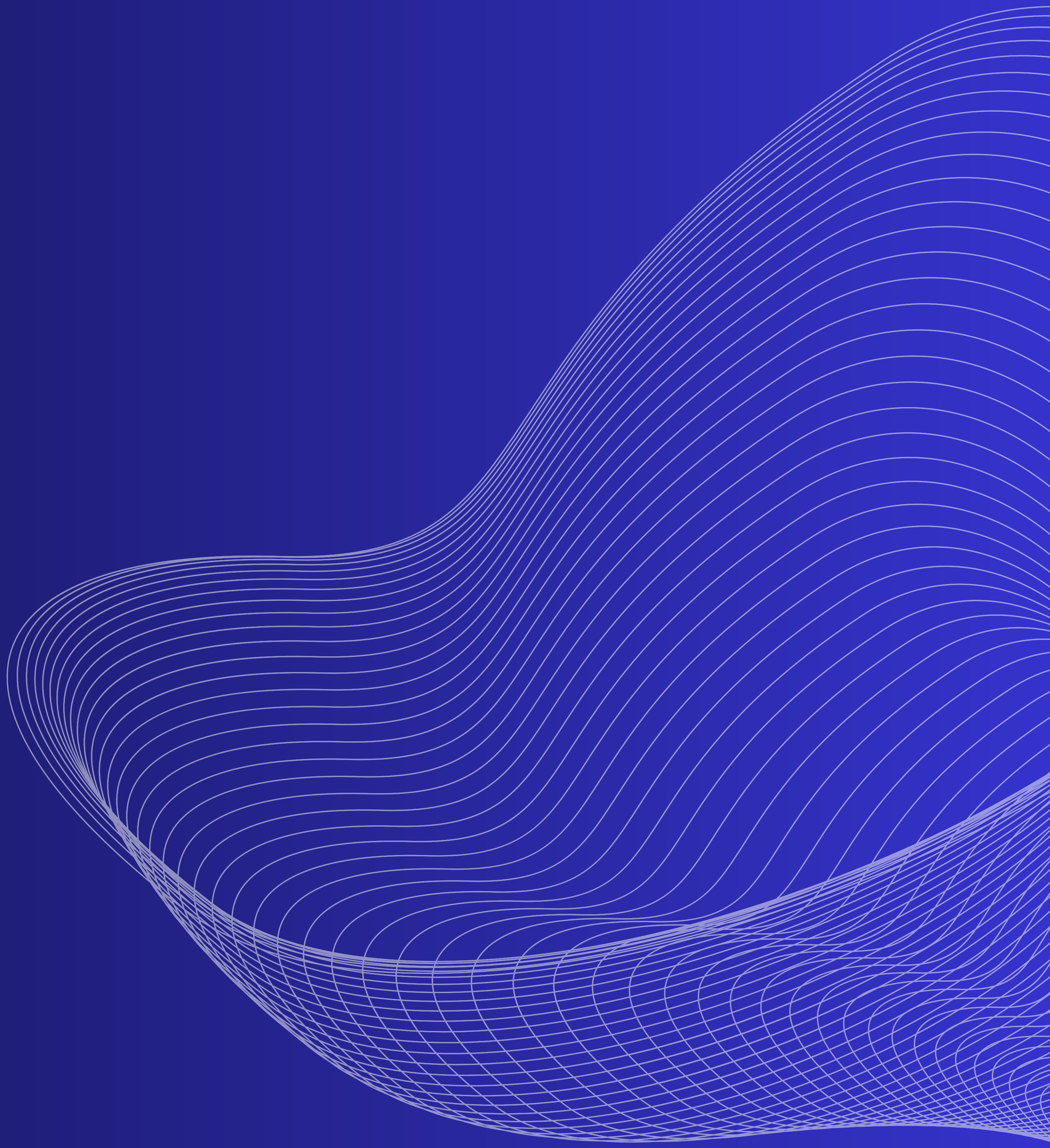
GROUP 26

CHALLENGE-2

Anomalous Sound Detection



ANOMALOUS SOUND DETECTION

INTRODUCTION

ASD involves determining if a machine's emitted sound is normal or anomalous, essential for AI-based factory automation in the fourth industrial revolution. Prompt detection of machine anomalies by observing sounds is useful for monitoring the condition of machines.

HOW TO SOLVE IT?

1. Data Exploration

2. Data Preparation

3. Model Selection

4. Model Performance

DATA EXPLORATION

INTRODUCTION

- Each recording is a single-channel audio lasting around 10 seconds.
- The recordings capture both the operating sound of the target machine and the environmental noise in the surroundings.
- The data is sourced from ToyADMOS and the MIMII Dataset.
- The dataset consists of normal and anomalous operating sounds from six types of toy and real machines.
- Our focus will be solely on the Slide rail machine.

DATA EXPLORATION

INTRODUCTION

Additional types of toy and real machines include:

- **ToyADMOS**
  - Toy-car
  - Toy-conveyor

- **MIMII Dataset**
  - Fan
  - Pump
  - Slide Rail
  - Valve

DATA EXPLORATION

INTRODUCTION

Power Spectral Density: Analyzing the distribution of power across different frequencies in the sound signal to understand its frequency characteristics.

Waveform: Visual representation of the sound signal in the time domain, showing the amplitude of the signal over time.

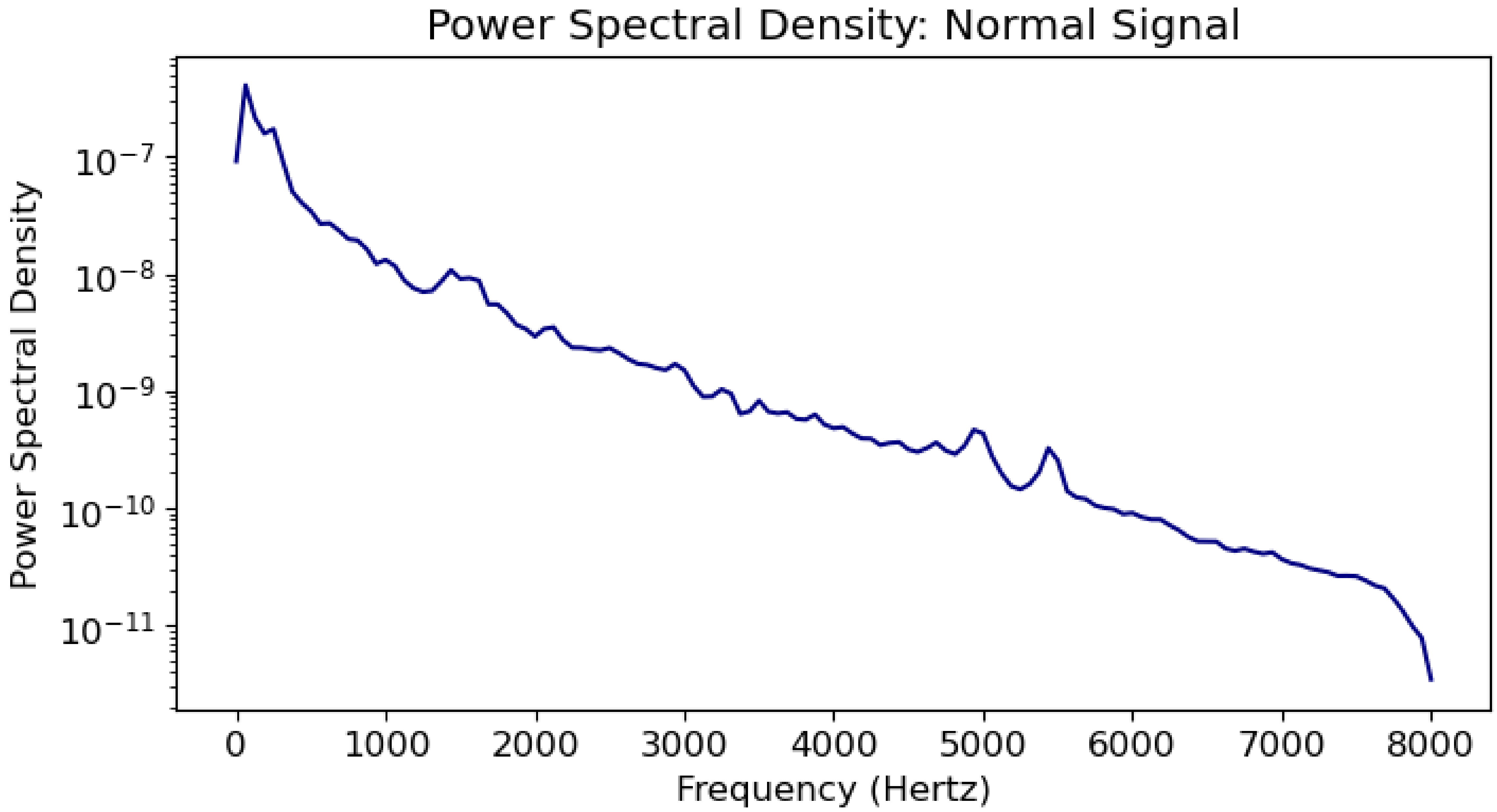
Spectrogram: A 2D representation of the sound signal that displays how the frequency content of the signal changes over time.

Chromagram: A representation of the sound signal that highlights the presence of different musical notes or pitches over time.

Mel Spectrogram: A spectrogram that uses the Mel scale to transform the frequency axis, providing a more perceptually relevant representation of the sound signal's frequency content.

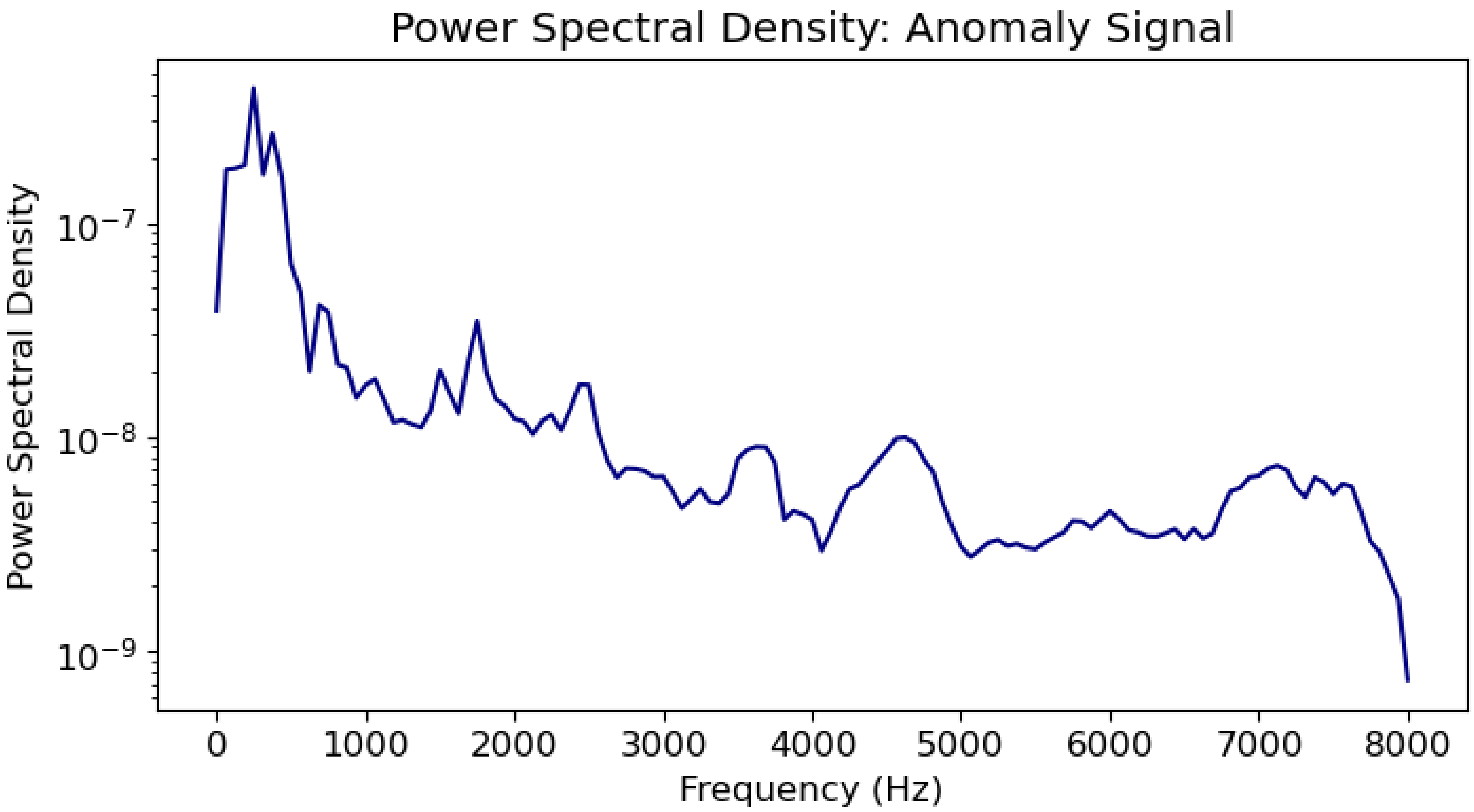
DATA EXPLORATION

POWER SPECTRAL DENSITY



DATA EXPLORATION

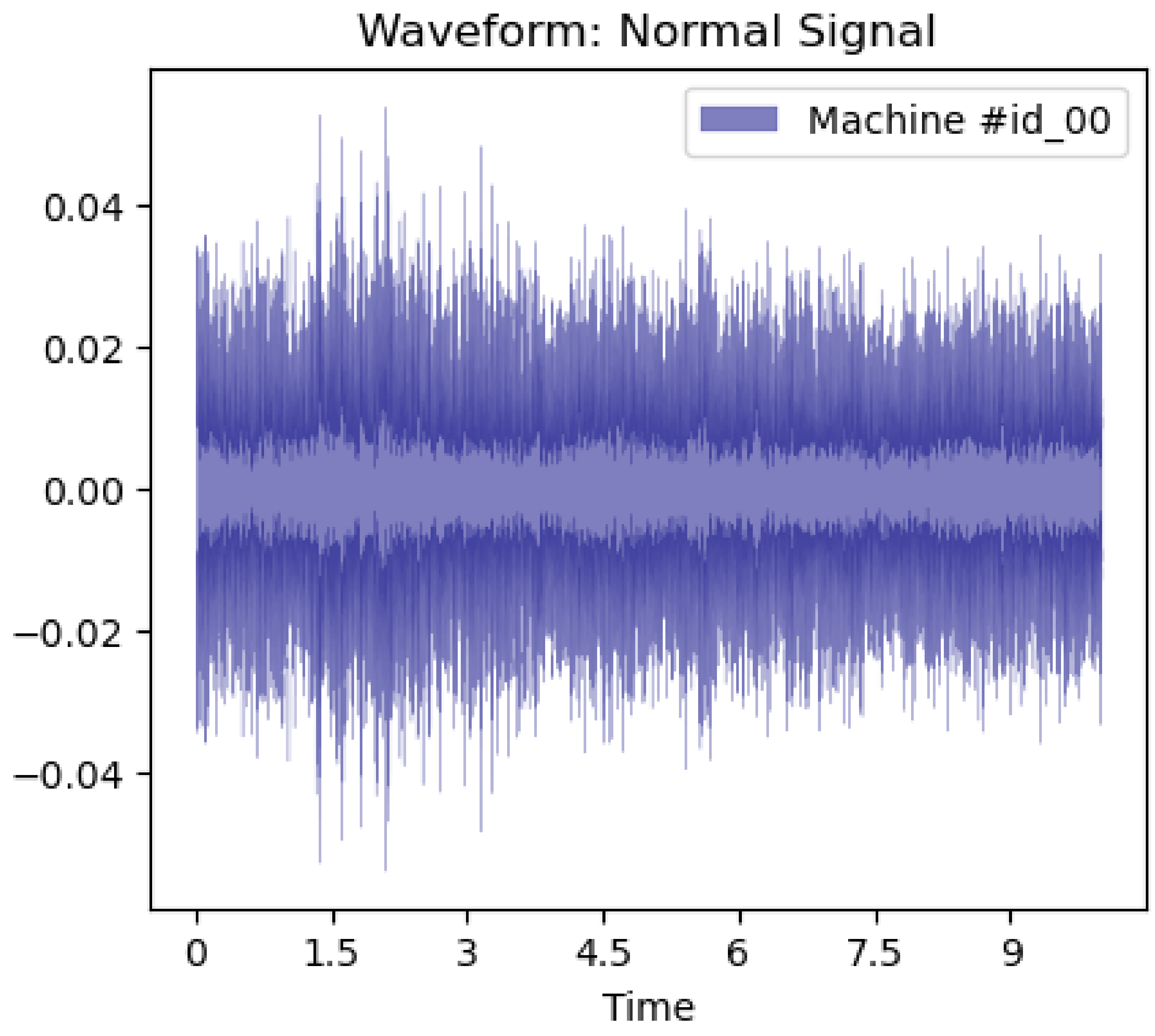
POWER SPECTRAL DENSITY

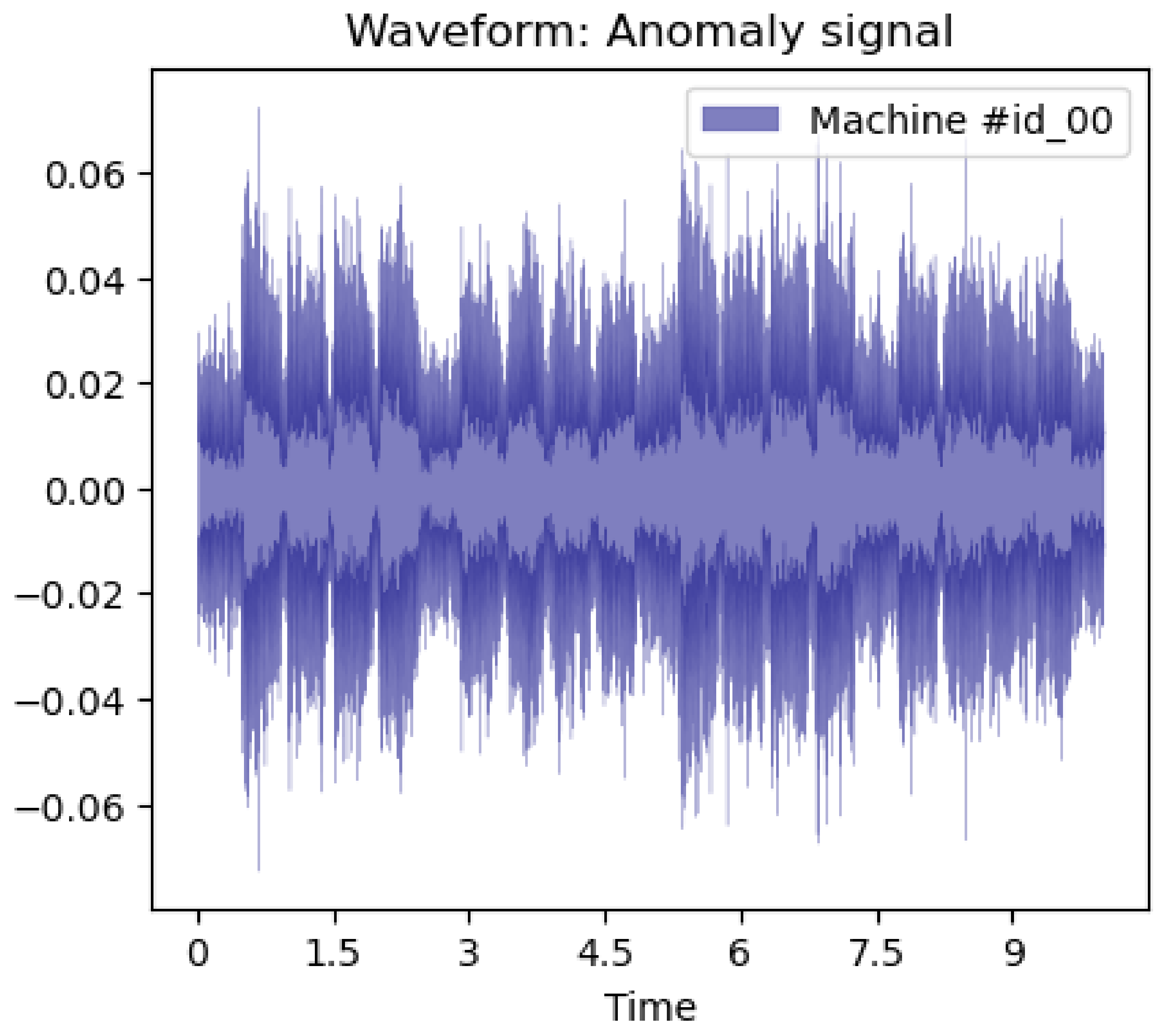


DATA EXPLORATION

WAVE FORM

The audio files represent time-series data, providing amplitude information of the sound throughout the duration.





DATA EXPLORATION

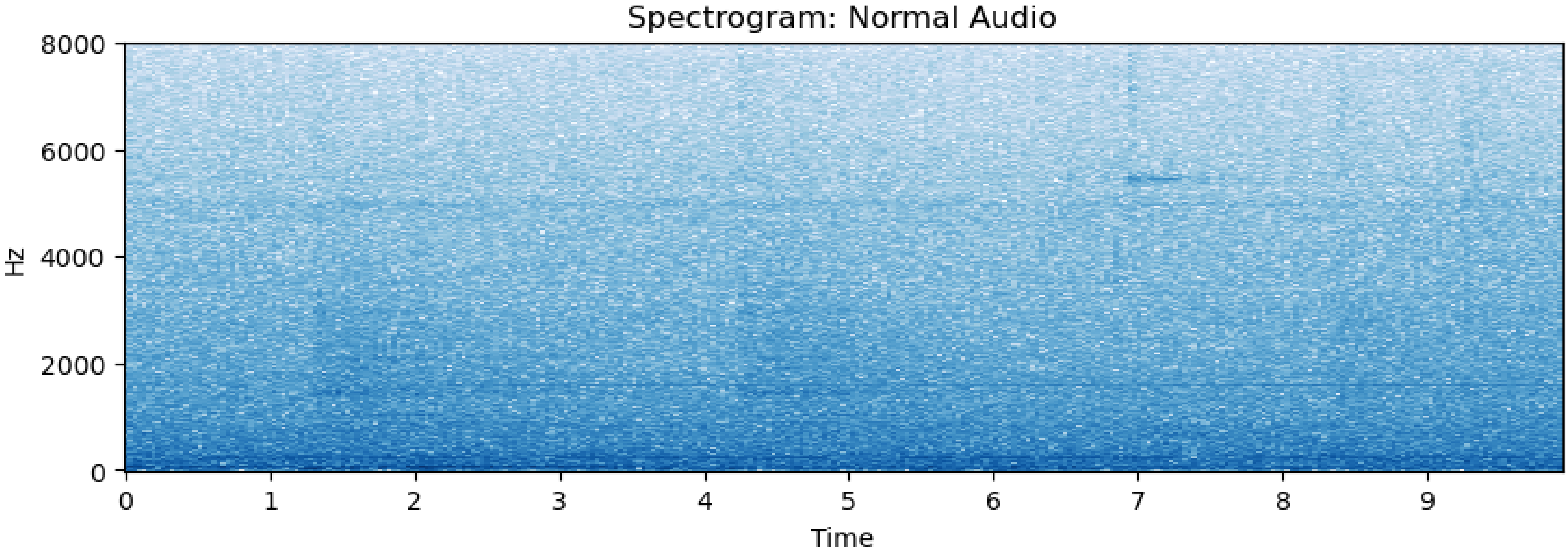
WAVE FORM

However, it is customary in deep learning models to transform the audio into a spectrogram, which can be described as:

- A concise representation of the audio waveform.
- Resembling an image, making it suitable for inputting into CNN-based architectures designed for image processing.

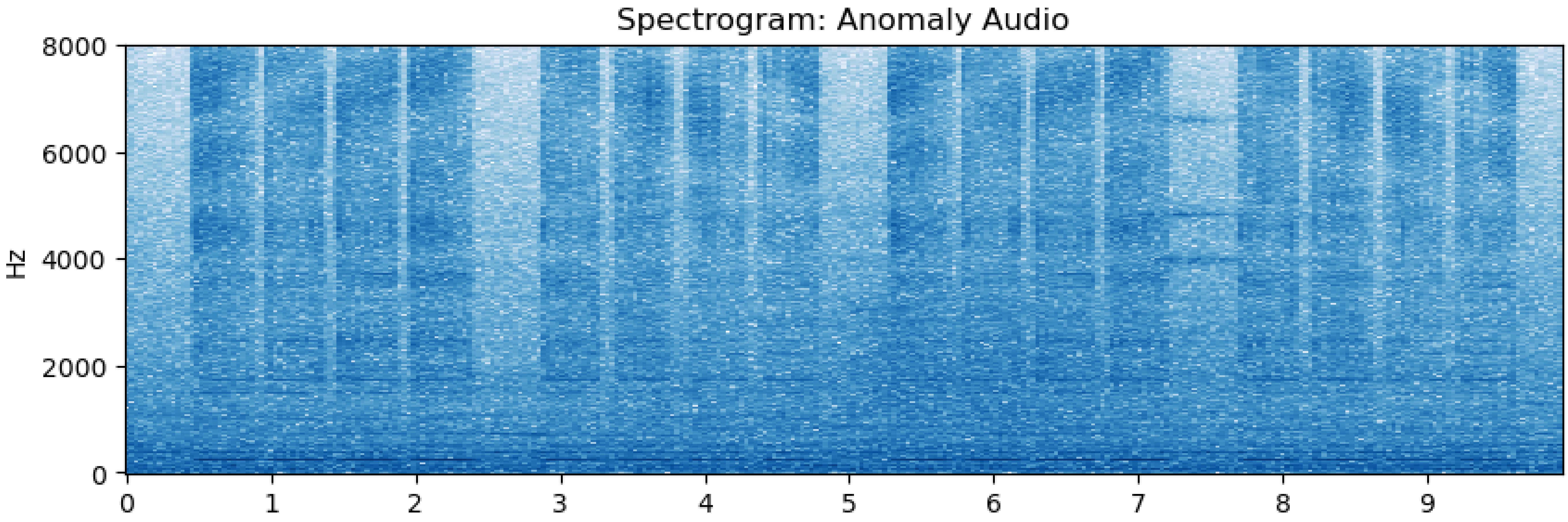
DATA EXPLORATION

SPECTROGRAM



DATA EXPLORATION

SPECTROGRAM



DATA EXPLORATION

CHROMAGRAM

- The Chromagram is an alternative method to examine the distinctions between the two types of signals.

- It focuses on perceiving the variations in pitch within an octave and the similarity of pitches that are separated by one or more full octaves.

- Pitch refers to the relative highness or lowness of a sound.

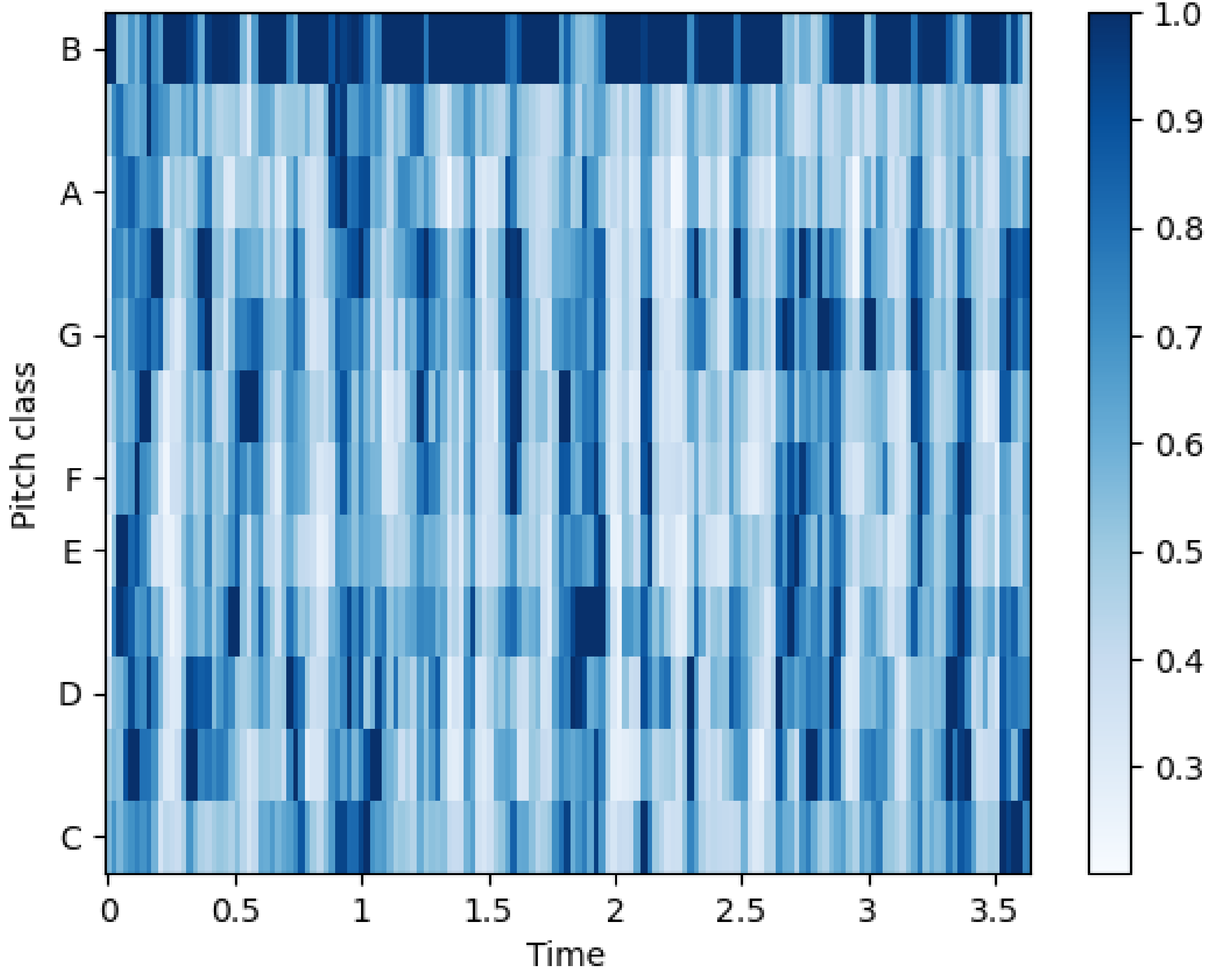
- A higher sound corresponds to a higher pitch, while a lower sound corresponds to a lower pitch.

- The Chromagram helps in capturing the perceived differences in pitch between different sounds and identifying the similarities between pitches that are octaves apart.

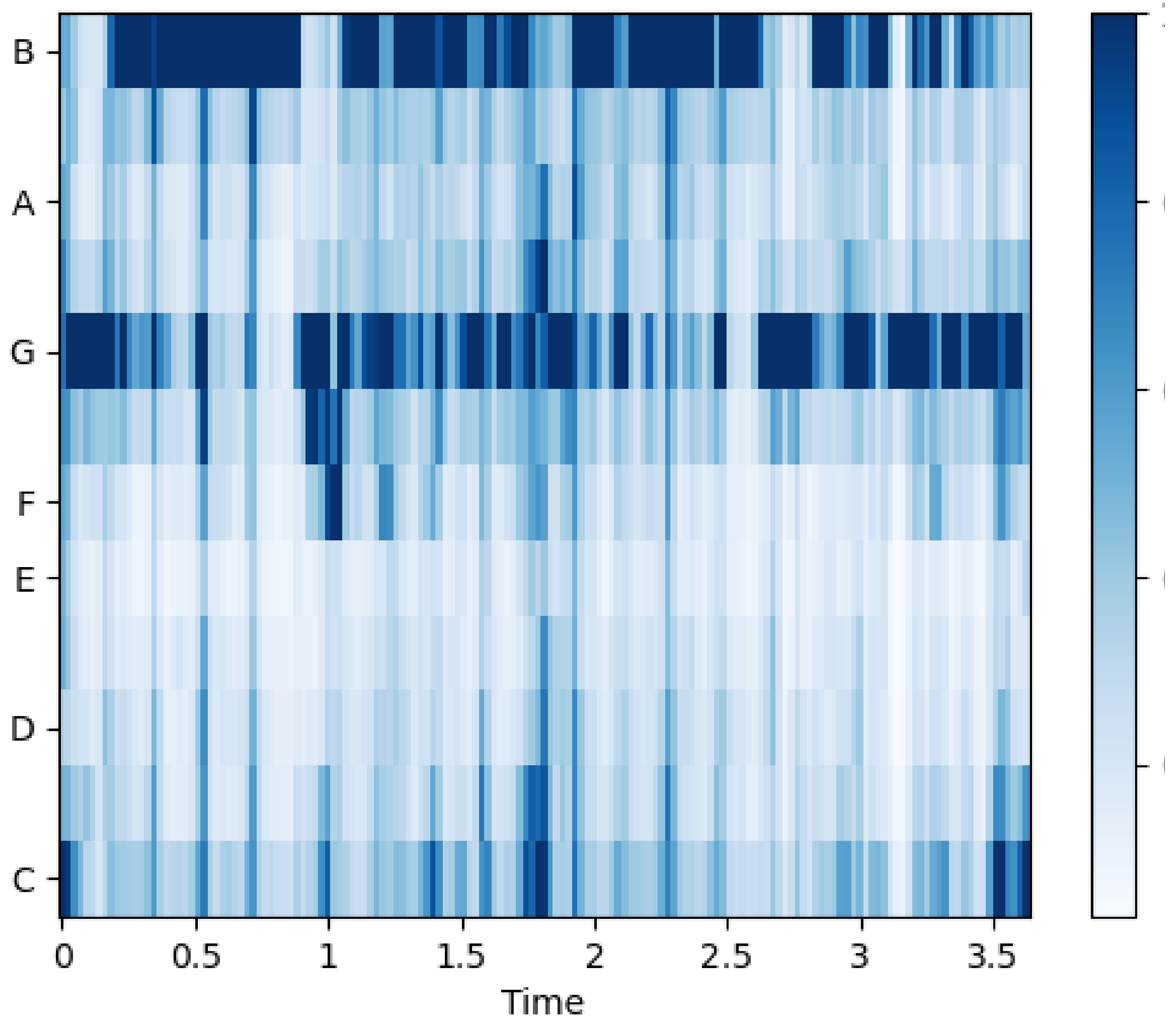
DATA EXPLORATION

CHROMAGRAM

Chromagram: Normal Signal



Chromagram: Anomaly Signal



DATA EXPLORATION

CHROMAGRAM

The chromagram analysis reveals that normal sounds encompass a wide range of notes, while the chromagrams of anomalous sounds exhibit a distinct emphasis primarily on notes **B and G**.

This observation reinforces the notion that anomalous sounds exhibit higher frequencies in their acoustic characteristics.

DATA EXPLORATION

MEL SPECTROGRAM

WHAT IS MEL SCALE?

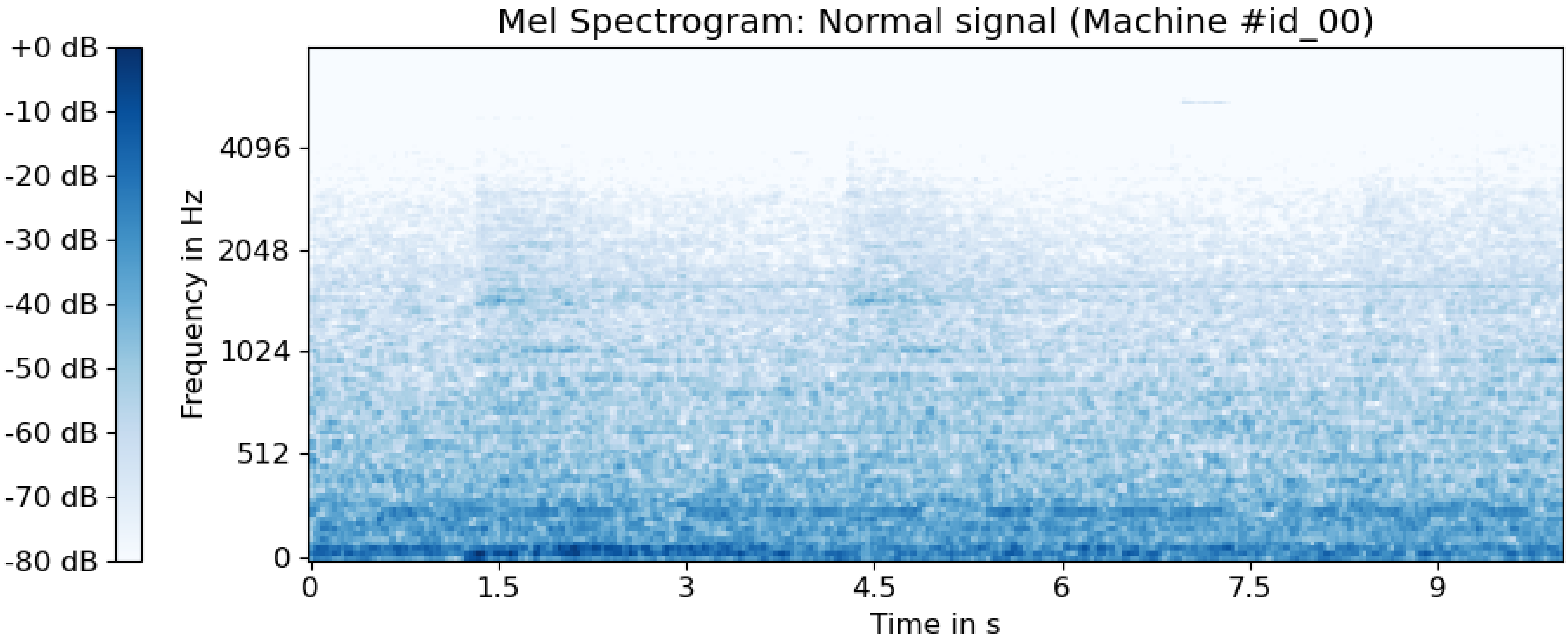
The Mel scale is a perceptual scale of pitches that approximates the human auditory system's response to different frequencies.

WHAT IS MEL SPECTROGRAM?

By mapping the linear frequency scale to the Mel scale, the Mel spectrogram provides a more meaningful and perceptually relevant representation of the audio signal's frequency content.
By dividing the frequency range into perceptually equal intervals, it captures important frequency components and patterns.

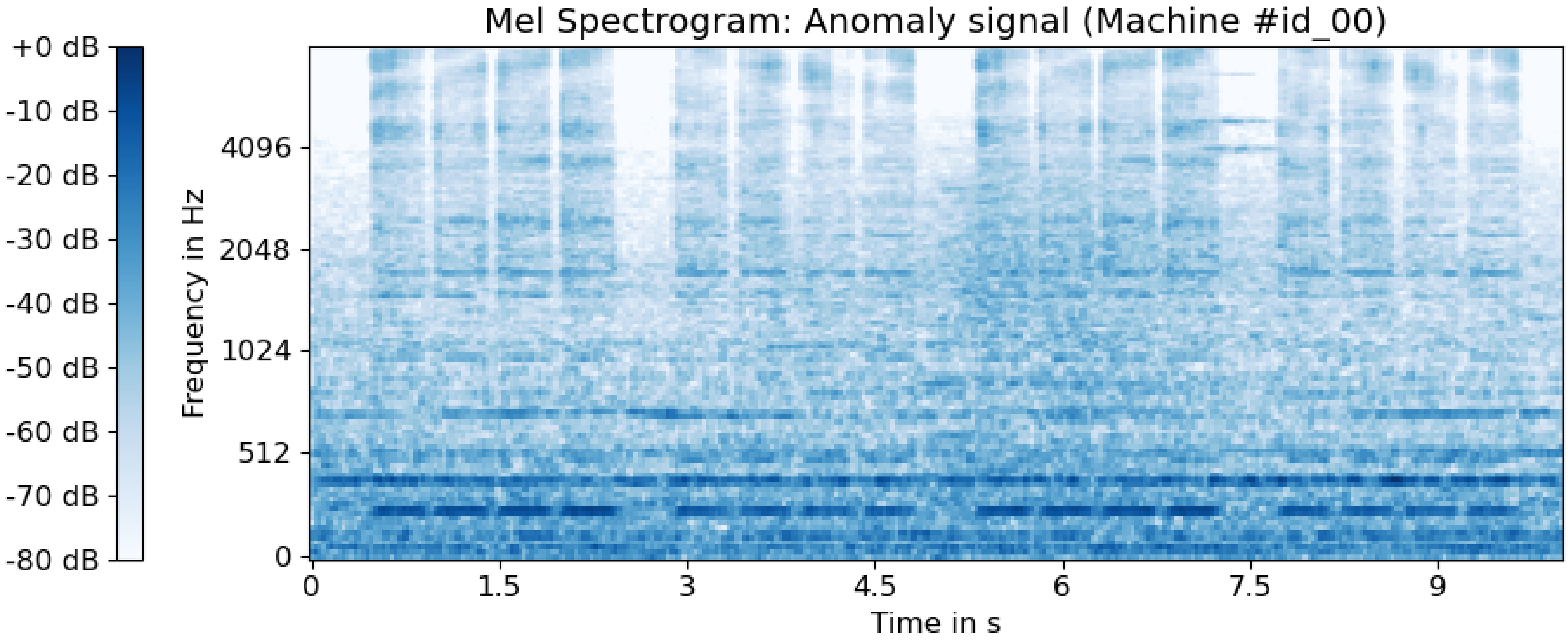
DATA EXPLORATION

MEL SPECTROGRAM



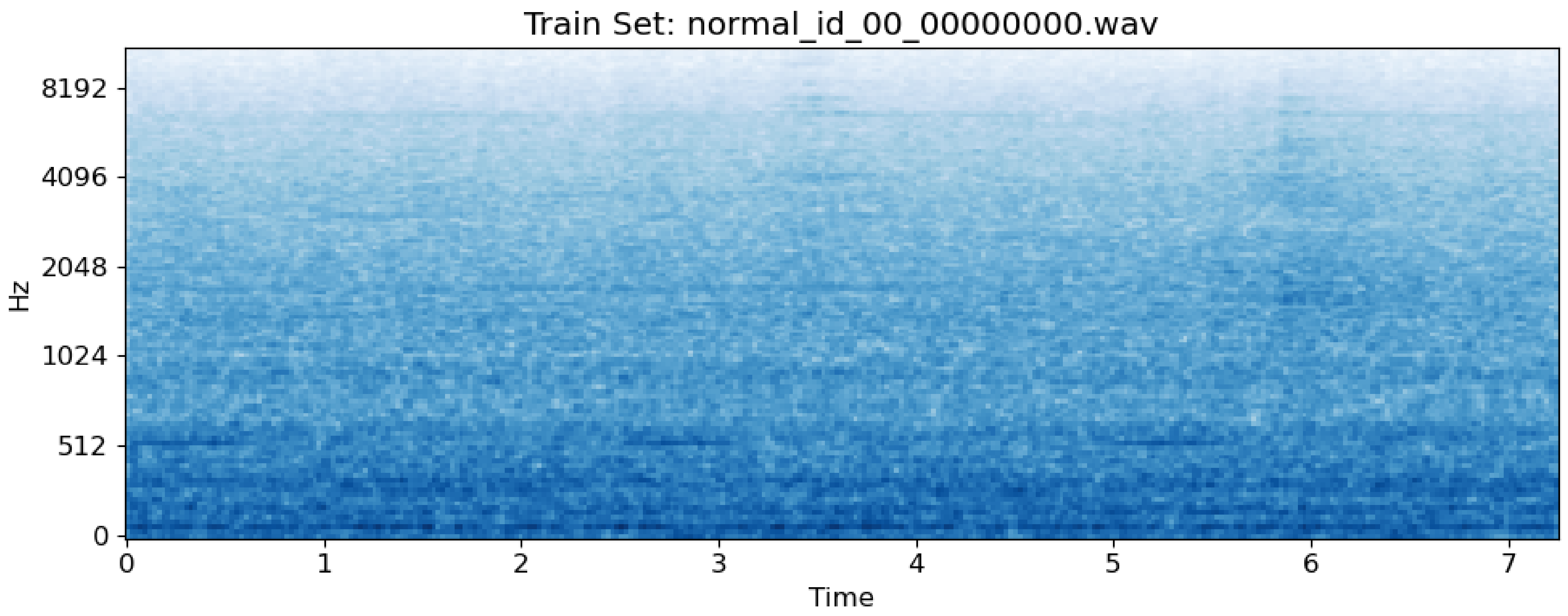
DATA EXPLORATION

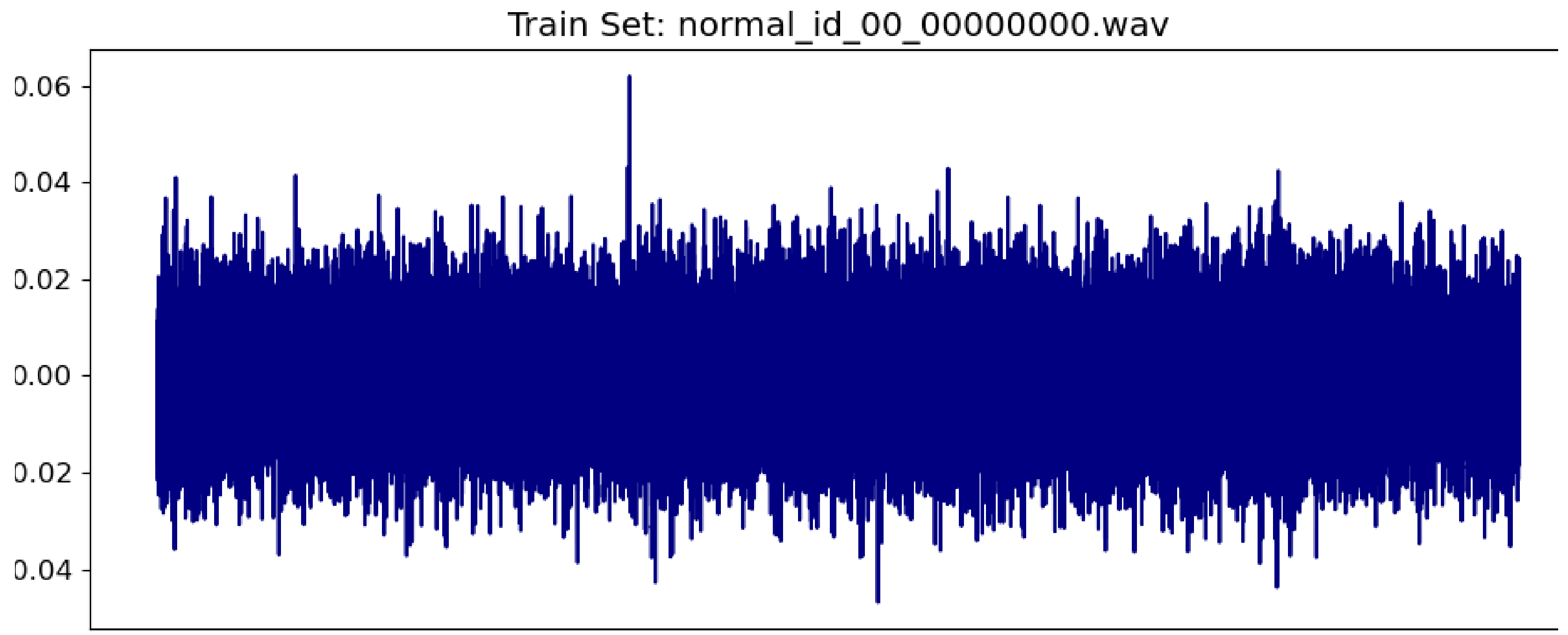
MEL SPECTROGRAM

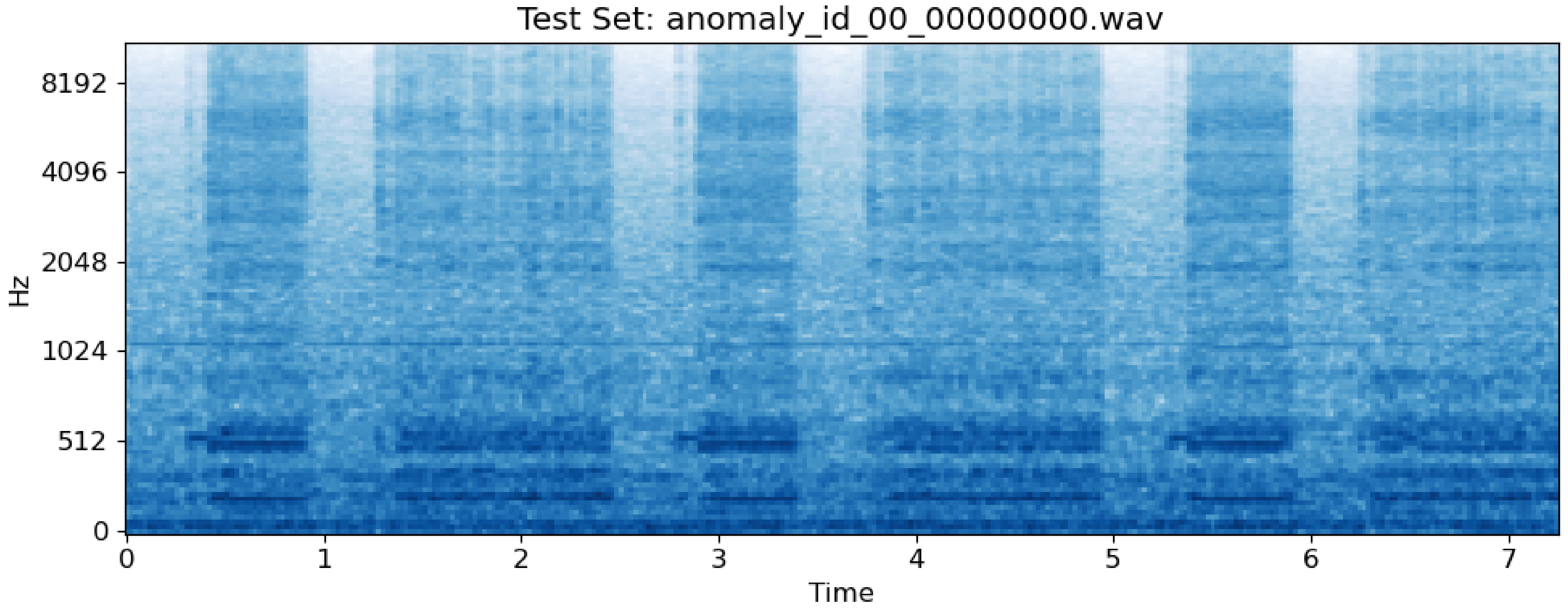


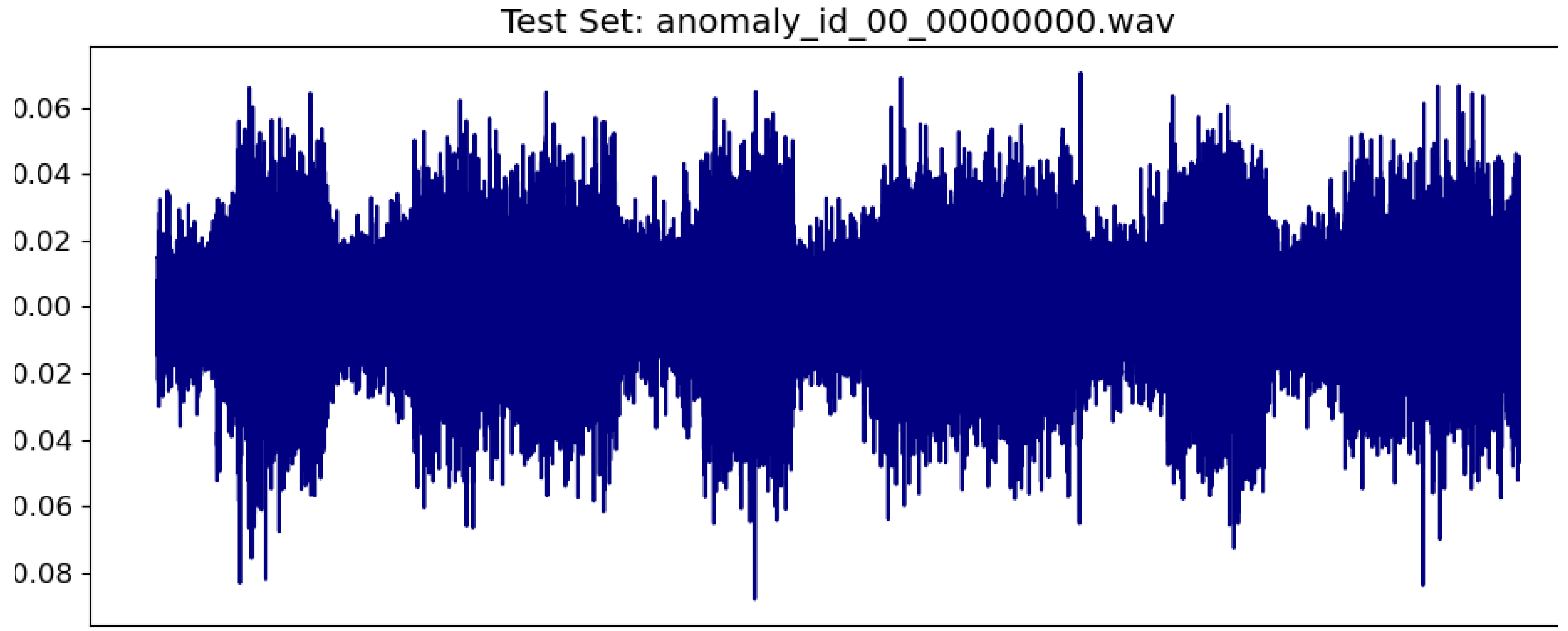
DATA EXPLORATION

TRAIN SET VS TEST SET


Train Set: normal_id_00_00000000.wav


Train Set: normal_id_00_00000000.wav


Test Set: anomaly_id_00_00000000.wav


Test Set: anomaly_id_00_00000000.wav

DATA PREPARATION

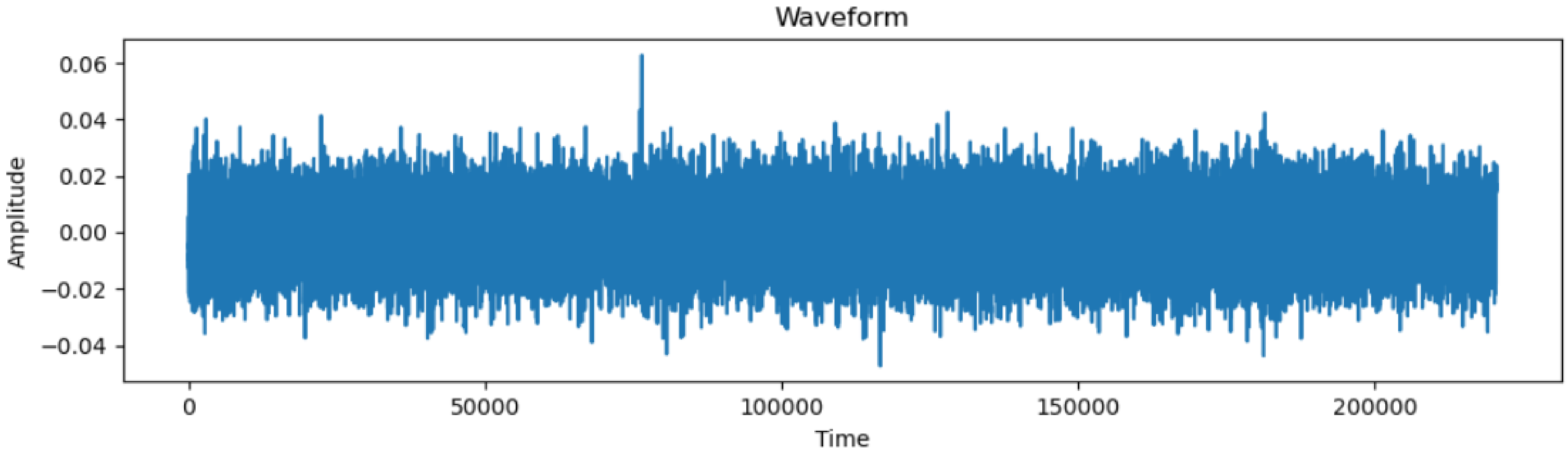
DATA PRE-PROCESSING FOR BOTH TRAINING AND EVALUATION DATA:

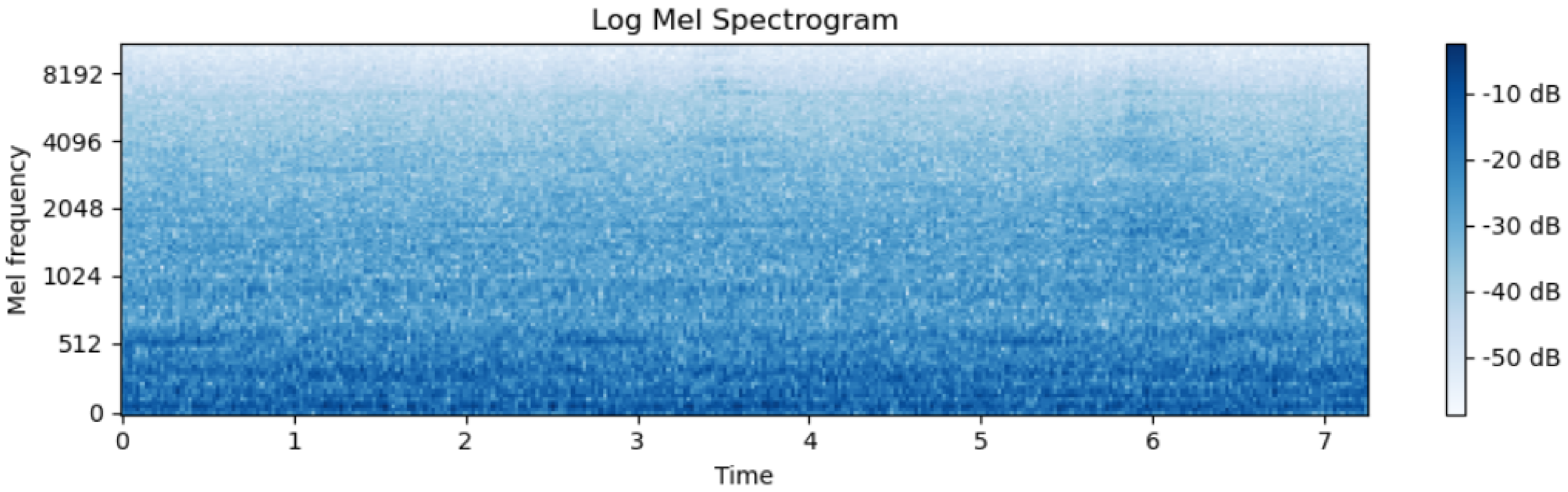
- We convert the sound files into log mel spectrogram data.
- The dataset includes only normal sound samples for training.
- For each machine type and split, the dataset is filtered, and the log mel spectrogram is calculated for the files.
- The spectrogram data is stored in an array.
- The array is then saved as a NumPy file.

DATA PREPARATION

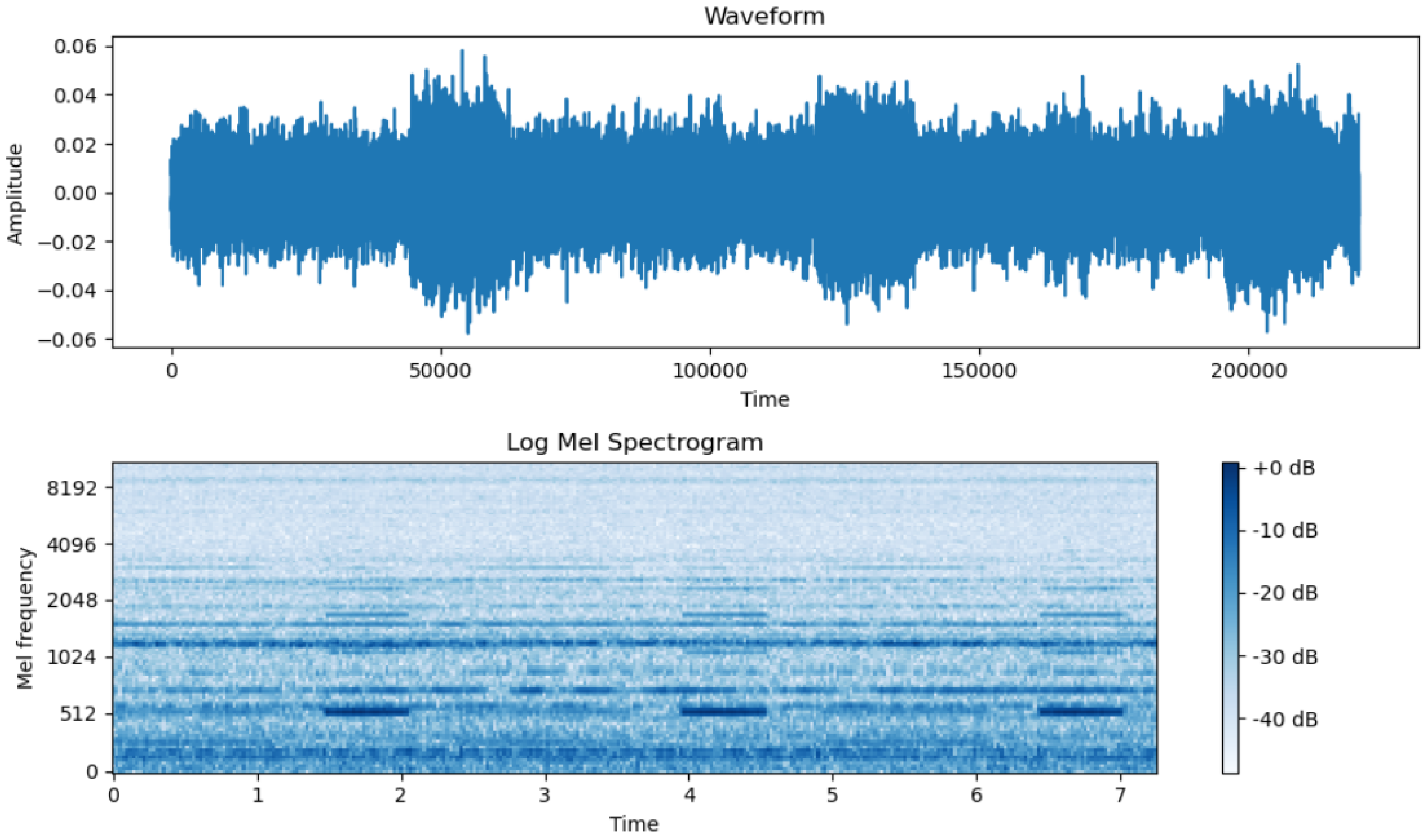
- Converting the sound files into log mel spectrogram data helps in distinguishing between normal and anomalous sounds emitted by the target machine.
- The main challenge in ASD is detecting unknown anomalous sounds when only normal sound samples are available for training.
- The data pre-processing step prepares the dataset for training a model capable of identifying unknown anomalous sounds not observed during training.

DATA PREPARATION





DATA PREPARATION



DATA PREPARATION

WHY LOG MEL SPECTROGRAMS PERFORM BETTER?

- The mel scale aligns frequency bins with human auditory perception.
- Log mel spectrograms improve accuracy in detecting unknown and diverse anomalies.
- Log mel spectrograms provide a compact representation by reducing dimensionality.
- They enable standardization and comparability across different audio samples.

MODEL SELECTION

ALTERNATIVE APPROACHES ATTEMPTED

Convolutional Neural Network:

- CNNs do not capture temporal dynamics and subtle variations in sound data very well.
- CNNs struggle to capture the global context and long-term dependencies required for accurate anomaly detection.
- Sound data is typically represented as sequential data or spectrograms. CNNs are not specifically designed to handle sequential data, and they may not be able to capture the temporal dependencies present in sound signals.
- CNNs lack the inherent ability to capture the temporal dependencies present in sound signals, limiting their effectiveness in anomaly detection for sound data.

MODEL SELECTION

ALTERNATIVE APPROACHES ATTEMPTED

Vanilla Auto-Encoders:

- Vanilla Auto-encoders can be sensitive to normal variations, leading to false positives or difficulty in distinguishing anomalies.
- Unsupervised training of autoencoders may not effectively leverage anomaly labels for explicit anomaly detection.
- Limited representation capacity of autoencoders may not capture the complexity and variability in sound data.
- The encoding and decoding process of autoencoders may not effectively capture the distinguishing features that separate normal and anomalous sound patterns.

MODEL SELECTION

ALTERNATIVE APPROACHES ATTEMPTED

Variational Auto-Encoders:

- The reconstruction loss in VAEs prioritize capturing normal variations, resulting in false positives in anomaly detection.
- The latent space dimensionality of VAEs is not sufficient to capture the variability and complexity of anomalous sound patterns.
- The limited representation capacity of VAEs do not capture the complexity and variability present in sound data very well, making it challenging to distinguish anomalies.

MODEL SELECTION

CONVOLUTIONAL AUTOENCODERS

WHAT ARE AUTO-ENCODERS?

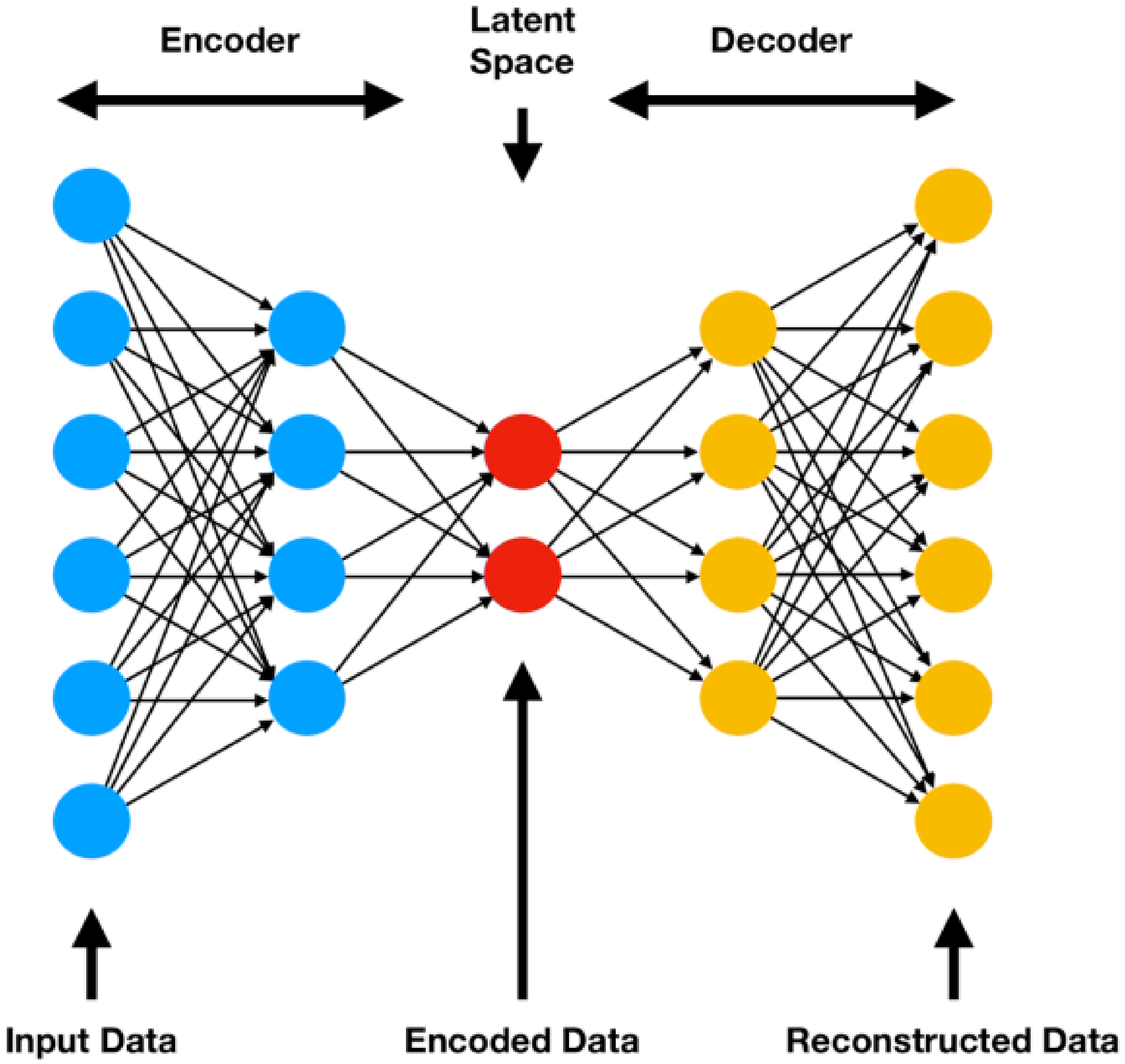
Autoencoders are neural network models that are trained to learn a compressed representation of the input data, often referred to as a latent space or bottleneck. They consist of an encoder network that maps the input data to a lower-dimensional representation and a decoder network that reconstructs the original input from this representation.

OBJECTIVE OF AUTO-ENCODERS:

The goal of autoencoders is to capture the most important features of the input data in the latent space and to reconstruct the input as accurately as possible, allowing for tasks like dimensionality reduction, denoising, and anomaly detection.

MODEL SELECTION

AUTOENCODERS



MODEL SELECTION

CONVOLUTIONAL AUTOENCODERS

WHAT ARE CONVOLUTIONAL AUTO-ENCODERS?

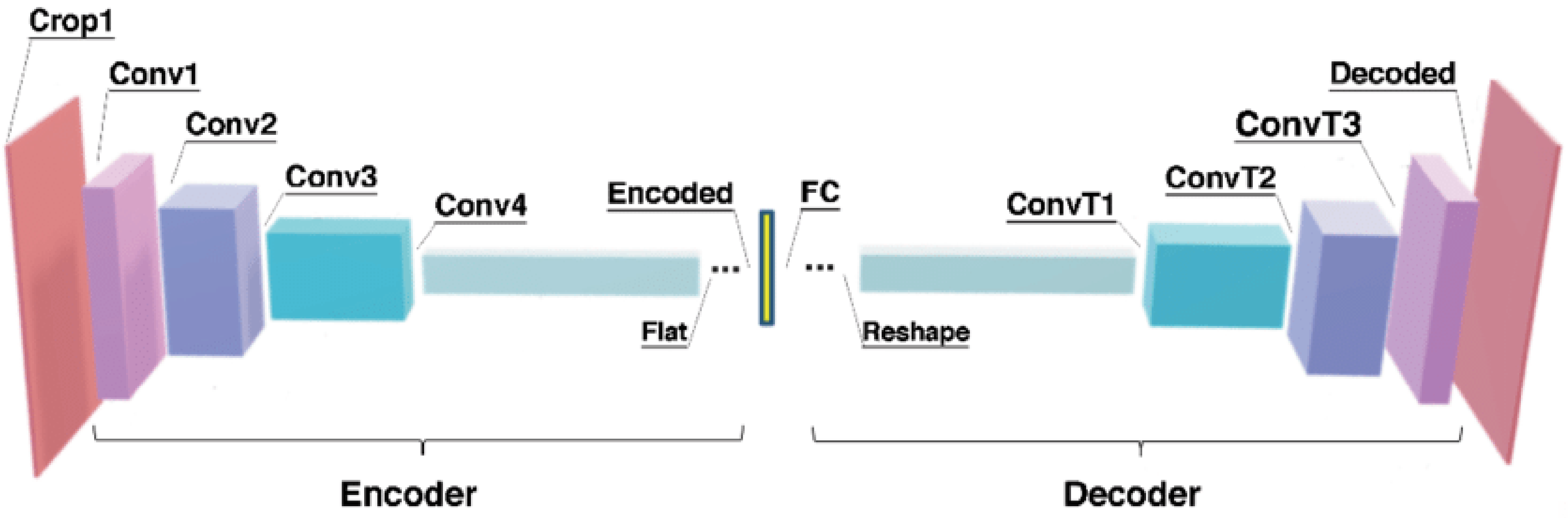
Convolutional Auto-Encoders extract hierarchical features from the input through convolutional layers in the encoder, compressing the data into a lower-dimensional latent space. The decoder then uses transposed convolutional layers to reconstruct the original input, aiming to minimize the reconstruction error during training.

WHY CONVOLUTIONAL AUTO-ENCODERS?

CNN autoencoders work well in anomaly detection because they capture the spatial structure and hierarchical features of the input data. By learning a compressed representation of normal data during training, they become sensitive to deviations from the learned patterns. When presented with anomalous data during inference, the reconstruction error between the input and the output of the autoencoder tends to be significantly higher, enabling effective identification and detection of anomalies.

MODEL SELECTION

CONVOLUTIONAL AUTOENCODERS



MODEL SELECTION

CONVOLUTIONAL AUTOENCODERS

PARAMETERS DEFINED FOR THE SPECTROGRAMS:

- **n_mels:** By specifying an appropriate value, we aim to capture a sufficient amount of frequency information in the input data.

- **n_fft:** By setting an appropriate value, we ensure that the spectrogram captures frequency details with the desired level of precision.

- **hop_length:** By adjusting this value, we control the temporal resolution of the spectrogram and the amount of overlap between adjacent frames.

By defining and fixing these parameters, we ensure that the model can extract and utilize as much relevant information as possible from the spectrograms. This allows the CNN Auto-Encoder to learn meaningful representations and perform effective anomaly detection based on the processed audio data.

MODEL SELECTION

CONVOLUTIONAL AUTOENCODERS

MODEL ARCHITECTURE:

- **Encoder:**
  - Conv2D
  - L2 regularization
  - MaxPooling2D
  - ReLU activation

- **Latent Space:** The compressed representation of the input obtained from the encoder.

- **Decoder**:
  - Conv2D
  - UpSampling2D
  - L2 regularization
  - ReLU activation

MODEL SELECTION

CONVOLUTIONAL AUTOENCODERS

HOW CONVOLUTIONAL AUTOENCODERS WORK?

- **ENCODER**
  - Convolutional layers apply filters to extract spatial features from the input data.
  - Pooling layers reduce the spatial dimensions by downsampling the features.
  - Activation functions introduce non-linearity to the encoder's output.

- **LATENT SPACE:**
  - The latent space is a compressed representation of the input data.
  - It is created by reducing the dimensions of the features in the bottleneck layer.
  - The latent space captures the most important and salient features of the input.

- **DECODER:**
  - Convolutional layers process the compressed features from the latent space.
  - Upsampling layers increase the spatial dimensions by interpolating the features.
  - Activation functions generate the output of the decoder.

MODEL SELECTION

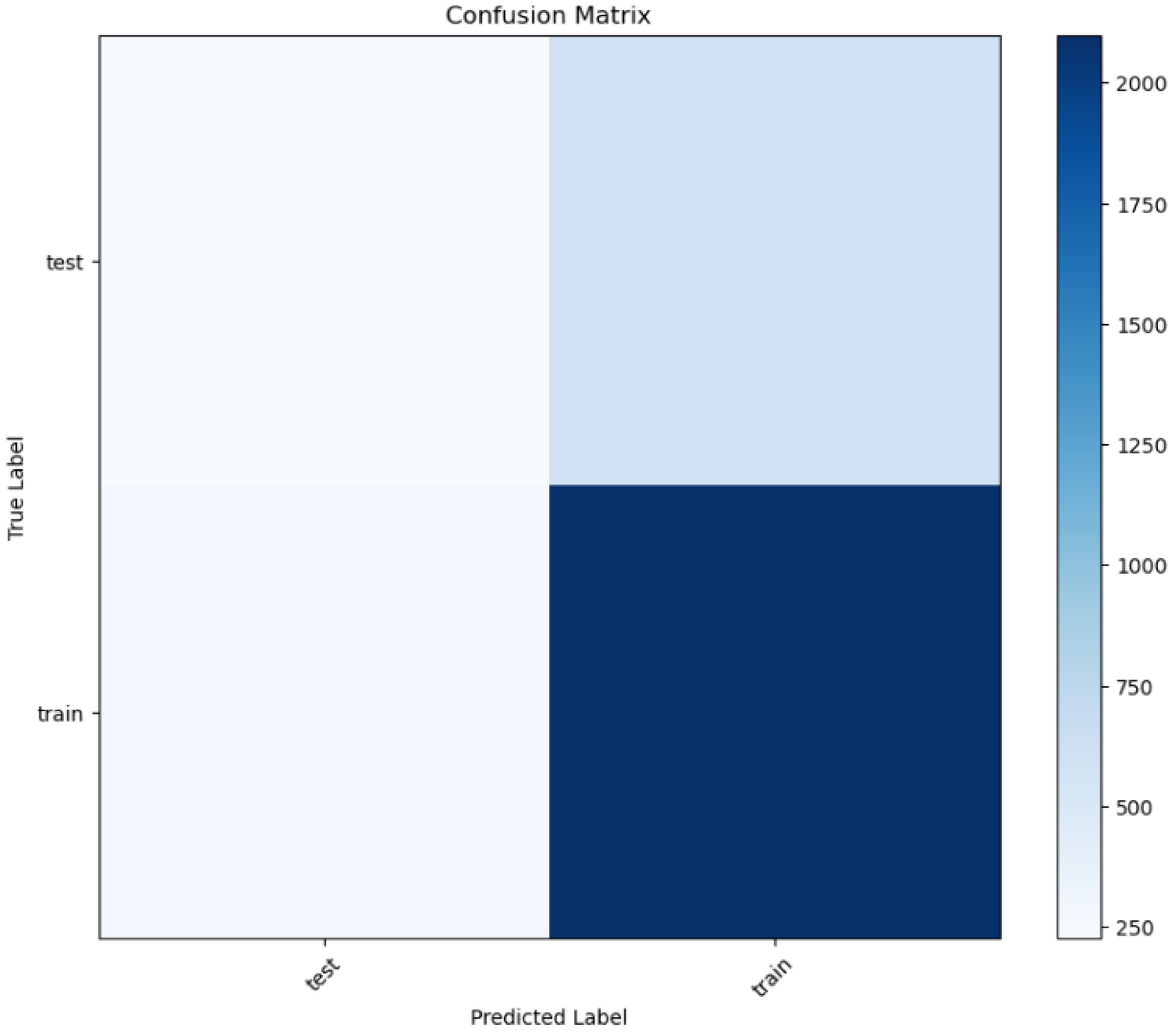
CONVOLUTIONAL AUTOENCODERS

WHY L2 REGULARIZATION?

- L2 regularization helps prevent overfitting in our model.
- It adds a penalty term to the loss function based on the L2 norm of the weights.
- The penalty discourages large weight values, promoting more general feature learning.
- It enhances the model's robustness in detecting unknown anomalous sounds.
- By controlling the complexity of learned features, L2 regularization improves performance in anomaly detection tasks.

MODEL PERFORMANCE

CONFUSION MATRIX



MODEL PERFORMANCE

PERFORMANCE METRICS

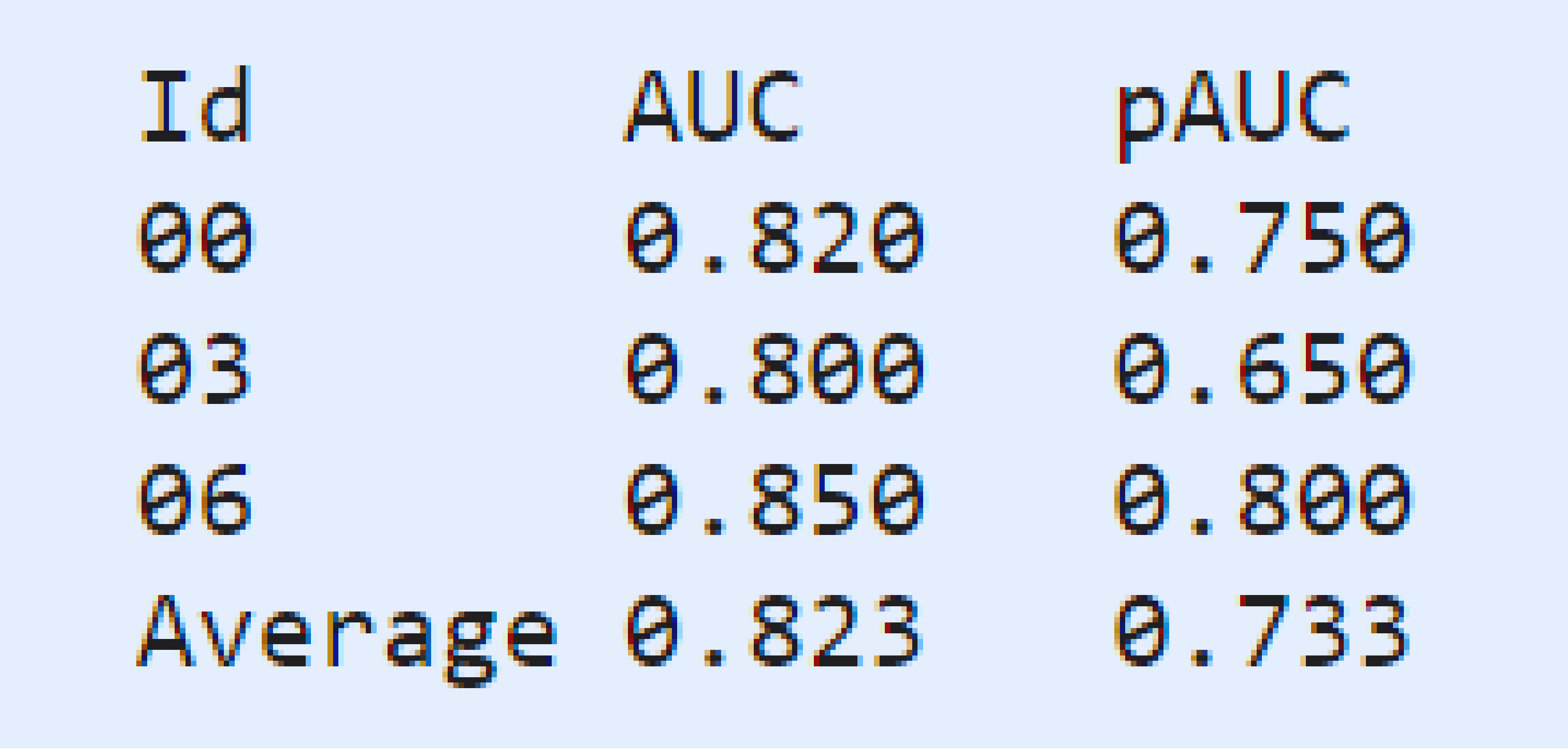
- **Precision** measures what proportion of predicted positive label is actually positive.
- **Recall** measures what proportion of actual positive label is correctly predicted as positive.
- **F1 score** is the harmonic mean of precision and recall, providing a balanced measure of a model's accuracy in binary classification tasks.

	PRECISION	RECALL	F1 SCORE
TRAIN	0.80	0.85	0.82
TEST	0.75	0.70	0.72

MODEL PERFORMANCE

- **AUC (Area Under the Curve):** AUC is a metric that measures the overall performance of a binary classification model by calculating the area under the Receiver Operating Characteristic (ROC) curve.

- **pAUC (Partial Area Under the Curve):** pAUC is a variant of AUC that focuses on a specific region of the ROC curve, typically a subset of false positive rates, to evaluate the model's performance within that region.



Id	AUC	pAUC
00	0.820	0.750
03	0.800	0.650
06	0.850	0.800
Average	0.823	0.733

CONCLUSION

- We explored different architectures like CNN, Vanilla Auto-Encoders and Variational Auto-Encoders for the task of anomalous sound detection.

- Among these architectures, the Convolutional autoencoders model with L2 regularization demonstrated superior performance.

- Our experimentation was constrained by limited computational resources, which prevented us from conducting extensive hyperparameter tuning.

- Further evaluation on diverse and larger datasets, including different types of machinery and environmental conditions, is necessary to assess the model's robustness and real-world applicability.

- The availability of only normal sound samples for training and the scarcity and diversity of actual anomalous sounds in real-world factories posed challenges in capturing and detecting unknown anomalous sounds effectively.

FUTURE WORK

- We could focus on integrating ensemble techniques such as model averaging or stacking, to combine multiple CNN AE models and improve the overall detection performance through ensemble predictions.
- Ensemble methods can leverage the diversity of different CNN AE models to capture a wider range of features and improve overall detection accuracy.
- Due to computational resource constraints, we were unable to explore hyperparameter tuning using techniques such as grid search. However, our future work might prioritize hyperparameter tuning as it can significantly impact the performance of the model and potentially improve the accuracy and effectiveness of the anomaly detection system.

REFERENCES

RESEARCH PAPERS:

- Wei, Shengyun, Shun Zou, and Feifan Liao. "A comparison on data augmentation methods based on deep learning for audio classification." In Journal of Physics: Conference Series, vol. 1453, no. 1, p. 012085. IOP Publishing, 2020.

- Koizumi, Yuma, Yohei Kawaguchi, Keisuke Imoto, Toshiki Nakamura, Yuki Nikaido, Ryo Tanabe, Harsh Purohit et al. "Description and discussion on DCASE2020 challenge task2: Unsupervised anomalous sound detection for machine condition monitoring." arXiv preprint arXiv:2006.05822 (2020).

- Cai, Xinyu, Heinrich Dinkel, Zhiyong Yan, Yongqing Wang, Junbo Zhang, Zhiyong Wu, and Yujun Wang. "A Contrastive Semi-Supervised Learning Framework For Anomaly Sound Detection." In DCASE, pp. 31-34. 2021.

- Kuroyanagi, Ibuki, Tomoki Hayashi, Yusuke Adachi, Takenori Yoshimura, Kazuya Takeda, and Tomoki Toda. "Anomalous sound detection with ensemble of autoencoder and binary classification approaches." In DCASE2021 Challenge. 2021.

REFERENCES

WEBSITES:

- https://towardsdatascience.com/audio-deep-learning-made-simple-part-1-state-of-the-art-techniques-da1d3dff2504
- https://importchris.medium.com/how-to-create-understand-mel-spectrograms-ff7634991056
- https://www.audiolabs-erlangen.de/resources/MIR/FMP/C3/C3S1_SpecLogFreq-Chromagram.html
- https://github.com/aldente0630/sound-anomaly-detection-with-autoencoders/tree/main
- https://www.kaggle.com/code/residentmario/autoencoders
- https://stackoverflow.com/questions/44787437/how-to-convert-a-wav-file-to-a-spectrogram-in-python3
- https://towardsdatascience.com/applied-deep-learning-part-3-autoencoders-1c083af4d798
- https://www.analyticsvidhya.com/blog/2021/06/complete-guide-on-how-to-use-autoencoders-in-python/
- https://github.com/JudeWells/keras_anomaly_detection

THE TEAM

SHUBHIKA GARG

Master's in Data Science and Engineering

YOSSI TAPIERO

Telecom SudParis

DORIAN LEBRU

Telecom SudParis

JORDAN ALLEMAND

Telecom SudParis