-end features, is implemented for AAD in industrial machines. Experiments are conducted on all machines in the MIMII DG datasets. The use of inverse-Mel scales, along with combinational maximum and concatenated spectrograms, has been shown to enhance source domain AUC, target domain AUC, and pAUC scores by 8%, 9%, and 2%, respectively, across all machines. The improvement in target domain AUC is particularly significant as it demonstrates the effectiveness of the proposed method in identifying challenging operational and environmental domain shifts. The higher ranking awarded by the official challenge evaluator in the evaluation datasets reflects the system's capability to effectively capture domain shifts. The results indicate that IMS contain a considerable amount of energy in higher frequency ranges that standard Mel scales fail to detect. Inverse-Mel scales are more efficient in capturing these high-frequency components and are hence advised to be used in AAD for industrial machines.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Kader Basha Tajuddin Shaikh	✓	✓	✓	✓	✓	✓	✓	✓	✓					
Naresh P. Jawarkar	✓	✓			✓	✓				✓		✓	✓	
Vasif Ahmed	✓									✓		✓		
Nadir Nizar Ali Charniya	✓									✓				

C	: Conceptualization	I	: Investigation	Vi	: Visualization
M	: Methodology	R	: Resources	Su	: Supervision
So	: Software	D	: Data Curation	P	: Project Administration
Va	: Validation	O	: Writing - Original Draft	Fu	: Funding Acquisition
Fo	: Formal Analysis	E	: Writing - Review & Editing		

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

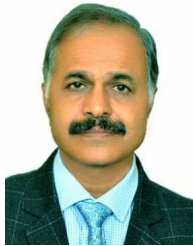The supporting data of this study are openly available at https://zenodo.org/record/6529888 [4].

REFERENCES

[1] K. B. T. Shaikh, N. P. Jawarkar, and V. Ahmed, "Machine diagnosis using acoustic analysis: a review," in *2021 IEEE Conference on Norbert Wiener in the 21st Century (21CW)*, Chennai, India: IEEE, Jul. 2021, pp. 1–6, doi: 10.1109/21CW48944.2021.9532537.

[2] T. Salm, K. Tatar, and J. Chilo, "Real-time acoustic measurement system for cutting-tool analysis during stainless steel machining," *Machines*, vol. 12, no. 12, Dec. 2024, doi: 10.3390/machines12120892.

[3] G. Li, X. Shang, L. Sun, B. Fu, L. Yang, and H. Zhou, "Application of audible sound signals in tool wear monitoring: a review," *Advanced Manufacturing Science and Technology*, vol. 5, no. 1, 2025, doi: 10.51393/j.jamst.2025003.

[4] K. Dohi *et al.*, "MIMII DG: Sound dataset for malfunctioning industrial machine investigation and inspection for domain generalization task," in *Proceedings of the 7th Detection and Classification of Acoustic Scenes and Events 2022 Workshop (DCASE2022)*, Nancy, France, Nov. 2022, pp. 1–5.

[5] K. Dohi *et al.*, "Description and discussion on DCASE 2022 challenge Task 2: unsupervised anomalous sound detection for machine condition monitoring applying domain generalization techniques," *Detection and Classification of Acoustic Scenes and Events 2022*, Nov. 2022, pp. 1-5.

[6] N. Harada, D. Niizumi, D. Takeuchi, Y. Ohishi, M. Yasuda, and S. Saito, "ToyADMOS2: Another dataset of miniature-machine operating sounds for anomalous sound detection under domain shift conditions," *arXiv-Electrical Engineering and Systems Science*, pp. 1-5, Jun. 2021, doi: 10.48550/arXiv.2106.02369.

[7] K. Dohi *et al.*, "Unsupervised anomalous sound detection for machine condition monitoring applying domain generalization techniques," *DCASE Community*, 2022. Accessed: Aug. 20, 2025. [Online]. Available: https://dcase.community/challenge2022/task-unsupervised-anomalous-sound-detection-for-machine-condition-monitoring.

[8] Y. Liu, C. Miao, X. Li, J. Ji, and D. Meng, "Research on the fault analysis method of belt conveyor idlers based on sound and thermal infrared image features," *Measurement*, vol. 186, Dec. 2021, doi: 10.1016/j.measurement.2021.110177.

[9] Y. Liu, Y. Chen, X. Li, X. Zhou, and D. Wu, "MPNet: A lightweight fault diagnosis network for rotating machinery," *Measurement*, vol. 239, Jan. 2025, doi: 10.1016/j.measurement.2024.115498.

[10] J. Liu, S. Fu, F. Liu, and X. Cheng, "Intelligent fault diagnosis of belt conveyor rollers using a polar KNN algorithm with audio features," *Engineering Failure Analysis*, vol. 168, Feb. 2025, doi: 10.1016/j.engfailanal.2024.109101.

[11] H. Zhou, Z. Gao, H. Li, and Y. Zhang, "State identifying method for rolling tire in lab test using acoustic signal," *Applied Acoustics*, vol. 231, Mar. 2025, doi: 10.1016/j.apacoust.2024.110487.

[12] Y. Zhao, B. Qin, Y. Zhou, and X. Xu, "Bearing fault diagnosis based on inverted Mel-scale frequency cepstral coefficients and deformable convolution networks," *Measurement Science and Technology*, vol. 34, no. 5, Feb. 2023, doi: 10.1088/1361-6501/acb0ea.

[13] L. Ma, A. Jiang, and W. Jiang, "The intelligent diagnosis of a hydraulic plunger pump based on the MIGLCC-DLSTM method using sound signals," *Machines*, vol. 12, no. 12, Nov. 2024, doi: 10.3390/machines12120869.

[14] S. Young *et al.*, *The HTK book*, Cambridge, United Kingdom: Cambridge University Engineering Department, 2002.

[15] M. Slaney, "Auditory toolbox: a MATLAB toolbox for auditory modeling work," *Interval Research Corporation*, pp. 1-41, 1998.

[16] Latha, "Robust speaker identification incorporating high frequency features," *Procedia Computer Science*, vol. 89, pp. 804–811, 2016, doi: 10.1016/j.procs.2016.06.064.

[17] H. K. Kathania, S. Shahnawazuddin, W. Ahmad, and N. Adiga, "Role of linear, mel and inverse-mel filterbanks in automatic recognition of speech from high-pitched speakers," *Circuits Systems Signal Process*, vol. 38, no. 10, pp. 4667–4682, Oct. 2019, doi: 10.1007/s00034-019-01072-7.

[18] S. Lalitha, S. Tripathi, and D. Gupta, "Enhanced speech emotion detection using deep neural networks," *International Journal of Speech Technology*, vol. 22, pp. 497–510, Sept. 2019, doi: 10.1007/s10772-018-09572-8.

[19] Z. Wang, J. Yan, Y. Wang, and X. Wang, "Speech emotion feature extraction method based on improved MFCC and IMFCC fusion features," in *2023 IEEE 2nd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*, Feb. 2023, pp. 1917–1924. doi: 10.1109/EEBDA56825.2023.10090810.

[20] S. Aziz and S. Shahnawazuddin, "Effective preservation of higher-frequency contents in the context of short utterance based children's speaker verification system," *Applied Acoustics*, vol. 209, June 2023, doi: 10.1016/j.apacoust.2023.109420.

[21] S. Aziz and S. Shahnawazuddin, "Experimental studies for improving the performance of children's speaker verification system using short utterances," *Applied Acoustics*, vol. 216, Jan. 2024, doi: 10.1016/j.apacoust.2023.109783.

[22] S. Aziz and S. Shahnawazuddin, "Role of data augmentation and effective conservation of high-frequency contents in the context children's speaker verification system," *Circuits Systems Signal Process*, vol. 43, pp. 3139–3159, May. 2024, doi: 10.1007/s00034-024-02598-1.

[23] S. Chakroborty, A. Roy, S. Majumdar, and G. Saha, "Capturing complementary information via reversed filter bank and parallel implementation with MFCC for improved text-independent speaker identification," in *2007 International Conference on Computing: Theory and Applications (ICCTA'07)*, Mar. 2007, pp. 463–467, doi: 10.1109/ICCTA.2007.35.

[24] S. Chakroborty, A. Roy, and G. Saha, "Improved closed set text-independent speaker identification by combining MFCC with evidence from flipped filter banks," *International Journal of Electronics and Communication Engineering*, vol. 2, no. 11, pp. 2554–2561, 2008.

[25] D. Sharma and I. Ali, "A modified MFCC feature extraction technique for robust speaker recognition," in *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Aug. 2015, pp. 1052–1057, doi: 10.1109/ICACCI.2015.7275749.

[26] B. McFee *et al.*, "Librosa: 0.10.0.post2," *GitHub*, 2023. [Online]. Available: https://github.com/librosa/librosa/releases/tag/0.10.0.post2

[27] K. Dohi, "Dcase2022 task 22 evaluator," *GitHub*, 2022. Accessed: Aug. 20, 2025. [Online]. Available: https://github.com/Kota-Dohi/dcase2022_evaluator

BIOGRAPHIES OF AUTHORS

Kader Basha Tajuddin Shaikh is currently pursuing a Ph.D. at Sant Gadge Baba Amravati University, Amravati. He holds both bachelor's and master's degrees in Instrumentation and Control Engineering from Mumbai University. Currently, he serves as an Assistant Professor in the Department of Automation and Robotics Engineering at Vivekanand Education Society's Institute of Technology, Mumbai. His research interests focus on the implementation of artificial intelligence and machine learning techniques in industrial applications. He is a certified LabVIEW associate developer (CLAD), a sun certified programmer for the Java platform (SCJP), and a Microsoft technology associate (MTA) in "Programming using Python". He can be contacted at email: kader.shaikh@ves.ac.in.

Naresh P. Jawarkar received his Ph.D. in Electronics from S.R.T.M. University, Nanded. He received his bachelor's degree in Electrical Engineering and master's degree in Integrated Power Systems from Nagpur University, Nagpur. He served as Assistant Professor, Associate Professor, and Professor at B. N. College of Engineering, Pusad, from July 1986 to August 2021. Currently, he is serving as an Adjunct Professor at Government College of Engineering, Amravati. He is a fellow of both the Institution of Engineers and the Institution of Electronics and Telecommunication Engineers. His research interests include speech signal processing, artificial neural networks, fuzzy logic, pattern recognition, and control systems. He can be contacted at email: npjawarkar@rediffmail.com.

Vasif Ahmed received bachelor's degree in Electrical and Electronics Engineering and master's degree in Industrial Electronics from the University of Baroda, Vadodara. He received Ph.D. in Electronics Engineering from Sant Gadge Baba Amravati University, Amravati. He served as Assistant Professor, Associate Professor, and Professor at B. N. College of Engineering, Pusad, from September 1994 to November 2024. He is a fellow of the Institution of Engineers and the Institution of Electronics and Telecommunication Engineers, and a life member of the Indian Society of Technical Education. His research interests encompass embedded systems, internet of things, and smart instrumentation. He can be contacted at email: ahmedvasif@gmail.com.

Nadir Nizar Ali Charniya received master's degree in Electronics Engineering from V.J.T.I., Mumbai, and Ph.D. in Electronics Engineering from Sant Gadge Baba Amravati University, Amravati. He has about 35 years of teaching and research experience. He is currently working as Professor in the Department of Electronics and Telecommunication Engineering at Vivekanand Education Society's Institute of Technology in Mumbai. His research interests include intelligent sensors and systems, neurocomputing techniques, signal processing, and their applications. He is a member of the Institution of Electronics and Telecommunication Engineers. He can be contacted at email: nncharniya@gmail.com or nadir.charniya@ves.ac.in.