# 迈向多模态工业故障分析：包含音频和振动信号的单速链条输送机数据集

张晨[1,*]，张宇聪[1,4,*]，缪潇潇[1],李明[1,2,3,**]

[1] 杜克昆山大学数字创新研究中心，中国昆山[2] 香港中文大学（深圳）人工智能学院，中国深圳[3] 武汉大学人工智能学院，中国武汉[4] 武汉大学计算机学院，中国武汉

zc199@duke.edu,ming.li.cuhksz@gmail.com

arXiv:2603.07130v1 [cs.SD] 7 Mar 2026

## 摘要

我们介绍了一个从单速链条输送机（SSCC）系统采集的多模态工业故障分析数据集，旨在检测生产线的系统级故障。该数据集包含多模态信号，包括三路音频和四路振动信号。数据集涵盖正常运行状态，以及在多种速度和负载条件下的四种代表性故障类型；同时，还包含在现场复现的无噪声条件和真实工厂噪声条件。该数据集专为支持通道级分析和多模态融合研究而设计。我们建立了标准化评估协议，涵盖仅使用正常数据训练的无监督故障检测，以及采用均衡数据集划分、覆盖不同运行条件和故障类型的有监督故障分类。我们还提供了统一的通道级 kNN 基线，以便在不针对特定任务进行训练的情况下，公平比较表征质量。该数据集为稳健的多模态工业故障分析提供了实用且可扩展的基准。

**索引词**：工业状态监测、故障分析、多模态数据集、音频与振动

## 1. 引言

早期故障分析是工业状态监测的关键组成部分，因为它能够及时安排维护，并防止自动化生产线发生意外系统故障 [1]。随着信号处理和机器学习技术的最新进展，故障分析系统日益采用数据驱动方法，其有效性在很大程度上取决于能否获得高质量且具有代表性的数据集。

尽管人们对数据驱动的工业状态监测兴趣日益浓厚，但大多数公开可用的故障分析数据集都是在实验室条件下采集的，且聚焦于有限类型的机器，例如电动机或滚动轴承。此外，许多现有数据集仅依赖单一传感模态，通常是音频或振动信号，因此可能无法捕捉多模态传感器所提供的互补故障相关信息。数据集往往完全排除噪声条件，或仅通过事后合成数据增强来引入噪声，导致基准评估与实际部署场景不匹配；在实际部署中，强烈且具有结构性的环境噪声不可避免。

为解决这些局限，我们推出了一个新的多模态工业故障分析数据集，该数据集采集自一条特定生产线的 SSCC 系统，记录了整个系统运行期间的行为。新的 SSCC 数据集包含从异构设备采集的同步多通道音频和振动记录，涵盖正常运行以及不同运行条件下的多种故障类型。虽然数据是在实验室环境中采集的，但通过对现场录制的工厂噪声进行受控复现，纳入了真实部署中的挑战，从而能够在噪声条件下开展系统且可复现的评估；这一研究方向在工业领域日益受到关注 [2]。此外，我们建立了统一的评估协议，支持无监督故障检测和有监督故障分类任务。所提出的 SSCC 数据集和协议共同构成了一个实用基准，可用于评估工业状态监测中的多模态信号表征与故障分析方法。代码[1], 以及附带演示的数据集[2]均已公开。

## 2. 相关工作

诸如凯斯西储大学（CWRU） [3]轴承故障数据集、IDMT-ISA 电动机（IIEE） [4] 数据集和 IDMT-ISA 压缩空气（IICA） [5]数据集等经典基准，主要针对基于振动的故障分析而设计，且是在受控实验室条件下采集的。这些数据集在推动数据驱动方法发展方面发挥了重要作用，但通常只关注有限种类的机器部件，例如轴承或电动机，并且依赖单一传感模态。

据我们所知，MaFaulDa [6] 和 HUSTmotor 多模态数据集 [7]是少数公开可用、同时提供振动和音频信号以用于工业故障分析的数据集。两者均基于旋转机械平台，并采用单个音频通道以及多轴振动测量。MaFaulDa 在相对无噪声的实验室环境中采集，未对环境噪声进行显式建模；而 HUSTmotor 引入噪声的方式主要是信号级加性处理，而非录制工厂环境中的真实声学干扰。

相比之下，SSCC 数据集构建于一套 SSCC 系统之上，并并行纳入三种不同的音频录制设备。此外，它将真实的工厂环境噪声融入数据采集过程。这一设计有助于在实际工业声学条件下对多模态故障分析进行更贴近现实的评估。

与聚焦于单个部件的数据集不同，SSCC 数据集面向完整的输送机系统，并采用系统级故障分析视角。在机器的不同位置布置了多个传

*这些作者贡献相同。**表示通讯作者。

[1]代码可于 https://github.com/Anonymous2417/SSCC-Fault-Benchmark.git 获取。[2]数据集和演示可于https://anonymous2417. github.io/SSCC-Dataset获取。

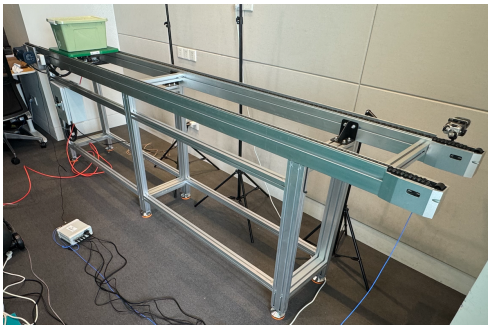

图 1：单速链式输送机实验平台的物理装置。

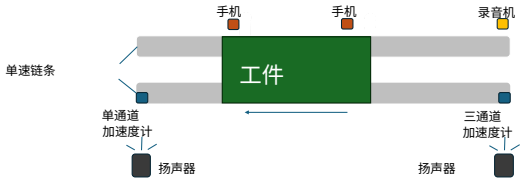


图2：总体系统布局和传感器配置。

感器。这种设置更贴近真实工业部署，因为故障可能在相互耦合的部件之间传播。该数据集支持在统一评估协议下开展故障检测和故障分类任务，并推动利用更丰富的通道多样性开展工业故障分析中的多模态学习研究。

## 3. 数据集概述

### 3.1. 机器系统描述

所提出的 SSCC 数据集采集自单速链式输送系统，这是一种生产线专用机械结构，广泛应用于工业输送和传动应用。与专注于轴承或电动机等独立部件的实验室测试台不同，本文所研究的系统代表了一种完整的链条驱动输送机构，常见于自动化生产线。

如图1所示，实验平台旨在模拟真实的生产线运行。在实际工业环境中，生产线通常较长且呈线性布局，物料单向流动，而非闭环循环 [8]。为体现这一特点，实验系统通过在输送机末端安装车间中常用的气动止挡机构，防止工件返回，从而确保运行过程中实现单向输送。

通过改变输送速度，在多种受控运行条件下采集数据。输送速度划分为五个等级（*20*、*40*、*60*、*80*和*100*）；负载等级分为（重、中等和轻）。这里，速度值是不同运行工况的名义指标，而非绝对物理单位；负载类别则表示 SSCC 系统的相对负载条件。这些运行属性独立于故障条件进行变化，并同时出现在正常状态和故障状态中。

### 3.2. 模态与传感器

SSCC 数据集包含两种感知模态，即音频和振动。图2展示了传感器布置和数据采集装置的概况。

使用三种录音设备采集音频信号 : a Zoom H5便携式录音机、iPhone 11 和小米 Mi 9 SE 智能手机。所有设备均安装在固定支架上，并朝向运行中的链条系统，以确保声学采集条件一致。由于麦克风特性和录音流程存在差异，使用多种设备自然会带来通道多样性。

除音频外，移动设备还会在数据采集期间录制视频。这些视频录制数据未用于本文所述的实验，而是作为补充数据发布，以支持未来可能开展的视觉分析或视听分析研究。

振动信号通过两种配置不同的传感器采集。单轴振动传感器安装在电机上，用于捕捉局部的电机相关振动；三轴振动传感器则安装在系统另一端，用于测量传动结构沿三个正交方向的系统级振动响应。

通过结合不同的音频设备与空间分布的振动传感器，所有感知模态均以同步方式记录，并作为并行通道处理，从而支持逐通道表征学习以及面向工业故障分析的多模态融合研究。

### 3.3. 噪声建模与注入

在真实工业环境中，环境噪声不可避免，并且已被证明会显著降低工业故障分析模型的性能 [9, 10]。因此，该数据集纳入了在实际工厂环境中录制的真实环境噪声。数据采集期间，通过放置在实验系统周围的两个扬声器重现录制的噪声，如图2所示。

对于每种运行工况，数据均在无噪声和含噪设置下采集，涵盖正常状态和故障状态。这种设计能够系统地评估噪声鲁棒性，并支持对模型在含噪工业条件下的性能进行可复现的基准测试。

### 3.4. 故障类型与标注

该数据集涵盖正常运行以及链式输送机和传动系统中常见的多种异常状态 [11, 12, 13]。具体而言，研究考虑了四种代表性故障类型：偏斜，对应导轨错位；干摩擦，表示链条润滑不足；松弛，指两侧链条过度松弛；以及螺丝滴，对应可能导致链传动路径局部阻塞或卡滞的异物侵入事件。

所有样本均带有样本级故障标签，可用于无监督故障检测和有监督故障分类研究。

### 3.5. 数据集组织与信号模态

在此数据集中，工况 定义为运行属性的特定组合，包括输送速度、负载等级、运行类别（正常或故障类型）以及噪声设置。因此，每种工况都对应单速链式系统的一种独特运行场景。

每种工况下，数据均在稳态运行期间采集，并切分为5秒片段。数据由所有感知设备同时采集，并作为同一样本的同步通道处理。

该数据集在所有运行条件下共包含6,669个样本。数据集的详细分布汇总于表1。对于低输送速度（20和40），数据仅在正常且无噪声的设置下采集，因为这类速度

表1： *SSCC* 数据集概览。行表示故障类别（定义见第 *3.4* 节）和负载条件，列对应无噪声和环境噪声设置下的输送速度。每个单元格列出样本数量，“–”表示该条件不可用。对于低速（*20* 和 *40*），数据仅在正常无噪声设置下采集，因为此类速度在实际工业运行中并不常见。

| 类别 | Load | 速度=20 | | 速度=40 | | 速度=60 | | 速度=80 | | 速度=100 | | 总计 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 无噪声 | 噪声 | 无噪声 | 噪声 | 无噪声 | 噪声 | 无噪声 | 噪声 | 无噪声 | 噪声 | |
| 正常 | 重 | 69 | – | 68 | – | 69 | 96 | 77 | 97 | 68 | 96 | 640 |
| 正常 | med | 73 | – | 66 | – | 78 | 96 | 67 | 94 | 66 | 99 | 639 |
| 正常 | 轻 | 68 | – | 86 | – | 69 | 97 | 66 | 97 | 67 | 97 | 647 |
| dry | 重 | – | – | – | – | 57 | 57 | 57 | 57 | 56 | 58 | 342 |
| dry | med | – | – | – | – | 69 | 57 | 56 | 57 | 64 | 56 | 359 |
| dry | 轻 | – | – | – | – | 62 | 58 | 49 | 56 | 57 | 57 | 339 |
| lean | 重 | – | – | – | – | 57 | 57 | 57 | 57 | 57 | 57 | 342 |
| lean | med | – | – | – | – | 58 | 56 | 57 | 57 | 56 | 55 | 339 |
| lean | 轻 | – | – | – | – | 57 | 56 | 69 | 60 | 56 | 56 | 354 |
| 松弛 | 重 | – | – | – | – | 96 | 99 | 97 | 97 | 97 | 96 | 582 |
| 松弛 | med | – | – | – | – | 73 | 96 | 96 | 96 | 96 | 99 | 556 |
| 松弛 | 轻 | – | – | – | – | 96 | 96 | 96 | 96 | 96 | 97 | 577 |
| 螺丝滴 | 重 | – | – | – | – | 57 | 56 | 57 | 55 | 56 | 54 | 335 |
| 螺丝滴 | med | – | – | – | – | 57 | 47 | 52 | 51 | 87 | 54 | 348 |
| 螺丝滴 | 轻 | – | – | – | – | 56 | 56 | 56 | 56 | 45 | 56 | 325 |
| **总体** | **合计** | **210** | – | **220** | – | **1,011 1** | **,080** | **1,009** | **1,083** | **1,024 1** | **,032** | **6,669** |

表2： 预训练音频编码器在不同视图下的故障检测性能（曲线下面积）。

| 特征 | 音频 | 振动 | 融合 |
|---|---|---|---|
| BEATs [14] | 0.868 | 0.784 | 0.891 |
| CED [15] | 0.835 | 0.772 | 0.854 |
| DaSheng [16] | 0.760 | 0.301 | 0.681 |
| EAT [17] | 0.747 | 0.291 | 0.666 |
| ECHO [18] | 0.807 | 0.201 | 0.658 |
| FISHER [19] | 0.954 | 0.546 | 0.882 |

在真实的工业运行条件下并不常见。

每个数据实例由通过异构传感器同时记录的七个信号通道组成。具体而言，其中包括三路立体声音频流：以 44.1 kHz 采样并以 WAV 格式存储的录音机音频、以 44.1 kHz 采样的 iOS 音频，以及以 48 kHz 采样的 Android 音频。尽管 iOS 和 Android 的录音以 MP4 格式存储，但建模流程仅使用其中的音轨。此外，振动数据以 CSV 格式存储，由四个相互独立的通道组成，采样率均为 100 kHz。

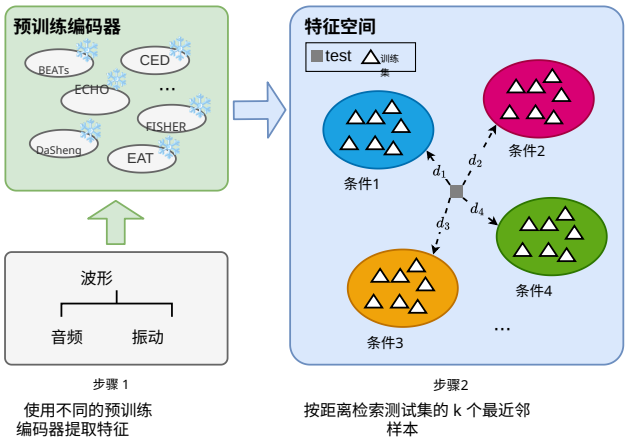


图 3： *SSCC* 数据集上的评估流程。对于异常检测任务，使用与最近邻之间的距离作为异常分数。对于故障诊断任务，根据 $k$ 个最近邻的标签进行多数投票，以确定预测结果。

# 4. 实验设置

## 4.1. 任务设计和数据划分

遵循工业故障分析领域的常见做法，我们考虑两项基本任务：故障检测和分类 [18, 19], , 并据此采用特定任务的数据划分策略。

故障检测采用仅正常样本训练机制进行评估，并针对运行速度进行零样本划分。

具体而言，训练集仅包含在非目标转速（vel $\neq 100$）下采集的正常样本，而测试集由在不同负载和噪声条件下以目标转速（vel $= 100$）采集的所有正常样本，以及来自所有运行工况的所有异常样本组成。

故障分类被设定为有监督多分类任务。为增加泛化难度，速度为 80 时采集的所有样本均完全不纳入训练。此外，速度为 100 时属于选定故障类别（干摩擦和偏斜）的样本也从训练集中移除，并重新划分至测试集。

数据划分在工况层面进行。属于同一工况的所有信号片段仅分配至训练集或测试集中的一方，以防止数据泄漏。

## 4.2. 评估协议

为确保在不同任务和数据集之间进行公平且统一的评估，我们采用基于 kNN 的推理框架作为故障检测任务的基准方法 [20]。这种选择能够评估所提取嵌入的内在质量，而不会引入额外的特定任务模型偏差；这是测试编码器能力的一种常见评估方法 [21]。

如图 3 所示，所有模型均在相同的基于 kNN 的流程下进行评估，特征嵌入由开源 SIREN 工具包 [18]，提取，该工具包为使用不同预训练模型提取特征提供了统一接口。正常数据集中的嵌入构成一个记忆库。推理时，将每个测试集嵌入与记忆库进行比较，并计算余弦距离。除非另有说

表3： 不同视图下预训练音频编码器的故障分类性能。*Acc* 表示总体准确率，*BalAcc* 表示各类别的平均召回率，*MacroF1* 表示各类别的平均 *F1* 分数。

| 特征 | View | Acc | 平衡准确率 | 宏平均 F1 | F1（干燥） | F1（倾斜） | F1（松动） | F1（正常） | F1（螺钉脱落） |
|---|---|---|---|---|---|---|---|---|---|
| BEATs [14] | 音频 | 0.914 | 0.900 | 0.906 | 0.954 | 0.935 | 0.908 | 0.959 | 0.774 |
| | 振动 | 0.869 | 0.873 | 0.884 | 0.985 | 0.984 | 0.803 | 0.839 | 0.811 |
| | 融合 | 0.941 | 0.928 | 0.937 | 0.996 | 0.996 | 0.903 | 0.993 | 0.796 |
| CED [15] | 音频 | 0.817 | 0.777 | 0.787 | 0.846 | 0.760 | 0.876 | 0.861 | 0.592 |
| | 振动 | 0.843 | 0.842 | 0.849 | 0.996 | 0.998 | 0.753 | 0.891 | 0.609 |
| | 融合 | 0.905 | 0.893 | 0.902 | 0.995 | 0.997 | 0.849 | 0.963 | 0.705 |
| DaSheng [16] | 音频 | 0.924 | 0.904 | 0.913 | 0.922 | 0.864 | 0.956 | 0.951 | 0.870 |
| | 振动 | 0.910 | 0.914 | 0.922 | 0.999 | 0.999 | 0.861 | 0.881 | 0.869 |
| | 融合 | 0.967 | 0.959 | 0.965 | 0.997 | 0.997 | 0.950 | 0.989 | 0.891 |
| EAT [17] | 音频 | 0.850 | 0.831 | 0.832 | 0.832 | 0.776 | 0.918 | 0.898 | 0.733 |
| | 振动 | 0.871 | 0.873 | 0.883 | 0.960 | 0.961 | 0.808 | 0.860 | 0.826 |
| | 融合 | 0.927 | 0.916 | 0.924 | 0.962 | 0.962 | 0.895 | 0.973 | 0.830 |
| ECHO [18] | 音频 | 0.975 | 0.971 | 0.973 | 0.963 | 0.973 | 0.983 | 0.983 | 0.963 |
| | 振动 | 0.715 | 0.690 | 0.682 | 0.707 | 0.809 | 0.720 | 0.784 | 0.391 |
| | 融合 | 0.907 | 0.886 | 0.894 | 0.954 | 0.963 | 0.871 | 0.990 | 0.693 |
| FISHER [19] | 音频 | 0.969 | 0.969 | 0.968 | 0.970 | 0.951 | 0.970 | 0.977 | 0.970 |
| | 振动 | 0.886 | 0.886 | 0.890 | 0.991 | 0.994 | 0.801 | 0.951 | 0.710 |
| | 融合 | 0.955 | 0.946 | 0.953 | 0.997 | 0.997 | 0.923 | 0.998 | 0.849 |

明，故障检测的最近邻数量固定为 $k = 1$，故障分类的最近邻数量固定为 $k = 11$；这一选择在平衡偏差-方差权衡的同时，能在不同设置下提供稳定且一致的性能。未针对特定数据集或模型进行参数调优。

故障检测和故障分类均采用统一的通道级 kNN 推理框架，覆盖七个通道。对于每个通道，使用仅基于训练正常样本拟合的通道专属标准缩放器对特征进行标准化。

为比较不同编码器的表征质量，我们选取六种预训练基础模型进行比较：BEATs [14], CED [15], EAT [17], Dasheng [16], FISHER [19], 和 ECHO [18]。所有模型均采用 ViT 风格的 [22] 架构作为骨干网络，并在涵盖多个领域的大规模开源音频数据集上进行训练。

对于每个测试样本，按通道计算其与第 $k$ 个最近邻的欧氏距离，并将其用作异常度量。为确保不同通道之间的可比性，使用基于经验累积分布函数（ECDF）的分位数映射对距离进行归一化 [23, 24], ，从而得到取值范围为 (0, 1) 的通道级异常分数。最终异常分数通过后期融合取平均值得出，并使用受试者工作特征曲线下面积（AUROC）评估性能；该指标常用于故障检测任务 [25]。对于故障分类，则为每个通道检索 $k$ 个最近邻，并通过距离加权投票计算类别概率。通道级概率向量经后期融合汇总，最终标签通过 arg max 得出。性能指标包括准确率、平衡准确率和宏平均 F1 分数。

## 5. 结果与分析

完成特征提取以及下游故障检测和分类任务后，结果如表2和表3所示。

在故障检测方面，基于音频的方法在大多数特征提取器上通常优于基于振动的方法，而振动在 BEATs 和 CED 上取得略高的性能。在所提出的仅正常样本训练与跨速度评估协议下，这表明声学信号能为所研究系统中的异常工况提供更具判别性的线索。一种合理的解释是，常见故障类型会引入细微的声音变化，这些变化更容易被音频表征捕捉，而振动信号往往被稳态运行中强烈的周期性成分所主导。决策层融合通常能够提升相对于仅振动视图的性能，并且在某些情况下达到或超过仅音频的结果，体现了结合音频与振动信号的益处。

在所有评估的特征提取器中，故障分类均取得了较高的准确率。在此任务中，音频和振动各具优势且互为补充：基于振动的特征在螺丝滴 故障上的表现更好，这可能是由于冲击性机械响应；而基于音频的特征在松弛 故障上的 F1分数始终更高，该故障与摩擦和 rattling 声有关。得益于这些互补特性，融合后的多模态系统在大多数情况下都比单模态系统表现更好，ECHO 和 FISHER 除外，这表明整合音频和振动信号有助于故障分类。

总体而言，这些结果揭示了该数据集中不同模态的有效性存在明显的任务依赖差异。在无监督评估协议下，音频信号对故障检测更具信息量；而带有明确标签的故障分类则受益于音频和振动共同捕获的互补物理线索。这表明，模态的贡献取决于目标任务和故障特征。尽管简单的基于距离的基线方法已能取得合理表现，但该数据集仍为未来探索更先进的声学表征学习和多模态融合方法提供了广阔空间。

## 6. 结论

本研究提出了一个包含七个同步传感通道的多模态工业故障分析数据集，采集了在 SSCC 系统中观测到的代表性故障条件。通过在多种运行工况下联合提供音频和振动测量，包括不同速度、负载水平和环境噪声，该数据集能够对速度链输送机故障分析的多模态方法进行系统且可复现的评估。我们采用统一的逐通道kNN评估探针，在特定任务协议下研究故障检测与分类，并且不同感知模态捕捉了故障行为的互补方面。这些发现表明，该数据集包含丰富的多模态信息，尚未被当前编码器充分利用，仍有可观的性能提升空间。因此，该数据集可作为工业故障分析的实用且可扩展基准，为未来更先进的表征学习与融合策略研究奠定坚实基础。

## 7. 生成式人工智能使用披露

生成式人工智能工具仅用于本稿的语法纠正和语言润色。所有研究内容、分析与结论均由作者独立完成。

## 8. 参考文献


[1] J. Garcia、L. Rios-Colque、A. Pe˜na 和 L. Rojas，《工业设备的状态监测与预测性维护：一篇关于信号处理、混合模型和实施挑战的 NLP 辅助综述》，《应用科学》，第15卷，第10期，第5465页，2025年。[2] D. A. Moysidis、G. D. Karatzinis、Y. S. Boutalis 和 Y. L. Karnavas，《考虑不同代表性特征模型的电机轴承故障检测与诊断中的噪声影响研究》，《*Machines*》，第11卷，第11期，2023年。[3] 凯斯西储大学，《轴承数据中心：预置故障测试数据》，凯斯工程学院。 [在线]。网址：https://engineering.case.edu/bearingdatacenter。[4] S. Grollmisch、J. Abeßer、J. Liebetrau 和 H. Lukashevich，《工业声音：工业声音分析的挑战与数据集》，载于*EUSIPCO*会议论文集，2019年，第1–5页。[5] D. Johnson、J. Kirner、S. Grollmisch 和 J. Liebetrau，《利用声发射和神经网络检测压缩空气泄漏》，载于*Inter-Noise*会议论文集，韩国首尔，2020年，第5662–5673页。[6] 信号、多媒体与电信实验室（SMT），《MAFAULDA：机械故障数据库》，SMT实验室。 [在线]。网址：http://www02.smt.ufrj.br/~offshore/mfs/page 01.html。[7] C. Zhao、W. Shen、E. Zio 和 H. Ma，《动态环境下用于智能故障诊断的多模态统一泛化与转换网络》，《人工智能工程应用》，第162卷，第112559页，2025年。[8] R. Deka、P. P. Borthakur、E. Baruah、P. Sarmah 和 M. Saikia，《机械输送系统综合综述：演进、类型与应用》，《国际工程科学杂志》，第18卷，第164–183页，2024年。[9] J.-G. Lee、K. S. Kim 和 J. H. Lee，《考虑环境噪声的基于声音的工业设备无监督故障诊断》，《*Sensors*》，第24卷，第22期，2024年。[10] Y. Yao、G. Gui、S. Yang 和 S. Zhang，《真实工业噪声环境下采用递归注意力机制的齿轮故障自适应抗噪网络》，《*Measurement*》，第186卷，第110169页，2021年。[11] H. Wen、B. Hou 和 X. Jin，《基于函数型数据特征工程与优化多层核极限学习机的链式输送机故障识别》，《机械科学与技术》，第37卷，2300，2023年5月，第2289–页。[12] L. Shuai、Z. Zhencai、L. Hao 和 S. Gang，《基于系统可靠性的刮板输送机刮板链设计优化：考虑相关故障模式》，*Eksploatacjai Niezawodno´s´c*，第21卷，第3期，第392–402页，2019年。[13] Z. Wang、B. Li、C. Liang、X. Wang 和 J. Li，《基于MBD-DEM-FEM的刮板输送机链条故障响应分析》。《机械工程学报》，第67卷，第10期，第501–516页，2021年。[14] S. Chen、Y. Wu、C. Wang、S. Liu、D. Tompkins、Z. Chen、W. Che、X. Yu 和 F. Wei，《BEATs：利用声学标记器进行音频预训练》，载于*ICML*会议论文集，第202卷，2023年，第5178–5193页。[15] H. Dinkel、Y. Wang、Z. Yan、J. Zhang 和 Y. Wang，《CED：用于音频标注的一致性集成蒸馏》，载于*ICASSP*会议论文集，2024年，第291–295页。

[16] H. Dinkel、Z. Yan、Y. Wang、J. Zhang、Y. Wang 和 B. Wang，《扩展掩码音频编码器学习以实现通用音频分类》，载于*Interspeech*会议论文集，2024年，第547–551页。[17] W. Chen、Y. Liang、Z. Ma、Z. Zheng 和 X. Chen，《EAT：采用高效音频Transformer的自监督预训练》，载于*IJCAI*会议论文集，2024年。[18] Y. Zhang、J. Liu 和 M. Li，《ECHO：面向变长信号的频率感知分层编码》，载于*ICASSP*会议论文集，2026年。[19] P. Fan、A. Jiang、S. Zhang、Z. Lv、B. Han、X. Zheng、W. Liang、J. Li、W.-Q. Zhang、Y. Qian、X. Chen 和 J. Liu，《FISHER：面向多模态工业信号综合表征的基础模型》，*arXiv*预印本 *arXiv:2507.16696*，2025年。[20] A. Jiang、B. Han、Z. Lv、Y. Deng、W.-Q. Zhang、X. Chen、Y. Qian、J. Liu 和 P. Fan，《AnoPatch：提升机器异常声音检测的一致性》，载于*Interspeech*会议论文集，2024年，第107–111页。[21] J. Zhang、H. Dinkel、Y. Niu、C. Liu、S. Cheng、A. Zhao 和 J. Luan，《X-ARES：音频编码器性能评估综合框架》，载于*Interspeech*，2025年，第4868–4872页。[22] A. Dosovitskiy、L. Beyer、A. Kolesnikov、D. Weissenborn、X. Zhai、T. Unterthiner、M. Dehghani、M. Minderer、G. Heigold、S. Gelly、J. Uszkoreit 和 N. Houlsby，《一张图像胜过16×16个词：用于大规模图像识别的Transformer》，载于*ICLR*会议论文集，2021年。[23] Z. Li、Y. Zhao、X. Hu、N. Botta、C. Ionescu 和 G. H. Chen，《ECOD：利用经验累积分布函数进行无监督离群点检测》，*IEEE*知识与数据工程汇刊，第35卷，第12期，第12 181–12 193页，2022年。[24] B. Ge、Y. Li 和 G. Yin，《DCCOPGAN：用于工业齿轮箱无监督多传感器异常检测的深度卷积Copula-GAN》，*Electronics*，第14卷，第13期，第2631页，2025年。[25] T. Nishida、H. Noboru、D. Niizumi、D. Albertini、R. Sannino、S. Pradolini、F. Augusti、K. Imoto、K. Dohi、H. Purohit、T. Endo 和 Y. Kawaguchi，《DCASE 2025挑战任务2说明与讨论：面向机器状态监测的首次无监督异常声音检测》，载于*DCASE*会议论文集，2025年，第55–59页。