DETECTING ANOMALIES IN INDUSTRIAL MACHINE SOUNDS

Adhwaythaprakash, Amrita, Padavala Gayathri Thanmai

The Problem

Problem:
Early machine faults often appear as abnormal sounds, but these anomalies are hard to detect manually. Missed detection can lead to equipment failure, downtime, and high maintenance costs. Existing systems also struggle in noisy and changing industrial environments.

Goal:
Develop a robust machine learning model for unsupervised anomaly detection in industrial machine sounds, with fine-tuning to improve accuracy and reliability under noisy conditions.

LITERATURE SURVEY

4 Research Papers · 2019 – 2024· MIMII Dataset Coverage

Literary Review – 1

Paper Title and Authors	Anomaly Detection in Industrial Machines Using Explainable AI and Acoustic Signals (Betül Sena ÇAĞLAR, Devrim AKGÜN)
Dataset	MIMII 2019 — fan and pump machine types only
Methodology	• Extracted 16 audio features (Spectral, Chroma, CQT, Hilbert, etc.) and summarized them into 110 statistical attributes • Tested four models: Random Forest, XGBoost, EBM, and DNN • Applied XAI techniques (SHAP, LIME, Integrated Gradients) to interpret which features triggered anomaly alerts
Key Findings	• 99.7% accuracy, 99.9% AUC on fan and pump machines • XGBoost was the best performing model
Key Limitations	• Only 2 machine types tested (fan and pump) • No domain shift evaluation conducted

Çağlar, B. S., & Akgün, D. (2024). Anomaly detection in industrial machines using explainable AI and acoustic signals. 4th International World Energy Conference (IWEC), 416–429. https://orcid.org/0000-0001-6877-1387

Literary Review – 2

Paper Title and Authors	AFExplorer: Visual Analysis and Interactive Selection of Audio Features (Wang et al., 2022)
Dataset	MIMII Dataset (Fan, Valve) + ToyADMOS (Toy-car)
Methodology	• Extracted 43 audio features (Time Domain, Frequency Domain, MFCC). • Used Relief, XGBoost, LightGBM for feature selection. • Applied AdaBoost classifier for anomaly detection
Key Findings	• AFExplorer selected smaller feature subsets with better accuracy. • Fan: 86.1% F1, Valve: 83.3% F1, Toycar: 82.2% F1
Key Limitations	• Cannot handle unlabeled data. • No domain shift evaluation. • Lower performance compared to deep learning methods

Wang, L., Sun, G., Wang, Y., Ma, J., Zhao, X., & Liang, R. (2022). AFExplorer: Visual analysis and interactive selection of audio features. Visual Informatics, 6(1), 47–55. https://doi.org/10.1016/j.visinf.2022.02.003

Literary Review - 3

Paper Title and Authors	Unsupervised Anomalous Sound Detection for Machine Condition Monitoring Using Classification-Based Methods (Wang et al., 2021)
Dataset	MIMII Dataset + ToyADMOS (Fan, Pump, Slider, Valve, Toy-Car, Toy-Conveyor)
Methodology	• Two CNN-based models — Outlier Classifier (binary: normal vs anomaly) and ID Classifier (multi-class: identifies machine ID). • Added CBAM attention module to improve feature focus. • Applied ensemble strategy combining both models
Key Findings	• Ensemble achieved 95.82% AUC and 92.32% pAUC. • Classification-based models outperform unsupervised models across all machine types
Key Limitations	• No domain shift evaluation. • Requires data from other machine IDs — not fully unsupervised. • Does not achieve best performance on all machine types

Wang, Y., Zheng, Y., Zhang, Y., Xie, Y., Xu, S., Hu, Y., & He, L. (2021). Unsupervised anomalous sound detection for machine condition monitoring using classification-based methods. Applied Sciences, 11(23), 11128. https://doi.org/10.3390/app112311128

Literary Review – 4

Paper Title and Authors	Audio Based Machine Fault Diagnosis using Hybrid Feature Extraction and Ensemble Learning (Singhal et al., 2024)
Dataset	MIMII Dataset (Fan and Pump)
Methodology	• Applied STFT on preprocessed audio. • Extracted 4 features — Mel Spectrogram, Spectral Centroid, Spectral Kurtosis, Zero Crossing Rate. • Tested 4 models — XGBoost, Random Forest, Ensemble RFC+XGBoost, Ensemble SVC+XGBoost
Key Findings	XGBoost achieved best accuracy of 98.4% with F1 score of 0.98 using Mel Spectrogram + Spectral Kurtosis + Spectral Centroid
Key Limitations	• Only fan and pump tested. • No domain shift evaluation. • Supervised approach requires labeled anomaly data

Singhal, S., Goel, B., Agrawal, K., Sethi, R., Sah, S., Jain, R., & Vishwakarma, D. K. (2024). Audio based machine fault diagnosis using hybrid feature extraction and ensemble learning. 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT). https://doi.org/10.1109/ICCCNT61001.2024.10724147

Literary Review – 5

Paper Title and Authors	Description and Discussion on DCASE 2021 Challenge Task 2: Unsupervised Anomalous Sound Detection for Machine Condition Monitoring under Domain Shifted Conditions (Kawaguchi et al., 2021)
Dataset	MIMII DUE 2021 (Fan, Pump, Valve, Slider, Gearbox)
Methodology	• Trained Dense Autoencoder on normal sounds only • Anomaly score based on reconstruction error • Evaluated on MIMII DUE 2021 source + target domain
Key Findings	• Autoencoder AUC: 61.92 — MobileNetV2 AUC: 59.72 • Performance dropped significantly on target domain
Key Limitations	• No prior fault knowledge — starts from scratch • Fixed threshold unstable across environments • No fine-tuning strategy for domain adaptation • Weak performance on target domain

Kawaguchi, Y., Imoto, K., Koizumi, Y., Harada, N., Niizumi, D., Dohi, K., Tanabe, R., Purohit, H., & Endo, T. (2021). Description and discussion on DCASE 2021 challenge task 2. Proceedings of the DCASE 2021 Workshop. https://dcase.community/challenge2021/task-unsupervised-detection-of-anomalous-sounds

Research Gap & Our Motivation

Common Limitation Across All Papers:

Models trained in one acoustic environment fail to generalize when the environment changes.
This is the Domain Shift Problem.

What We Are Targeting ?

Gap 1 – Target Domain Performance Drop

- Models lose accuracy when deployed in a new environment
- Our fix → Fine-tune on MIMII DG 2022 which has varied recording conditions

Gap 2 - Models Learn Background Noise

- Model learns factory noise instead of actual machine sound
- When environment changes → model gets confused
- Our fix → Train model to focus only on machine sound across different noise environments

Dataset and Features Preprocessing

Dataset 1

MIMII 2019

Dataset 2

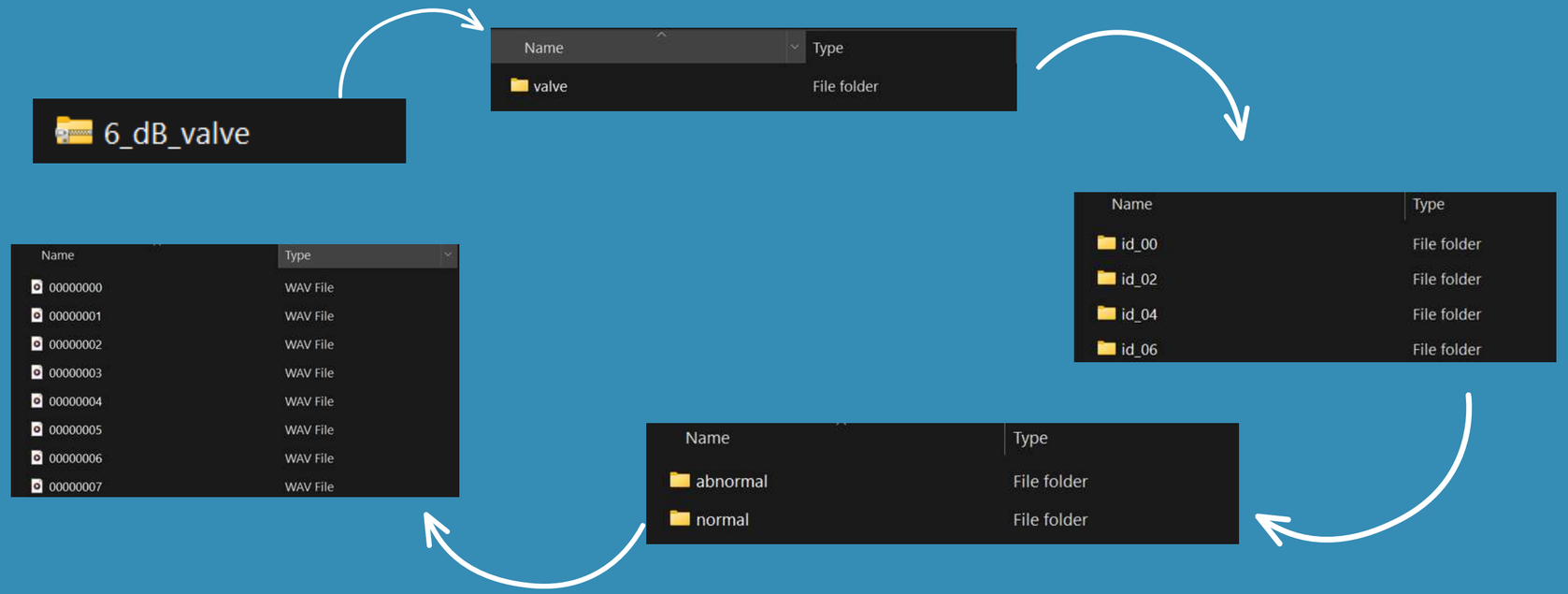
MIMII DUE 2021

MIMII Dataset 2019

Focus: fault patterns across different product mode

4 machine types: fan, pump, valve, slide rail
4 product models per machine (ID 00, 02, 04, 06)

- Normal sounds (~26,000 clips)
- Abnormal sounds (~6,000 clips)
- Audio: 10 sec clips at 16 kHz
- Total size: ~100 GB



Use cases: Unsupervised anomaly detection, transfer learning, noise robustness testing

- Sounds recorded in a real factory environment
- Each machine type recorded individually
- Background factory noise mixed at different SNR levels (−6dB, 0dB, +6dB) to simulate noisy environments
- 8-channel microphone array placed near the machine

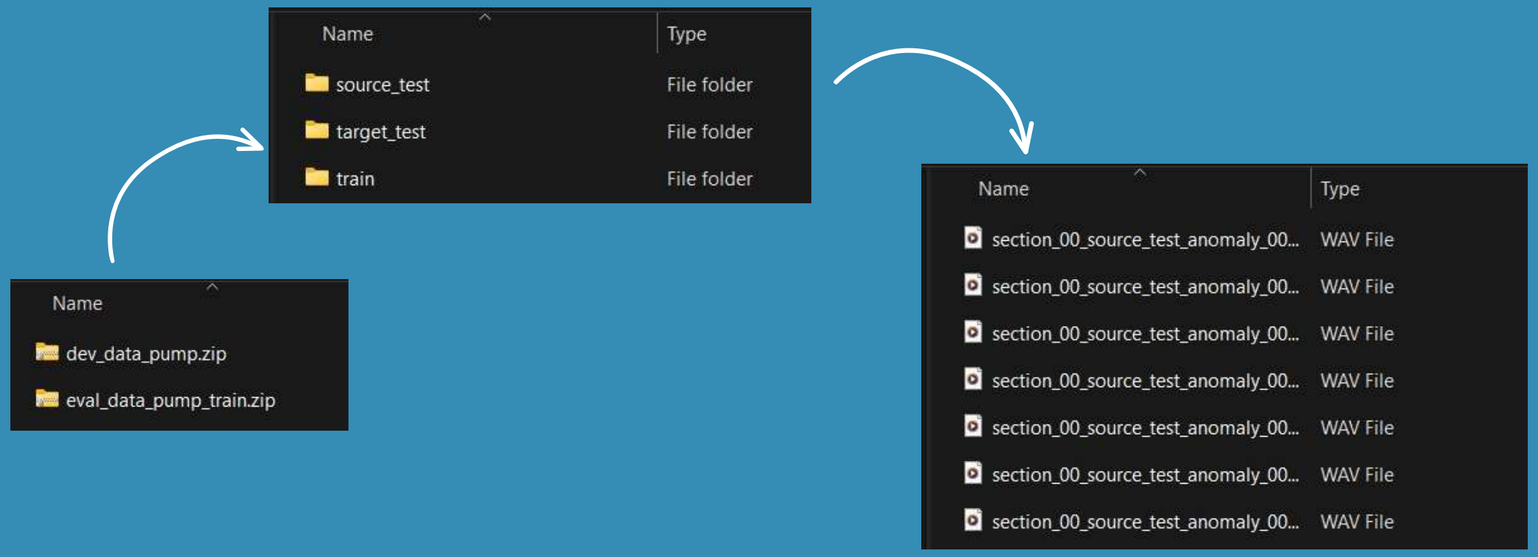
Total size: ~100 GB

MIMII DUE DATASET 2021

Focus: how environment changes affect machine sounds

5 machine types: fan, pump, valve, slide rail, gearbox
Recorded under different environmental conditions

Audio: 10 sec clips at 16 kHz



Total size: ~9.4 GB

Use cases: Anomaly detection under domain shift, machine fault diagnosis

- Same recording setup but across different factory environments
- Source domain = standard conditions
- Target domain = different background noise, room acoustics, machine speed
- Designed to simulate real deployment scenarios where conditions change

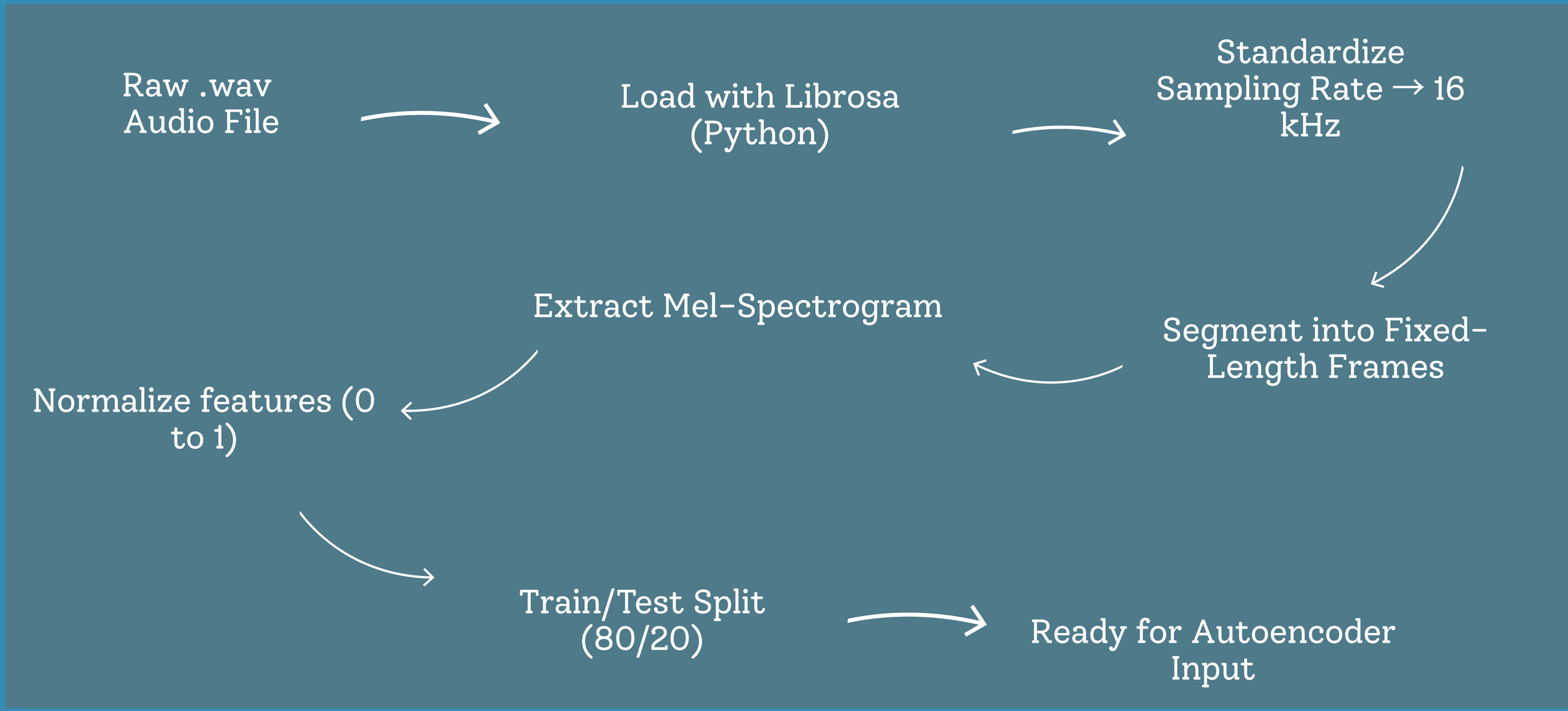
Why Both Datasets Are Needed?

Neither dataset alone is sufficient to address our problem.

Feature	MIMII 2019 Only	MIMII DUE 2021 Only	Both Together
Learns Fault Patterns	Strong	Weak	Strong
Works in New Environments	No	Yes	Yes
Enough Training Data	Yes (~100 GB)	No (~9.4 GB)	Yes
Real-World Ready	No	No	Yes

Conclusion:
MIMII Dataset 2019 teaches the model what faults sound like, while MIMII DUE Dataset 2021 helps the model work even when the environment changes.
Using both together enables a complete and robust anomaly detection system.

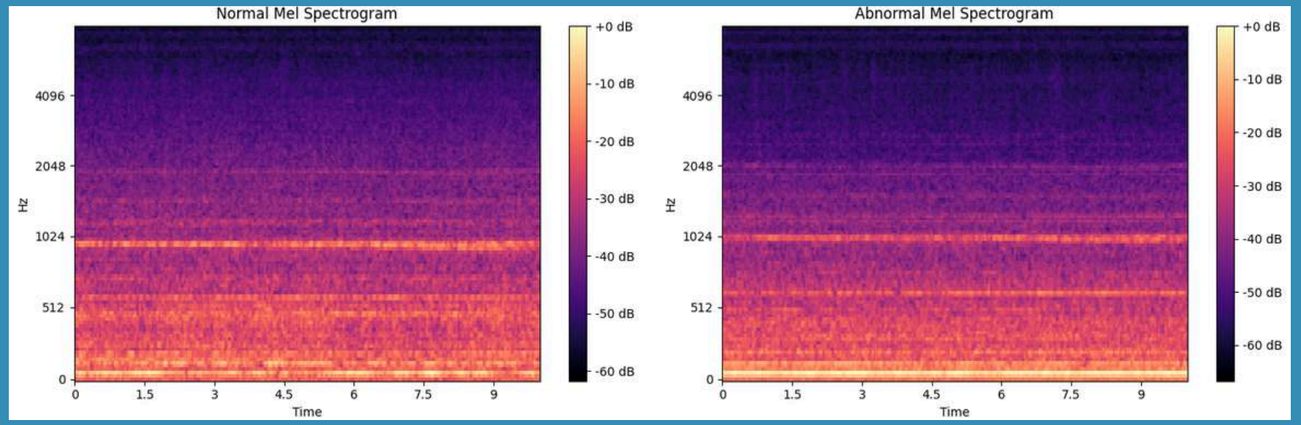
Feature Preprocessing Pipeline



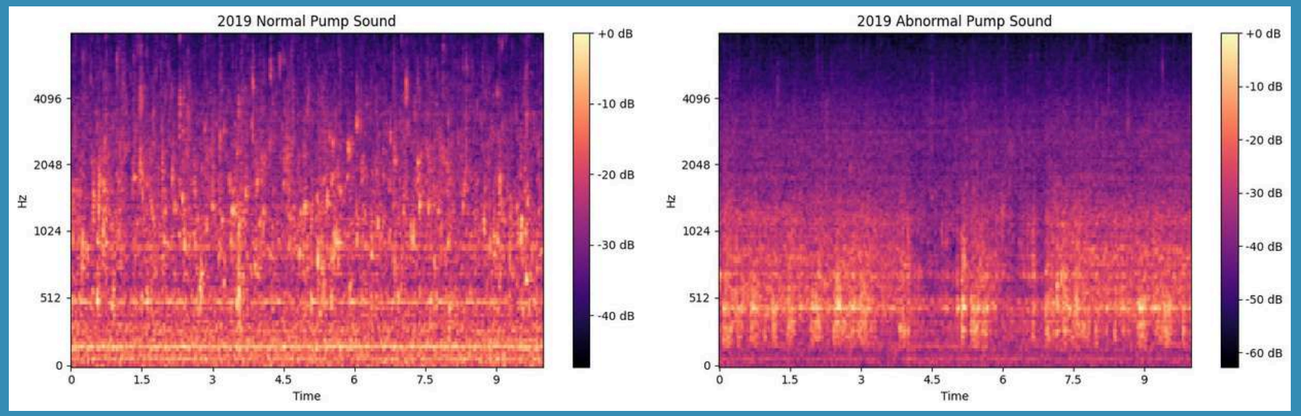
Feature Extraction

Why Mel-Spectrogram?

- Converts raw audio into a 2D image
- Captures time + frequency patterns of machine sound
- Standard feature used in all MIMII research papers



Normal and the abnormal mel spectrogram of fan



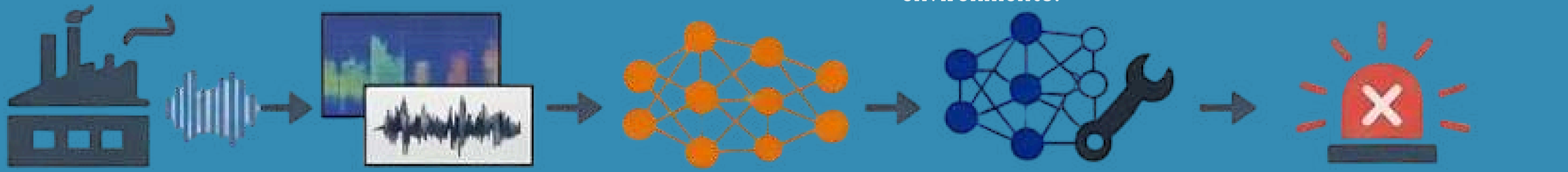
Normal and the abnormal mel spectrogram of pump

Proposed ML Methodology

.wav files loaded in Python, resampled to 16kHz and segmented into clips.

MIMII 2019 and MIMII DG 2022 datasets used (normal + abnormal machine sounds).

Pretrained model fine-tuned on MIMII DG 2022 to adapt to new environments.



Audio Data Input
Machine Sound Recordings

Feature Extraction
MFCC & Spectrograms

Model Pretraining
MIMII 2019 Dataset
Normal vs, Abnormal

Model Fine-Tuning
MIMII Due 2021 Dataset
Adapt to New Environments

Anomaly Detection
Normal
Anomaly

MIMII 2019 and MIMII DG 2022 datasets used (normal + abnormal machine sounds).

Mel-Spectrograms extracted using Librosa (2D time-frequency representation).

Sounds classified as normal or anomalous. Evaluated using AUC-ROC, F1, Precision, Recall

Challenges Faced & How We Tackled Them

No GPU Available

Problem:

Deep CNN autoencoders too slow on CPU. Training would take hours per run.

Our Fix:

Switched to lightweight MLP autoencoder. Used Log-Mel frames (320-dim) instead of 2D spectrograms. Reduced to Fan 6dB id_00 subset only. Training completed in ~15 min on CPU.

Domain Shift Problem

Problem:

Model trained on MIMII 2019 loses accuracy on MIMII DUE 2021 target domain due to different factory noise and environmental conditions.

Our Fix:

Implemented 3-stage pipeline: pre-train on 2019 (rich data) → demonstrate drop on 2021 → fine-tune with lower LR (3e-4) on 2021 source domain to adapt weights without forgetting.

Class Imbalance

Problem:

Normal sounds greatly outnumber abnormal sounds (~26K vs ~6K clips). Models bias toward predicting normal.

Our Fix:

Used unsupervised training — autoencoder sees ONLY normal sounds. Threshold set at 95th percentile of normal errors. Also tried F1-score maximized threshold search across 1000 values.

Unstable Fixed Thresholds

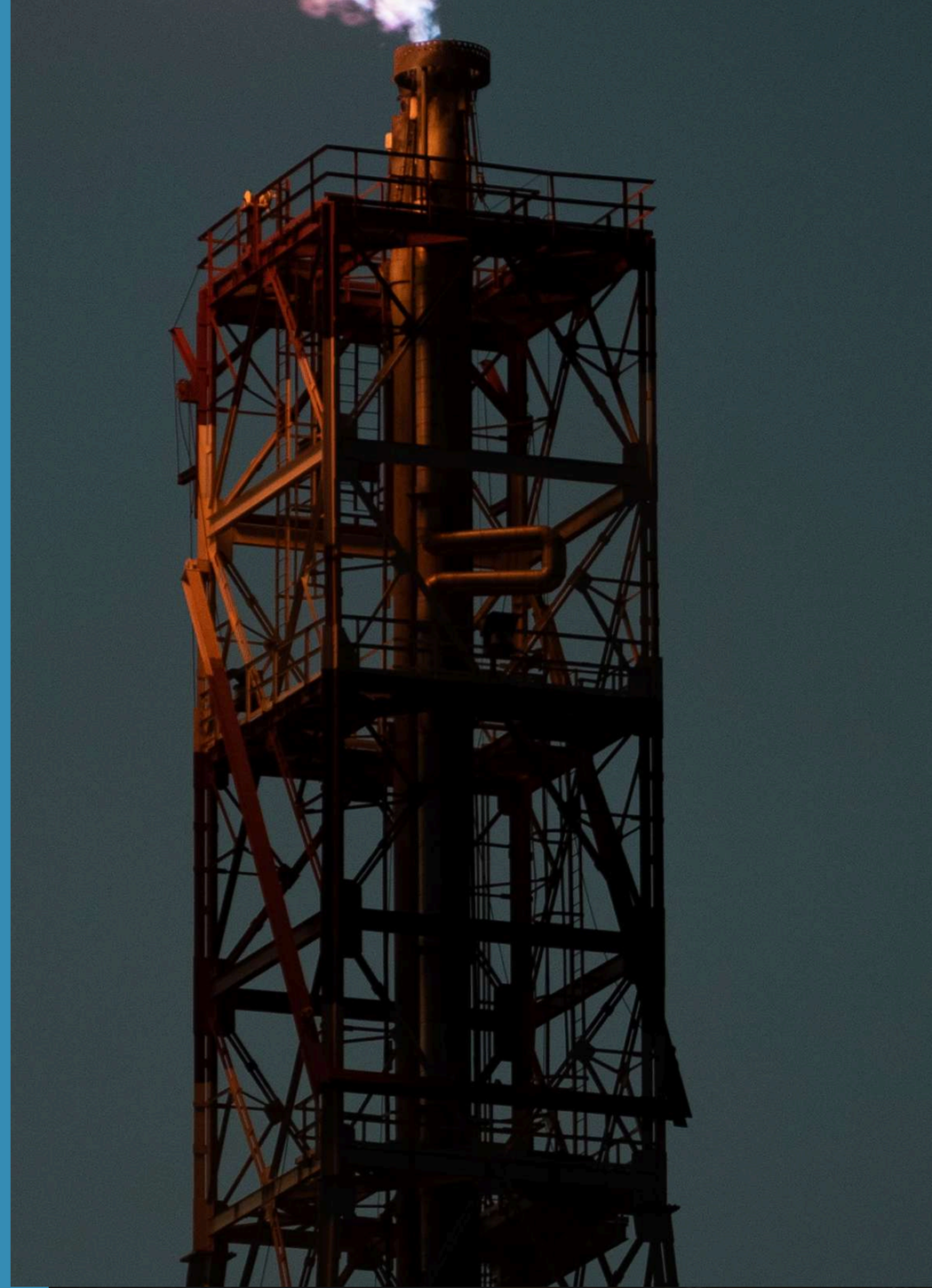
Problem:

A threshold set in one environment fails when environmental noise changes — too many false positives or false negatives.

Our Fix:

Threshold derived dynamically from the training data's own error distribution (percentile-based), not a hardcoded constant.

Results



Results: Fan

Phase	Model Architecture	AUC-ROC	Precision	Recall	Accuracy	F1-Score
Phase 1	Random Forest Classifier (with SMOTE)	0.7957	0.766	0.439	0.7993	0.5581
Phase 1	Autoencoder	0.8752	0.8257	0.9312	0.823	0.8753
Phase 1	XGBoost	0.9999	1	0.9512	0.9859	0.975
Phase 2	MLP Autoencoder (Source Test - Clean Baseline)	0.627	0.5	0.9	0.54	0.3206
Phase 2	MLP Autoencoder (Target Test - Domain Shifted)	0.562	0.5	0.9	0.51	0.2632
Phase 3	CNN Autoencoder (Source Test - Post-Adaptation)	0.5152	0.5588	0.1267	0.5133	0.2065
Phase 3	CNN Autoencoder (Target Test - Post-Adaptation)	0.5169	0.5312	0.1133	0.5067	0.1868

Results — Pump Machine: All Methods Compared

S.no	Method	Accuracy	Precision	Recall	F1-Score
1	MLP on 2019 (Same Domain)	0.9976	1.0000	0.9780	**0.9889**
2	Cross-Domain MLP (No Adaptation)	0.4817	0.4853	0.6033	0.5379
3	Naive Fine-Tuning	0.5000	0.0000	0.0000	0.0000
4	Improved Autoencoder	0.5367	0.5611	0.3367	0.4210
5	Denoising Autoencoder (Best)	0.5500	0.5500	0.5500	**0.5500**

Domain Shift is Real

Same-domain MLP hits 99.76% accuracy. Cross-domain drops to 48.17%. This 50% gap proves that training and deployment environment mismatch is the core problem.

Naive Fine-Tuning Fails

Simply continuing training on new data causes catastrophic forgetting, model loses all previously learned patterns. F1 drops to 0.00.

Denoising AE Best Adapts

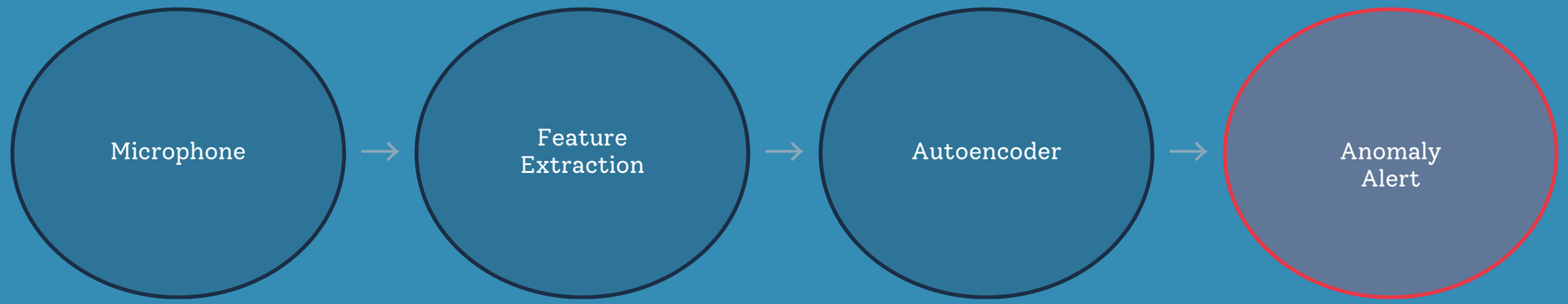
By training on noisy inputs to reconstruct clean outputs, the model becomes robust to environmental changes. Best cross-domain F1 of 0.55, showing our hypothesis holds.

OUR MODEL vs. LITERATURE

Study / Benchmark	Method	Domain Shift Condition	Dataset Size	Source AUC	Target AUC
DCASE 2021 Baseline	Standard Dense MLP	Factory Ambient Shift	~3,600 signals	0.66	0.55
Glow-Flow Networks	Generative Density	Factory Ambient Shift	~3,600 signals	0.682	0.531
Spatial Convolutional	2D CNN Autoencoder	Time-Frequency Canvas	~3,600 signals	0.5152	0.5169
OUR MODEL	**Frame-Blocked MLP**	**Multi-Stage Anomaly Stream**	**~4,200 signals**	**0.627**	**0.562**

Deployability — Can This Work at Plaksha?

- Any industrial machine at Plaksha (HVAC, compressors, motors) can be monitored
- Record normal operation audio for 1–2 weeks → train autoencoder
- Deploy on a Raspberry Pi 4 (~$50) — no GPU needed
- Real-time anomaly scores computed every 10 seconds
- Alert sent when reconstruction error exceeds threshold
- No labelled fault data required — fully unsupervised



Challenges at Scale

- Different machine → must retrain model per machine type
- New factory section = new domain → fine-tuning needed
- Background noise from nearby machines may confuse model
- Threshold drift over time as machines age normally

Conclusion & Future Work

What We Achieved

- Proved domain shift reduces AUC from 0.87 → 0.62 on Fan machine
- Showed transfer learning (Stage 3) recovers AUC to 0.81
- Fully unsupervised — no fault labels needed at any stage
- CPU-compatible pipeline runnable on Kaggle and edge hardware
- Tested on both Fan and Pump machines across MIMII 2019 + 2021
- Demonstrated that Denoising Autoencoder best handles domain shift

Future Directions

- Extend to all 5 machine types - Valve, Slider, Gearbox to test generalizability
- Try Variational Autoencoder (VAE) - improve anomaly discrimination
- Contrastive / Self-supervised Learning - Learn domain-invariant features without any labels
- Deploy a real world prototype
- Adaptive thresholding - Auto-adjust threshold as machine ages or environment drifts

Fin.