# Fault Detection in Air Compressors Using Acoustic Signals: A Python-Based Implementation


## Abstract

This report presents a Python-based implementation of an acoustic signal classification pipeline for air compressor fault detection. It reproduces key aspects of the IEEE methodology proposed by Verma et al. for condition-based monitoring. The pipeline includes signal loading, multi-domain feature extraction (FFT, DCT, WPT, STFT, time-domain), feature selection, and classification using a Support Vector Machine (SVM). The model achieves high accuracy on the IIT Kanpur dataset and demonstrates the effectiveness of multi-domain features in fault classification.


## 1. Dataset Description

Source: IIT Kanpur air compressor acoustic dataset.

Sampling: 50 kHz, 5 seconds per recording (250,000 samples).

Classes: 8 conditions (1 healthy + 7 faults: inlet valve, outlet valve, non-return valve, piston ring, flywheel, rider belt, bearing).

Size: 225 recordings per class (1,800 samples total).

Format: 24-bit PCM .dat files (comma-separated values).

The load_signal_dataset() function reads all .dat files under each class folder, extracting the acoustic waveform into a list X and class labels into y. Each class is perfectly balanced.

## 2. Preprocessing

Standard normalization:

- StandardScaler scales each feature to zero mean and unit variance before training.
- Improves SVM performance and ensures consistency in mutual information–based feature selection.

Note: Advanced preprocessing steps (bandpass filtering, windowing, smoothing, robust normalization) described in the IEEE paper are not implemented here.

## 3. Feature Extraction

Multi-domain features extracted from each signal:

- FFT (8 features): compute magnitude spectrum via FFT; divide spectrum into 8 bins and average energy in each bin.
- DCT (40 features): retain first 40 coefficients from Discrete Cosine Transform.

WPT (8 features): 3-level Wavelet Packet Transform using Daubechies-1 (db1); compute mean absolute values of 8 leaf nodes.

STFT (10 features): segment length = 64; mean magnitude over time for first 10 frequency bins.

Time-Domain Statistics (4 features): mean, standard deviation, kurtosis, skewness.

Total: 70 features per recording.

## 4. Feature Scaling and Selection

Scaling: All features normalized with StandardScaler.

Selection: SelectKBest with mutual_info_classif retaining top 25 features.

Note: SelectKBest is simpler and computationally efficient compared to mRMR but still captures highly informative features.

## 5. Classification (SVM Model)

Classifier: Support Vector Machine (SVM) with RBF kernel.

We employed a grid-search strategy over

- $\mathbf{C} \in \{0.1, 1, 10, 100, 1000\}$
- $\boldsymbol{\gamma} \in \{0.001, 0.01, 0.1, 1\}$

Using 5-fold cross-validation on the training set, the best average validation accuracy was achieved at $\mathbf{C = 100}$ and $\boldsymbol{\gamma} = \mathbf{0.01}$.

Hyperparameters: C = 100, $\gamma$ = 0.01.

Data split: 80% training (1,440), 20% testing (360), stratified by class.

Evaluation: classification report, confusion matrix, 5-fold CV accuracy, training time (~0.04s), prediction time (~0.01s).

## 6. Results

Test Accuracy: ~98.89%.

Mean 5-Fold CV Accuracy: ~97.5%.

Observations: high precision and recall across all 8 classes; minimal confusion between fault classes.

## 7. Comparison with IEEE Methodology

Table comparing techniques between Verma et al. and this implementation:

## 8. Conclusion

This implementation replicates the core idea of condition monitoring through acoustic features and classification. Using FFT, DCT, WPT, STFT, and time-domain features, the system achieves high accuracy with minimal training time. Simplifications still result in performance comparable to Verma et al.'s IEEE study. Future work could incorporate advanced preprocessing or feature engineering techniques.

## Table: Comparison of Techniques

| Aspect | IEEE Verma et al. | This Implementation |
|---|---|---|
| Preprocessing | Filtering, clipping, smoothing, robust normalization | Standard normalization only |
| FFT, DCT, WPT, STFT | Yes | Yes |
| Time-Domain Features | 8 features | 4 features (mean, std, kurtosis, skewness) |
| Wavelet Levels | 7 (db4) | 3 (db1) |
| Feature Selection | mRMR, NMIFS | SelectKBest (mutual information) |
| SVM Tuning | Extensive (grid search) | Grid search |
| Accuracy | >99% | ~98.89% |

## References


1. Nishchal K. Verma, R. K. Sevakula, S. Dixit and A. Salour, "Intelligent Condition-Based Monitoring Using Acoustic Signals for Air Compressors," IEEE Transactions on Reliability, 2016.
2. MathWorks: Acoustics-Based Machine Fault Recognition