CHALMERS
UNIVERSITY OF TECHNOLOGY

Physics-Informed Two-Stage Learning Framework for Engine Ignition Frequency and RPM Estimation

Synchronized Sound−Vibration Product Sensing and Multi-Representation Feature Ranking under LOSO/LOVO Generalization

Author: Nuree Kim

Supervisor: Tomas Frödin

Examiner: Dr. Mohsen Mirkhalaf

Industrial Partner: EnviroClean Sweden AB

Master's Thesis in Complex Adaptive Systems
(MSc Programme −MPCAS)

Department of Computer Science and Engineering

Chalmers University of Technology

Gothenburg, Sweden, 2026

This thesis was carried out at Chalmers University of Technology in collaboration with EnviroClean Sweden AB.

The company provided synchronized sound and vibration datasets of multi-cylinder four-stroke engines for research and analysis within this project.

Department of Computer Science and Engineering
Chalmers University of Technology, Gothenburg, Sweden

Physics-Informed Two-Stage Learning Framework for Engine Ignition Frequency and RPM Estimation

Synchronized Sound−Vibration Product Sensing and Multi-Representation Feature Ranking under LOSO/LOVO Generalization

Author: Nuree Kim

Supervisor: Tomas Frödin

Examiner: Dr. Mohsen Mirkhalaf

Master's Thesis 2026:XX

Master's Programme in Complex Adaptive Systems

Department of Computer Science and Engineering

Chalmers University of Technology

SE−412 96 Gothenburg, Sweden

*This master's thesis was carried out at Chalmers University of Technology in collaboration with **EnviroClean Sweden AB**, which provided synchronized sound and vibration datasets from real multi-cylinder four-stroke engines for research and analysis.*

Printed by Chalmers Reproservice, Gothenburg, Sweden, 2026.

Email: nuree@student.chalmers.se

Cover image: Conceptual schematic of the synchronized sound–vibration–product pipeline for ignition rhythm analysis.

Abstract

Accurate estimation of the ignition frequency (f_0) of internal combustion engines is essential for non-invasive rotational speed (RPM) monitoring and condition diagnosis. In practice, domain shifts caused by sensor placement, vehicle-specific resonances, and environmental noise can distort the harmonic structure of f_0, leading to multiple plausible spectral peaks within short analysis windows. A physics-informed, two-stage machine-learning framework is developed to estimate the ignition frequency of four-stroke engines using synchronized sound and vibration measurements. A nonlinear product signal ($x_{\text{prod}}(t) = x_{\text{sound}}(t)\, x_{\text{vib}}(t)$) is introduced to emphasize ignition events jointly detected by both sensors, providing an additional representation for joint analysis of ignition-related components. Features are extracted from multiple signal representations, including the FFT, Envelope FFT, Cepstrum, Autocorrelation (ACF), and Envelope–ACF, and combined into a unified feature space capturing harmonic consistency, periodicity, and peak morphology.

The framework consists of two learning stages. Stage 1 classifies global engine characteristics, including the cylinder count and ignition-frequency class. Stage 2 ranks local frequency candidates using a LightGBM-based LambdaMART ranker to identify the ignition frequency f_0. Generalization performance is evaluated using Leave-One-Source-Out (LOSO) and Leave-One-Vehicle-Out (LOVO) protocols.

Results show a mean LOSO accuracy of 95.5% for cylinder classification and a Top-1 accuracy of 84% within a ± 5 Hz tolerance for candidate ranking, with mean frequency errors below 2 Hz ($\approx$60 rpm). The model demonstrates consistent generalization behavior on unseen vehicles, indicating robustness to domain shifts across sensors and operating conditions.

The proposed approach therefore provides an interpretable and practically applicable solution for non-invasive RPM estimation and establishes a foundation for real-time diagnostic applications on embedded automotive systems.

Keywords: ignition frequency, RPM estimation, multi-sensor fusion, harmonic analysis, machine learning, LambdaMART.

Statement on the Use of AI Tools

AI-based tools, including ChatGPT, were used during the preparation of this thesis to support language refinement, clarity, and restructuring of the manuscript, solely as a writing aid. All scientific content, signal processing methods, machine learning models, experimental design, analyses, and interpretations were developed, implemented, and validated by the author. The author takes full responsibility for the correctness, originality, and integrity of the presented work. All reported results, performance claims, and generalization statements are limited to the evaluated datasets and operating conditions described in this thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The ignition rhythm of an internal combustion engine is a periodic signal determined by the number of cylinders (N_{cyl}) and the rotational speed (RPM). For four-stroke engines, the ignition frequency f_0 is defined as

$$f_0 = \frac{\text{RPM} \times N_{\text{cyl}}}{120}. \tag{1.1}$$

This fundamental frequency and its integer harmonics can be observed in both airborne acoustic signals and structural vibration signals. Accurate estimation of f_0 is essential not only for non-invasive RPM monitoring based on ignition-frequency analysis, but also for assessing engine condition in data-driven diagnostic applications.

In practical measurement environments, however, the ignition rhythm rarely appears as an ideal harmonic pattern. Sensor placement, vehicle-specific resonance characteristics, environmental noise, and hardware variability distort the spectral components around f_0. Even under near steady-state operating conditions, small throttle variations and RPM fluctuations can cause f_0 to drift within a few hertz. As a result, within short analysis windows such as the 0.25 s segments used in this study, multiple plausible spectral peaks may appear simultaneously. Under these conditions, traditional approaches based on single-sensor or single-representation analysis—such as direct FFT peak selection, envelope-based methods, or cepstrum-based periodicity detection—often exhibit limited robustness when applied to new vehicles or previously unseen measurement sessions [1, 2].

To address these challenges, this study synchronizes sound and vibration signals in time and introduces a nonlinear product signal obtained by sample-wise multiplication of the two signals. This product signal emphasizes ignition events that are consistently observed by both sensors. Subsequently, frequency-domain and periodicity-related features are extracted using

FFT, Envelope FFT, Cepstrum, Autocorrelation (ACF), and Envelope–ACF representations. Each representation captures a different aspect of the ignition process, and their combination forms a physically interpretable multi-channel, multi-representation feature space that jointly describes the harmonic structure and temporal characteristics of ignition events.

1.1 Problem Definition

This thesis focuses on the following key challenges:

- **Domain shift across sessions and vehicles**: Feature distributions vary with sensor placement, vehicle structure, and recording conditions, making generalization difficult.
- **Ambiguity in identifying the true** f_0: In short signal segments, multiple strong candidates may coexist due to subharmonics, superharmonics, and resonance effects, reducing the reliability of simple peak-selection strategies.
- **Limited exploitation of physical structure**: Many existing methods rely on a single signal representation and fail to systematically integrate harmonic consistency, periodicity, and peak-shape information into machine-learning models.

1.2 Limitations in Existing Research

A review of existing studies reveals several major limitations:

1. **Insufficient generalization evaluation**: In many studies, training and test splits are not strictly separated by source or vehicle. When data from the same vehicle or measurement environment appear in both sets, shared characteristics can inflate performance and obscure true generalization ability [3].
2. **Lack of candidate-frequency ranking**: Most approaches assume the presence of a single dominant spectral peak corresponding to the ignition frequency and rely on direct peak selection or tracking. In realistic conditions, however, multiple f_0 candidates may coexist within a short analysis window, and explicit ranking of these candidates remains relatively underexplored [1, 2].
3. **Incomplete integration of multiple signal representations**: Frequency-domain, envelope-based, cepstrum, and autocorrelation representations are widely used to analyze periodicity and harmonic structure in engine signals. However, these representations are typically applied independently or combined only at a late decision stage. Few

studies integrate them at the candidate-frequency level in a physically interpretable manner [4–6].

1.3 Research Objectives

To address these limitations, the objectives of this thesis are as follows:

1. Construct a **physics-informed multi-representation feature space** using sound, vibration, and product signals.
2. Develop a **Stage 1 classifier** that predicts global segment-level attributes, including cylinder count and ignition-frequency class.
3. Design a **Stage 2 ranking model** that evaluates the relative plausibility of competing frequency candidates to identify the true f_0.
4. Apply **LOSO and LOVO** validation protocols to evaluate generalization across sensor configurations, measurement sessions, and vehicles.

1.4 Contributions

The main contributions of this work are summarized as follows:

- A unified framework that combines **harmonic consistency, periodicity measures, and peak-quality indicators** extracted from FFT, Envelope FFT, Cepstrum, ACF, and Envelope–ACF representations.
- A **two-stage learning architecture** that separates global engine-structure classification from local candidate-frequency ranking, enabling stable identification of f_0 in multi-peak environments.
- Introduction of a **product signal** that complements single-sensor representations by capturing ignition responses common to both sound and vibration sensors.
- A dual generalization evaluation protocol based on LOSO (source-level separation) and LOVO (vehicle-level separation), demonstrating robust performance under domain shifts caused by sensor, session, and vehicle variations.

1.5 Thesis Structure

The remainder of this thesis is organized as follows. Chapter 2 reviews related work on signal processing, multi-sensor fusion, and machine-learning approaches for ignition-rhythm estimation. Chapter 3 describes the physical background of engine ignition rhythms and the signal representations used in this study. Chapter 4 presents the proposed methodology, including preprocessing, feature extraction, model design, and evaluation procedures. Chapter 5 reports experimental results under LOSO and LOVO conditions. Chapter 6 discusses the results, their physical interpretation, and the strengths and limitations of the proposed approach. Finally, the concluding chapter summarizes the main findings and outlines directions for future research.

Chapter 2

Related Work

This chapter reviews representative signal-processing techniques, multi-sensor fusion strategies, and machine-learning-based approaches for estimating engine ignition rhythm (f_0) and rotational speed (RPM). The discussion highlights key limitations of existing methods and positions the present study within the broader research landscape.

2.1 Signal-Processing Methods for f_0/RPM

This section reviews signal-processing methods that have been widely used to estimate ignition frequency (f_0) and rotational speed (RPM) from acoustic and vibration measurements. These approaches generally aim to detect periodic structures or harmonic patterns associated with engine operation. While they offer physical interpretability and computational efficiency, their reliability can degrade in real measurement environments due to resonance effects, noise, and variations in operating conditions.

Envelope Analysis (Hilbert Transform).

Envelope analysis is a well-established technique for extracting periodic impulsive components in rotating machinery. The signal is typically band-pass filtered to isolate a resonance band, after which the analytic signal is obtained via the Hilbert transform. The spectrum of the resulting envelope is then analyzed to identify characteristic frequencies or rotational orders. This approach is widely used in bearing and gear diagnostics [5, 7], and has also been applied to engine acoustic and vibration signals to emphasize combustion-related periodicity. However, in practical environments, the choice of resonance band strongly influences the results, and noise, subharmonics, and sensor-dependent resonances can distort the envelope spectrum. As a result, multiple plausible frequency candidates may emerge, leading to

reduced robustness when measurement conditions change.

Cepstrum and Cyclostationary Analysis.

Cepstrum analysis transforms spectral periodicity into peaks in the quefrency domain, thereby revealing harmonic spacing, and has been used for periodicity estimation and inference of combustion- or pressure-related information in engine signals [8]. Closely related cyclostationary approaches model engine vibration as an angle-synchronous process, providing a structured representation of order components [4]. These methods can exploit latent periodicity even under strong resonance or noisy conditions. However, their performance is highly sensitive to parameter selection, and ambiguity can arise when multiple harmonic structures coexist, which limits generalization across vehicles, sensor placements, and operating conditions.

Autocorrelation and Cross-correlation.

Autocorrelation (ACF) is a classical time-domain technique for detecting periodicity in measured signals. Cross-correlation between acoustic and vibration signals has also been used for non-contact RPM estimation by exploiting periodic components shared across different sensors [9]. In realistic environments with broadband noise and structural resonance, however, correlation peaks can become smeared or ambiguous. This often results in multiple periodicity candidates, making it difficult to reliably distinguish the true ignition frequency f_0 from subharmonic or superharmonic components.

Order Tracking and Tacholess Speed Estimation.

Tacholess order-tracking techniques aim to estimate rotational speed and track harmonic orders without physical rotation sensors. These methods are implemented using time–frequency analysis, singular value decomposition, or signal reconstruction techniques such as the Vold–Kalman filter [10]. They are particularly effective under variable-speed conditions, including acceleration and deceleration. However, their algorithmic complexity is high, they are sensitive to hyperparameter settings, and their robustness can degrade in the presence of non-stationary resonance or sensor-dependent variability.

2.2 Multi-sensor Fusion (Sound & Vibration)

Multi-sensor fusion has been studied in engine and diesel systems to improve fault detection performance and robustness under varying operating conditions [11]. By combining complementary sensor information, these approaches aim to mitigate the limitations of single-sensor

methods. More recently, fusion of acoustic and vibration signals has been explored for non-contact speed estimation and condition monitoring [12].

In general, multi-sensor fusion can provide improved robustness compared to single-sensor approaches. However, when domain shifts arise due to differences in sensor placement, mounting conditions, hardware characteristics, or signal synchronization, reliable generalization is not guaranteed. These factors can alter both spectral and temporal characteristics of the measured signals, even when the underlying engine operating state remains unchanged.

In this study, a product-based representation (sound$\times$vibration) is defined by sample-wise multiplication of synchronized acoustic and vibration signals. This representation emphasizes signal components that are common across different sensors. By exploiting sample-level interactions between synchronized sensors, the product signal effectively captures jointly occurring components across acoustic and vibration measurements.

2.3 Machine Learning for Engine/Vehicle Acoustics

Supervised machine-learning approaches have been extensively studied for vehicle and engine analysis using acoustic signals. Existing work includes speed estimation using a single microphone or smartphone sensors, mel-spectrogram-based vehicle sound detection, and direct regression or classification of RPM from engine sounds. Several studies have reported competitive performance of mel-spectrogram features for single-channel speed estimation [13, 14], and acoustic vector sensors have been employed to improve robustness in roadside measurement scenarios [15].

Related studies have also investigated acoustic-based speed-change analysis in pass-by scenarios [16] and supervised models for direct RPM estimation from engine sounds [17]. However, most existing approaches are optimized for driving or roadside measurement environments, where Doppler effects, distance variation, and traffic flow dominate the observed signals.

As a result, systematic evaluation of precise ignition-frequency (f_0) discrimination under stationary or quasi-steady operating conditions remains limited in the existing literature. Moreover, many studies do not explicitly validate generalization across vehicles and sensor configurations using strict protocols such as Leave-One-Source-Out (LOSO) or vehicle-level separation. This gap motivates the need for learning frameworks that can achieve both fine-grained frequency discrimination and robust cross-domain generalization.

2.4 Learning to Rank for Peak/Order Selection

Selecting the true ignition frequency (f_0) from multiple competing candidates can naturally be formulated as a ranking problem, in which candidate frequencies within a segment are ordered according to their likelihood of corresponding to the physical ignition process. This perspective aligns with learning-to-rank problems in information retrieval, which focus on relative relevance rather than absolute classification or regression.

LambdaMART is a gradient-boosted decision-tree framework designed for learning-to-rank tasks, directly optimizing list-wise ranking metrics such as normalized discounted cumulative gain (NDCG). Because such metrics are not differentiable, LambdaMART employs surrogate gradients, referred to as "lambdas," to guide optimization [18, 19]. This approach has demonstrated high accuracy and practical efficiency across a wide range of ranking applications.

In the field of rotating machinery and engine analysis, studies that explicitly integrate physics-based signal characteristics into learning-to-rank frameworks for ignition-frequency estimation remain relatively limited. Furthermore, existing work often lacks rigorous evaluation of robustness to sensor-domain shifts or systematic cross-domain validation. In this study, physics-informed features such as harmonic lattice consistency, periodicity measures, and peak-shape descriptors are incorporated as ranking inputs to address these limitations.

2.5 Summary of Limitations and Positioning of This Work

Existing approaches for estimating engine ignition frequency and RPM provide physically interpretable solutions, but each category exhibits inherent limitations under realistic operating conditions. Traditional signal-processing methods are sensitive to resonance effects, noise, and parameter selection, often leading to ambiguity among frequency candidates. Multi-sensor fusion can improve robustness, but reliable generalization is difficult to ensure in the presence of sensor-domain shifts. Machine-learning-based acoustic approaches are largely optimized for driving or pass-by scenarios, and often lack precise f_0 discrimination under stationary conditions as well as rigorous cross-domain evaluation. These limitations motivate the development of a physics-informed, multi-representation learning framework explicitly designed to resolve frequency ambiguity and to generalize across vehicles and sensor configurations.

Chapter 3

Background

This chapter explains the physical principles underlying engine ignition rhythm, reviews signal-processing techniques used to extract it, and introduces the theoretical foundations of the machine-learning approaches built upon these representations. In particular, this study proposes a *physics-informed two-stage learning* framework that combines physically interpretable features with learning-based inference, and summarizes the mathematical and statistical validity of each stage.

3.1 Engine Dynamics and Ignition Rhythm

In a four-stroke internal combustion engine, each cylinder fires once for every two revolutions (720°) of the crankshaft. Accordingly, the engine ignition frequency is defined as

$$f_0 = \frac{\text{RPM} \times N_{\text{cyl}}}{120}, \tag{3.1}$$

where RPM denotes the number of crankshaft revolutions per minute, and N_{cyl} denotes the number of cylinders.

Taking the ignition frequency (f_{ign}) as the fundamental frequency (f_0), integer-multiple components at $kf_0\,(k \in \mathbb{N})$ appear as harmonics in the frequency domain, and their collection constitutes the engine's ignition rhythm. This harmonic structure directly reflects the number and spacing of ignition events determined by N_{cyl}, and therefore exhibits physically fixed characteristics governed by engine configuration and rotational speed. In other words, the ignition rhythm corresponds directly to the periodic rotation of the crankshaft and the sequence of ignition events defined by the cylinder count.

In real measurement environments, slight RPM fluctuations, load variations, and the influ-

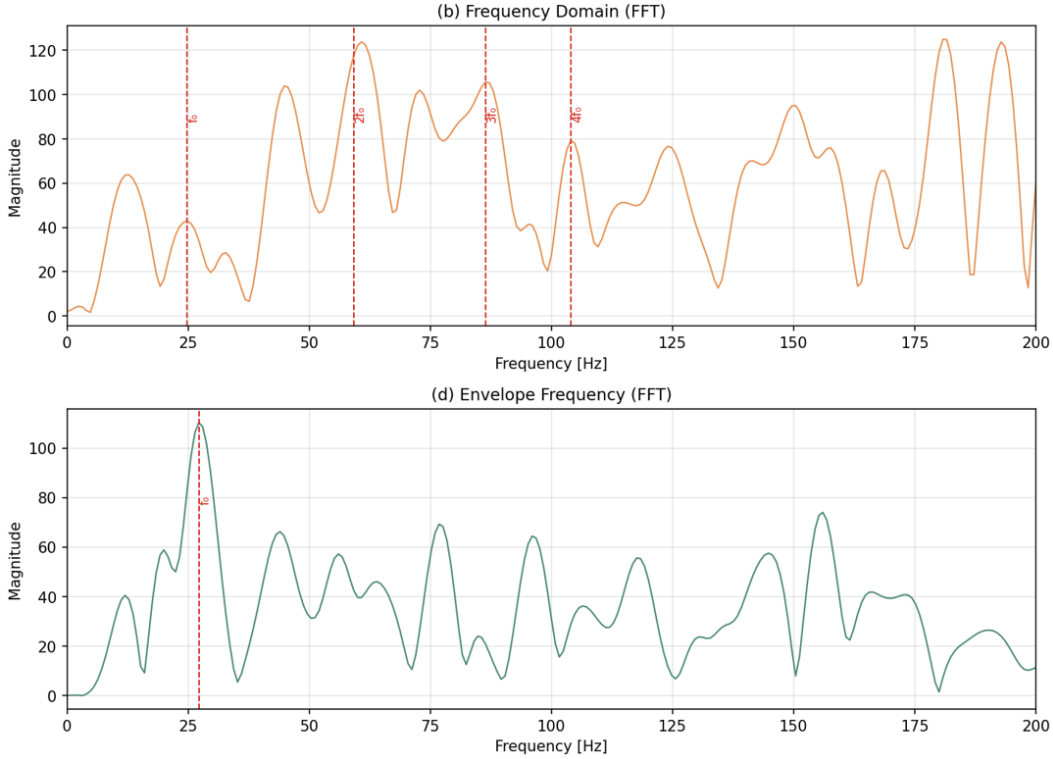
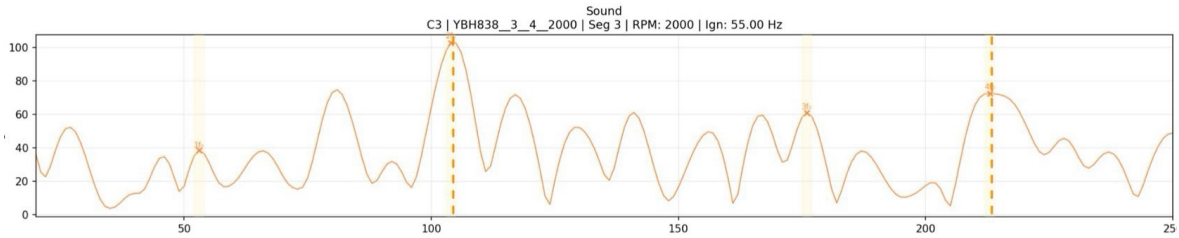


Figure 3.1: Comparison of frequency-domain representations under identical engine ignition conditions. The top panel shows the FFT of the raw signal, and the bottom panel shows the FFT of the envelope signal. Vertical dashed lines indicate the theoretical locations of the ignition frequency (f_0) and its integer multiples ($2f_0$, $3f_0$, $4f_0$).

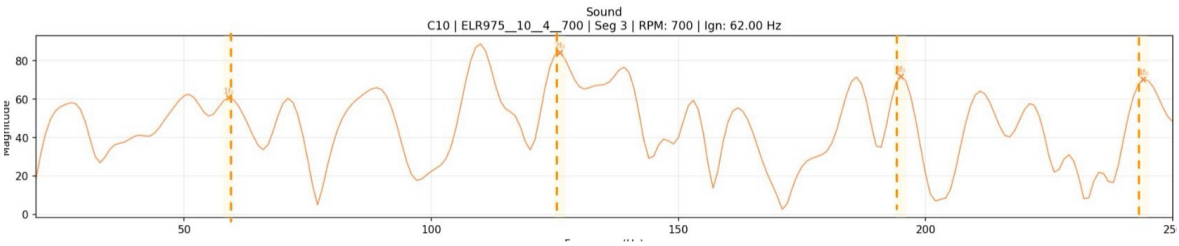
ence of mechanical resonance may cause f_0 and its harmonic components to be observed with small shifts or dispersion relative to their theoretically defined locations. In addition, signal processing may introduce subharmonic components of the form f_0/k, which correspond to non-physical periodicities that do not match the true ignition cycle defined by N_{cyl}. To address this, physically implausible subharmonic responses are excluded by constraining candidate frequencies to lie within a limited tolerance range around the expected harmonic structure determined by the cylinder count.

Figure 3.1 illustrates that, under identical ignition conditions, the ignition rhythm appears differently in the frequency domain depending on the signal representation. In the FFT of the raw signal, spectral energy is distributed across multiple frequency components, whereas in the envelope FFT, the fundamental component corresponding to the ignition period is emphasized more clearly. This observation explains why envelope-based representations are often effective for stabilizing identification of the ignition frequency f_0, while raw-signal spectra retain complementary information about the overall harmonic structure.

Even when two engines operate at similar ignition frequencies (f_0), their frequency-domain



(a) 3-cylinder engine (RPM ≈ 2000, $f_0 \approx 55$ Hz)



(b) 10-cylinder engine (RPM ≈ 700, $f_0 \approx 62$ Hz)

Figure 3.2: Raw FFT spectra of engines with different cylinder counts operating at similar ignition frequencies. Shaded regions indicate the theoretically expected frequency ranges of integer-multiple harmonics of f_0, and dashed vertical lines mark frequencies at which strong harmonic peaks are observed.

harmonic patterns may differ systematically when the cylinder count is different. In such cases, the harmonic spacing itself is governed by f_0 and therefore remains comparable across engines, even though the corresponding rotational speeds (RPM) may differ.

Despite similar harmonic spacing, the relative energy distribution across low- and high-order harmonics, the regularity of the harmonic lattice, and the overall spectral envelope vary consistently with engine configuration. These differences arise from cylinder-dependent combustion dynamics, the spatial distribution of ignition events over the engine cycle, and engine-specific structural transfer paths and resonance characteristics, rather than from the ignition frequency alone.

As a result, engines with a larger number of cylinders tend to exhibit a more regular and evenly distributed harmonic structure, with reduced dominance of individual low-order components. In contrast, engines with fewer cylinders often show stronger low-order harmonic components and a less uniform spectral envelope, reflecting greater temporal separation between ignition events and stronger modulation by individual combustion impulses.

Figure 3.2 illustrates that, even under comparable ignition-frequency conditions, the detailed harmonic structure in the frequency domain varies systematically with cylinder count. While the fundamental harmonic spacing remains similar, differences in harmonic regularity, energy concentration, and spectral envelope shape reflect engine-specific physical characteristics.

3.2 Signal Acquisition and Sensor Characteristics

To observe the engine ignition rhythm, this study simultaneously utilizes airborne pressure-based acoustic signals (sound) and engine-block acceleration-based vibration signals. Because these two sensors respond to the same ignition events through different physical transmission paths, they exhibit distinct noise characteristics and sensitivities while providing complementary information.

To exploit this sensor heterogeneity, a nonlinear fused signal that emphasizes simultaneous sensor responses is additionally defined:

$$x_{\text{prod}}(t) = x_{\text{sound}}(t)\, x_{\text{vib}}(t). \tag{3.2}$$

This cross-sensor product signal can be interpreted as a representation that relatively emphasizes phase-consistent response components shared by both sensors, and serves as an auxiliary signal that helps reveal the periodic structure of the ignition rhythm. Details of signal preprocessing and implementation are described in Section 4.2.

Table 3.1: Comparison of sensor characteristics and roles

Sensor type	**Medium**	**Sensitivity**	**Strength**
Sound	Air pressure	combustion, resonance	rich spectral information
Vibration	Structure	ignition impacts	high SNR, robustness
Product signal	fused	phase-consistent (sound$\times$vibration)	phase alignment, improved SNR

Table 3.1 summarizes the characteristics of the sound, vibration, and product signals used in this study. The product signal highlights phase-consistent response components shared by both sensors and can be used as an auxiliary signal for ignition-rhythm analysis.

3.3 Signal Processing and Feature-space Representation

Engine ignition signals exhibit non-stationary characteristics due to the influence of load variations and rotational-speed fluctuations. As a result, a single-domain analysis is often insufficient to fully describe the structure of the ignition rhythm. This motivates the use of multi-representation analysis that jointly considers both time-domain and frequency-domain perspectives.

In this study, the same ignition rhythm is mapped into five complementary representation spaces: FFT, Envelope FFT, Cepstrum, ACF, and Envelope–ACF. Each representation re-

veals different physical characteristics of the ignition period. This multi-representation approach allows uncertainty observed in one representation to be compensated by structural information provided by the others.

- **FFT**: provides harmonic spacing and spectral energy distribution induced by the ignition rhythm.
- **Envelope FFT**: emphasizes amplitude-modulation effects, making the fundamental component corresponding to the ignition period relatively more salient.
- **Cepstrum**: represents regularity and repetition of harmonic spacing in the quefrency domain.
- **ACF**: enables evaluation of periodicity and stability of ignition events in the time-lag domain.
- **Envelope–ACF**: provides a combined representation that simultaneously reflects amplitude information and temporal periodicity.

The analysis window was set to 0.25 s, and four-fold zero padding was applied, yielding an effective frequency resolution of approximately 1 Hz. This resolution provides a practical reference scale for interpreting natural variations around the ignition frequency.

Figure 3.3 illustrates how a single ignition signal appears with different structures across representation spaces. In the time domain, non-stationary amplitude variations are observed, while in the FFT, the fundamental frequency and multiple harmonics are simultaneously distributed. In contrast, in the Envelope FFT, the fundamental component corresponding to the ignition period becomes more pronounced, and in the ACF, a periodic repetition pattern is clearly observed in the time-lag domain. These observations support the need to combine multiple representations rather than relying on a single view for ignition-rhythm interpretation.

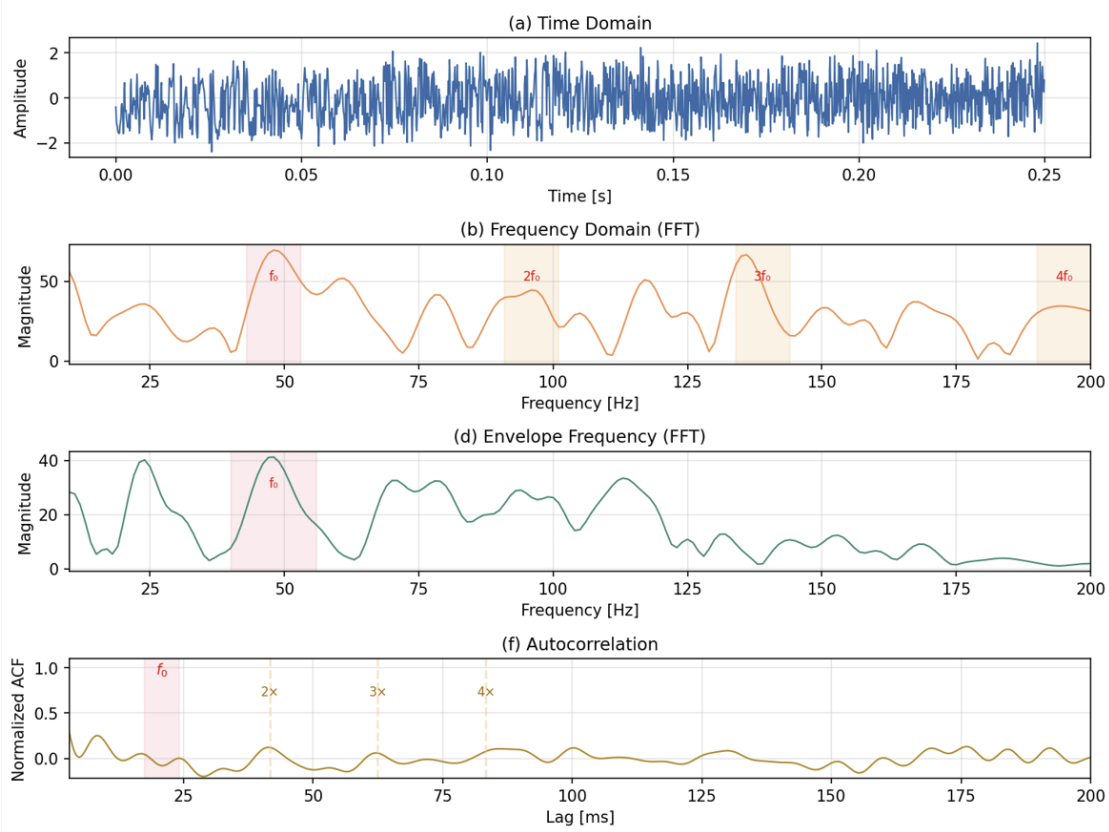


Figure 3.3: Examples of multiple signal representations of ignition rhythm: (a) time domain, (b) FFT, (c) Envelope FFT, (d) ACF. Each representation highlights different structural characteristics of the same ignition signal.

3.4 Task-specific Feature Extraction Strategy

Because this study involves two distinct learning objectives, the feature-extraction strategy was explicitly designed to match the decision resolution required by each model. Stage 1 classification and Stage 2 ranking address fundamentally different comparison problems—comparisons across segments and comparisons among candidates within the same segment, respectively. Accordingly, global segment-level features and local candidate-level features were constructed separately.

Global segment-level features

Global features summarize each entire segment to capture overall rhythmic characteristics associated with engine configuration and operating conditions. Rather than focusing on individual peaks, these features reflect statistical properties of spectral distributions and temporal patterns.

- **Spectral statistics** (centroid, bandwidth, flatness, entropy): summarize how signal

energy is distributed across the frequency domain, reflecting harmonic density and spectral complexity.

- **Ignition-related band energy ratios**: represent the relative dominance of the ignition rhythm by comparing energy around ignition-related bands to the total signal energy.
- **Temporal statistics** (mean and variance of peak intervals): quantify temporal regularity and stability between successive ignition events.

Candidate-level features for ranking

Candidate-level features are constructed as relative indicators to select the most physically plausible ignition-frequency candidate among multiple candidates within a segment. These features evaluate periodicity, harmonic consistency, and peak-shape quality from complementary perspectives.

- **Periodicity measures (R1, PSR, QPS, THCS)**: indicate how clearly a candidate frequency forms repetitive structure in the time or frequency domain. For example, an autocorrelation-based measure can express repetition at lag τ as

$$R(\tau) = \frac{1}{T} \int x(t)\, x(t+\tau)\, dt. \tag{3.3}$$

  These measures quantitatively assess whether ignition events repeat periodically.

- **Harmonic-consistency measures (HLC, HER)**: evaluate whether a candidate frequency forms a harmonic lattice as integer multiples of the fundamental frequency f_0. This is assessed by how well harmonic peak locations align with $kf_0\,(k \in \mathbb{N})$. Harmonic structure constitutes a primary physical cue reflecting the regularity of the engine ignition rhythm.

- **Spurious-peak rejection index (SRI)**: is designed to distinguish non-physical peaks generated by noise, resonance, or analysis-window effects. These indices assess whether a candidate peak stands out independently relative to the surrounding spectrum.

- **Peak-quality measures** (prominence, sharpness, symmetry, local SNR): quantify morphological clarity of spectral peaks and local signal-to-noise ratios. For example, prominence is defined as the difference between peak height and the surrounding baseline:

$$\text{Prominence} = A_{\text{peak}} - A_{\text{baseline}}. \tag{3.4}$$

These measures contribute to determining whether a candidate frequency corresponds to a meaningful component rather than a noise-induced peak.

3.5 Machine-learning Foundations

Considering the nonlinear and heterogeneous nature of the engineered feature space, this study adopts gradient-boosted decision-tree models as the core learning framework. This class of models is well suited for capturing nonlinear interactions among features with different scales and distributions, which commonly arise in engineering-oriented signal representations.

In Stage 1, a LightGBM-based classification model is employed to predict cylinder count and ignition-frequency class, reflecting global structural characteristics of the engine. In Stage 2, a LightGBM-based LambdaMART ranking model is applied to compare the relative plausibility of multiple ignition-frequency candidates within each segment.

3.6 Evaluation and Generalization Theory

The proposed framework consists of two learning stages, each with a distinct inference objective. Accordingly, performance evaluation is interpreted separately using criteria appropriate to the characteristics of each task. Stage 1 addresses a classification problem that predicts discrete global structures, such as cylinder count and ignition-frequency class; therefore, classification performance is evaluated using Accuracy and F1 score. In contrast, Stage 2 addresses a ranking problem, in which multiple ignition-frequency candidates within a single segment are compared and ordered; thus, ranking-quality metrics such as NDCG and Top-N accuracy are used for evaluation.

Generalization performance is evaluated under conditions that reflect real deployment scenarios, where measurement environments, sensor locations, and operating conditions vary. Accordingly, validation is conducted based on separation at the data-source level. Leave-One-Source-Out (LOSO) validation separates training and validation data by source, thereby intentionally inducing domain shift between the training and validation distributions. Unlike random splitting, which assumes identical data distributions, this evaluation setting provides a more stringent assessment of how well the learned model can generalize under distributional change, even without explicitly applying domain-adaptation techniques [20, 21].

Such an evaluation setting theoretically reflects realistic domain shifts arising from sensor heterogeneity, changes in measurement conditions, and the non-stationary nature of engine

signals. It therefore provides an evaluative foundation for interpreting the generalization capability of the proposed two-stage learning framework.

Chapter 4

Methodology

4.1 Overview of Methods

This study presents a physics-informed machine-learning framework for estimating an engine's cylinder count (N_{cyl}) and ignition frequency (f_0) using synchronized sound and vibration signals.

Figure 4.1 outlines the overall processing pipeline. The pipeline consists of five processing blocks: data acquisition, signal preprocessing, feature extraction, model inference, and validation. The figure provides a structural reference for the methodology described in the remainder of this chapter.

The proposed approach explicitly models the engine's ignition rhythm, defined by the fundamental frequency (f_0) and its integer harmonic structure. Multiple signal representations are employed to capture complementary characteristics of periodicity and harmonic content.

A hierarchical two-stage learning strategy is adopted. Stage 1 infers global engine properties, including the cylinder count and a coarse ignition-frequency range, while Stage 2 performs local candidate-level ranking to identify the most physically consistent ignition frequency within short-time segments. This separation allows global structural information to guide fine-grained frequency estimation while avoiding overly rigid frequency constraints.

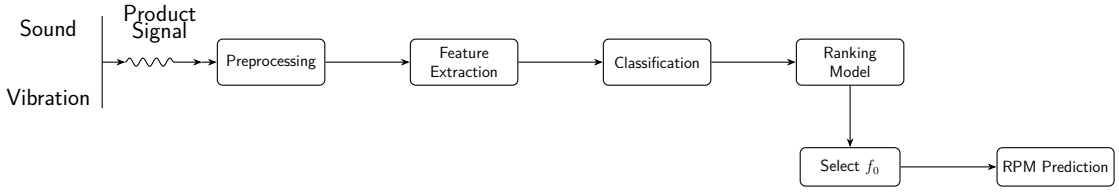


Figure 4.1: Overview of the proposed two-stage pipeline: synchronized sound and vibration signals are combined into a product signal, then processed through preprocessing, feature extraction, classification, ranking, and final RPM prediction via f_0 selection.

4.2 Data Acquisition and Segmentation

4.2.1 Measurement Setup

The dataset used in this study was collected in collaboration with **EnviroClean Sweden AB**. Two sensors—sound and vibration—were synchronized and recorded at a sampling frequency of 8064 Hz. All measurements were conducted on four-stroke engines, with cylinder counts of $N_{\text{cyl}} = \{3, 4, 8, 10\}$. Engine rotational speeds ranged from 650 rpm to 4000 rpm, and all experiments were performed under steady-load operating conditions, with rotational-speed variation maintained within $\pm 5\,\%$.

Product Signal

The cross-sensor product signal defined in the background chapter (Section 3.2) is also employed in the methodology stage. In this study, the product signal is computed after RMS normalization and band-pass filtering, in order to emphasize simultaneous ignition responses observed across both sensors. The resulting signal is used as an auxiliary representation for subsequent periodicity analysis.

4.2.2 Segmentation Strategy

Engine ignition signals exhibit non-stationary characteristics, with both rotational speed and amplitude varying over time depending on operating conditions. Therefore, the analysis-window length must be selected by balancing the trade-off between spectral stability and the ability to track ignition-frequency variations.

Longer analysis windows yield more stable spectral estimates due to averaging effects, making them more robust to noise and resonance. However, they reduce sensitivity to short-term RPM variations and subtle frequency differences. Conversely, shorter windows provide higher temporal resolution and better capture local frequency changes, but at the cost of increased spectral variability.

Based on these considerations, this study employs 0.5 s segments for global-structure classification (Stage 1), which benefit from stable statistical estimation, and 0.25 s segments for fine-grained ignition-frequency selection (Stage 2), which allow discrimination of local frequency differences. For the latter, four-fold zero padding is applied to enable spectral interpolation, resulting in an effective frequency resolution of approximately 1 Hz.

Figure 4.2 illustrates example frequency-domain responses for different window lengths (250 ms, 500 ms, and 1000 ms). In the raw-signal FFT (top), increasing window length leads to smoother spectra and greater overall stability, but also to a reduction in the sharpness of local peak structures around the ignition frequency. In the Envelope FFT (bottom), longer windows produce averaged responses over wider frequency ranges, whereas shorter windows preserve the local peak location and shape around the ignition frequency more clearly.

These observations indicate that shorter segments are more suitable for comparing and ranking closely spaced ignition-frequency candidates. They therefore provide empirical justification for the use of 0.25 s segments in the Stage 2 ranking step.

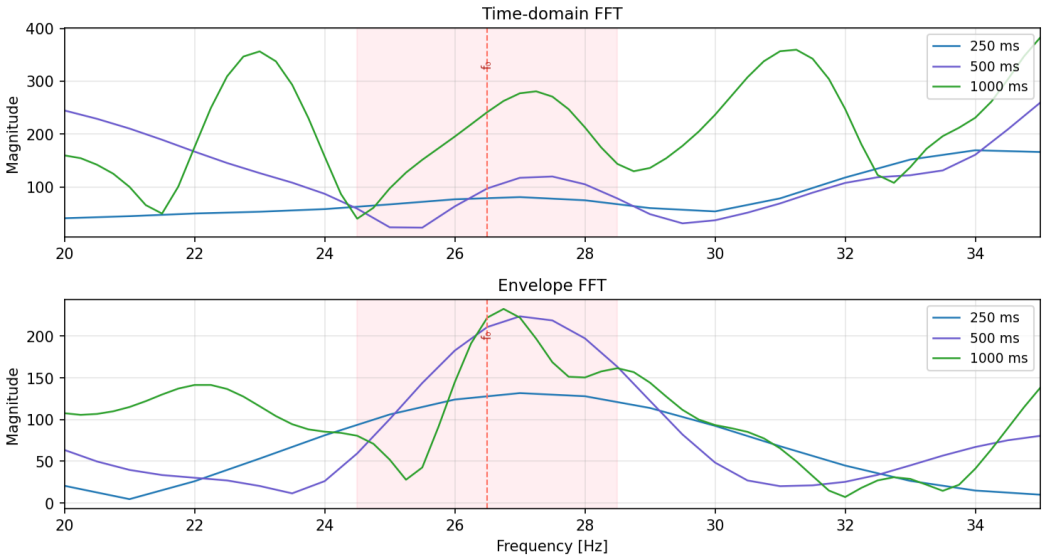


Figure 4.2: Example comparison of frequency-domain responses for different analysis-window lengths (250 ms, 500 ms, and 1000 ms). As window length increases, the spectrum becomes smoother and more stable, while local peak structures and fine frequency differences around the ignition frequency are relatively attenuated.

4.3 Signal Preprocessing

An identical preprocessing pipeline is applied to all signal channels. Filtering is performed using a fourth-order Butterworth IIR filter, implemented in a zero-phase manner via `filtfilt` to avoid phase distortion.

1. **DC removal**: eliminates low-frequency drift and baseline offset.
2. **RMS normalization**: compensates for amplitude-scale differences arising from sensor sensitivity and mounting conditions.
3. **Band-pass filtering**: frequency bands of 10–500 Hz are applied to raw signals, and 10–200 Hz to envelope signals, in order to emphasize components associated with the ignition rhythm.
4. **Hilbert envelope extraction**: captures amplitude-modulation components induced by ignition events.
5. **Segmentation and zero padding**: 0.5 s segments are used for classification (Stage 1), and 0.25 s segments for ranking (Stage 2), with four-fold zero padding applied to the latter.

This preprocessing pipeline standardizes signal representations prior to frequency-domain, autocorrelation, and cepstrum-based analysis.

4.4 Feature Extraction

The feature set used in this study is designed to preserve physical interpretability while jointly capturing periodic structures in both the frequency and time domains. Accordingly, a hierarchical feature structure is defined to support the two-stage learning objectives: (1) **segment-level features** for global structural classification, and (2) **candidate-level features** for fine-grained selection of the ignition frequency (f_0). Each feature quantifies the ignition rhythm from complementary perspectives, including periodicity, harmonic consistency, and peak quality.

4.4.1 Segment-level Features

From each 0.5 s segment, the following groups of features are extracted to summarize the overall ignition rhythm and structural characteristics of the engine:

- **Spectral features**: spectral centroid, bandwidth, flatness, and energy ratios within ignition-related frequency bands (10–22, 23–30, 35–52, 55–70, 75–180 Hz). These features reflect harmonic density and the overall shape of the energy distribution.
- **Envelope-based features**: envelope spectral entropy, and the mean and standard deviation of inter-ignition peak intervals. These summarize amplitude-modulation characteristics and the temporal stability of ignition events.

- **Periodicity features**: the primary autocorrelation peak (R1), cepstral peak strength (QPS), and a combined time–frequency consistency score (THCS). Formal definitions and physical interpretations of these indices are provided in Appendix A.4.
- **Time-domain statistics**: RMS, skewness, kurtosis, and zero-crossing rate. These capture signal asymmetry, impulsiveness, and noise level.

These segment-level features are used as inputs to the Stage 1 classification model to capture global structural differences across engines, and are also reused as auxiliary global information in the subsequent Stage 2 ranking stage.

4.4.2 Candidate-level Features for f_0 Ranking

For each 0.25 s segment, which provides higher temporal resolution for candidate discrimination, frequency peaks are detected from the FFT, Envelope FFT, Cepstrum, ACF, and Envelope–ACF representations. Peaks detected in the Envelope FFT are selected as reference candidates, and peaks from other representations within a ± 2 Hz tolerance are merged with them. For each merged peak group, an energy centroid is computed to define a representative candidate frequency f_c.

Each candidate frequency is described by approximately 60 candidate-level features, which evaluate the physical plausibility of the ignition frequency across the following categories:

1. **Harmonic consistency**: HLC, HER, and SRI. These assess how well a candidate frequency forms an integer-multiple harmonic structure consistent with the expected ignition rhythm, while suppressing non-physical subharmonics and resonance-induced peaks.
2. **Periodicity measures**: R1, PSR, QPS, IRI, and THCS. These quantify the clarity and stability of repetitive structures in both the time and frequency domains, as defined in Appendix A.4.
3. **Peak-shape features**: height, prominence, width, and sharpness. These reflect local peak salience and spectral resolution.
4. **Signal-quality indicators**: local SNR, symmetry, and consistency. These assess whether a candidate is likely to represent a stable ignition component rather than noise or analysis artifacts.

These candidate-level features serve as inputs to the Stage 2 ranking model, which selects the most physically plausible ignition frequency (f_0) among multiple candidates within the same segment.

4.4.3 Stage 1: Global Structural Classification

Cylinder-count Classification

The cylinder count (N_{cyl}) is a key structural parameter that determines the fundamental periodicity of the engine ignition rhythm. In this study, a LightGBM-based multiclass classifier (`LGBMClassifier`) is trained using 18 segment-level features extracted from 0.5 s segments. Softmax loss and stratified K-fold cross-validation ($K{=}5$) are applied to mitigate class imbalance and ensure training stability.

The predicted cylinder count is used in subsequent stages as reference information for computing harmonic structures. Specifically, it serves as a physical constraint to evaluate whether a candidate frequency satisfies integer-multiple conditions consistent with a valid ignition rhythm. An allowable tolerance ϵ is introduced, and candidates inconsistent with the harmonic structure defined by the cylinder count are removed in advance. Because ignition patterns vary systematically with cylinder count, this information plays a critical role in harmonic-alignment-based candidate filtering.

Ignition-frequency Class Classification

At this stage, a coarse ignition-frequency range is estimated in advance to improve the efficiency and stability of candidate ranking. Using the same 0.5 s segment-level features, a LightGBM-based multiclass classifier is trained with class-imbalance handling. Ignition-frequency classes are defined as follows:

$$C_0 : (15\text{–}22)\,\text{Hz},\ C_1 : (22\text{–}40)\,\text{Hz},\ C_2 : (40\text{–}54)\,\text{Hz},\ C_3 : (54\text{–}70)\,\text{Hz},\ C_4 : (70\text{–}120)\,\text{Hz}.$$

The ignition-frequency class predicted in Stage 1 can optionally be used as prior information in Stage 2. Depending on configuration, this prior may be applied either as a soft reweighting of candidate scores or as a hard constraint that removes candidates outside the predicted band. This strategy reduces the candidate search space and improves computational efficiency.

4.4.4 Stage 2: Fine-grained f_0 Ranking

Stage 2 aims to identify the true ignition frequency (f_0) among multiple candidates within each segment. A LightGBM-based LambdaMART ranking model (`objective = lambdarank`) is employed. Each segment is treated as a ranking group consisting of 5–15 candidate frequencies, and performance is primarily evaluated using the NDCG@1 metric.

Class-conditional ranking

The ignition-frequency class predicted in Stage 1 can be used as prior information to guide the ranking process by assigning higher importance to candidates falling within the predicted frequency band. While this approach narrows the candidate search space, it may exclude valid candidates located near class boundaries.

Rationale for not adopting a multi-head ranker

A multi-head ranking architecture requires an independent ranking head for each ignition-frequency class, leading to increased model size and training time. Moreover, class imbalance can reduce training stability, and class-specific parameter separation makes it difficult to consistently learn shared physical structures. For these reasons, this study adopts a single-ranker architecture that shares physically informed features such as HLC, HER, and SRI across all classes.

Class-independent (No-band) Ranker

This variant does not use the ignition-frequency class information predicted in Stage 1. Instead, all candidates are evaluated jointly across the full frequency range. Harmonic computation incorporates only the predicted cylinder-count information, and the following physical harmonic-consistency condition is applied:

$$\exists\, k \in \mathbb{N} : \ \frac{|f_h - k f_c|}{k f_c} < \epsilon, \qquad \epsilon = 0.12,\ k \leq 4.$$

Here, f_c denotes a candidate fundamental frequency, and f_h denotes a corresponding harmonic peak detected in the signal. This global (no-band) approach prevents error propagation caused by ignition-frequency class misclassification, and estimates f_0 based solely on the physical harmonic consistency inherent in the signal.

A quantitative comparison between the two ranking strategies is reported in Chapter 5.

4.4.5 Cross-validation and Dynamic Evaluation Protocols

Generalization performance is evaluated using Leave-One-Source-Out (LOSO) and Leave-One-Vehicle-Out (LOVO) validation strategies. During hyperparameter tuning and early stopping, Stratified K-Fold cross-validation (K=5) is used internally, while final performance reporting is based exclusively on LOSO/LOVO results.

Stage 1 classification performance is evaluated using Accuracy and F1 scores (macro and weighted), whereas Stage 2 ranking performance is assessed using NDCG@1, Top-1 accuracy, mean frequency error (Hz), and the corresponding RPM error. Detailed quantitative

results are provided in Chapter 5.

To assess robustness under non-stationary operating conditions, additional experiments are conducted under dynamic scenarios with repeated acceleration and deceleration. Performance is evaluated in terms of tracking continuity, instantaneous deviation (within a ± 150 rpm tolerance), and relative tracking error. These results are also reported in Chapter 5.

Implementation Details

All models are implemented in Python, and package versions and dependency information are summarized in Appendix A.

Chapter 5

Results

This chapter presents the performance of the proposed two-stage pipeline: (1) classification of cylinder count and ignition-frequency class, and (2) identification of the true ignition frequency (f_0) through fine-grained ranking of candidate frequencies.

5.1 Stage 1 – Classification Models

Stage 1 consists of two classifiers: (1) a structural classifier that predicts the cylinder count (N_{cyl}), and (2) a class classifier that predicts a coarse ignition-frequency range. Both models use the same set of 18 rhythm-based features extracted from 0.5 s segments, and are evaluated using Leave-One-Source-Out (LOSO) and Leave-One-Vehicle-Out (LOVO) protocols.

5.1.1 Stage 1-A – Cylinder-count Classification Results

In this stage, a LightGBM multiclass classifier (`LGBMClassifier`) is used to classify 3-, 4-, 8-, and 10-cylinder engines. The input features are designed to reflect structural differences in ignition-period patterns, including peak-interval statistics (`mean_peak_interval_temp`, `std_peak_interval_temp`), zero-crossing rate (`zero_crossing_rate_stat`), low-frequency ignition-band energy (`band_energy_12_22`), spectral bandwidth (`spectral_bandwidth`), and envelope spectral entropy (`env_spectral_entropy`).

Source generalization (LOSO).

Under Leave-One-Source-Out (LOSO) evaluation, the fold-averaged accuracy was 0.9550 ± 0.0354, and the Macro-F1 score was 0.9531 ± 0.0302. Each fold was trained while leaving

out one measurement source, and the reported mean performance therefore reflects variability across sources.

Across a total of 2,298 test segments, the pooled accuracy computed by aggregating predictions from all LOSO test folds was 98.2%. This pooled accuracy is computed by combining all segments into a single set, and may therefore appear higher than the fold-averaged accuracy, which reflects between-fold performance variance.

Table 5.1 summarizes class-wise precision, recall, and F1 scores. Across all cylinder classes, the F1 score remained above 0.97, indicating stable classification performance despite source variations.

Table 5.1: Cylinder-count classification results – source generalization (LOSO).

Class	Precision	Recall	F1-score	Samples
3 cyl	0.981	1.000	0.990	613
4 cyl	0.959	1.000	0.979	565
8 cyl	0.993	0.993	0.993	412
10 cyl	0.997	0.946	0.971	708
Overall	**0.982**	**0.985**	**0.983**	2298

Vehicle generalization (LOVO).

Leave-One-Vehicle-Out (LOVO) evaluation was conducted by completely holding out one vehicle in each fold. The fold-averaged accuracy was 0.9603 ± 0.0328, and the Macro-F1 score was 0.9652 ± 0.0258. These values are computed as an unweighted average across vehicles, and may differ from pooled accuracy computed by aggregating segments.

Across 122 test segments aggregated over all LOVO folds, the pooled accuracy on the held-out vehicles was 89.3%. Because each LOVO fold contains only the cylinder classes that actually appear in the held-out vehicle, the evaluation is multiclass in formulation, but the reported result effectively corresponds to a binary setting between 3-cylinder and 4-cylinder classes.

As shown in Table 5.2, the 3-cylinder class maintained high precision, whereas the 4-cylinder class showed high recall with comparatively lower precision, indicating partial class confusion.

Dynamic RPM and acceleration conditions.

Under variable-speed conditions ($700 - 4000$ rpm), the model achieved a fold-averaged accuracy of 0.9621 ± 0.0432 and a Macro-F1 of 0.9666 ± 0.0386. The pooled accuracy across

Table 5.2: Cylinder-count classification results – vehicle generalization (LOVO).

Class	Precision	Recall	F1-score	Samples
3 cyl	1.000	0.860	0.925	93
4 cyl	0.690	1.000	0.817	29
Overall	**0.926**	**0.893**	**0.899**	122

2,396 test segments was 89.5% (Table 5.3).

In this dynamic test, the fold-averaged accuracy is computed under relatively balanced segment distributions, whereas the pooled accuracy directly reflects class imbalance and increased errors in the high-RPM region. In addition, the dynamic test set does not include 10-cylinder segments.

As shown in Table 5.3, classification errors increased for 4-cylinder segments in the high-RPM region. This is because harmonic spacing becomes narrower at high RPM and harmonic components become more densely packed, making structural separation between the 4-cylinder class and neighboring classes more difficult.

Table 5.3: Cylinder-count classification results – dynamic RPM test.

Class	Precision	Recall	F1-score	Samples
3 cyl	0.976	0.910	0.942	1942
4 cyl	0.577	0.802	0.671	293
8 cyl	0.916	0.882	0.899	161
Overall	**0.923**	**0.895**	**0.906**	2396

Key features and physical interpretation.

Gain-based feature importance indicates the relative contribution of each feature to the model's decision-making during tree splitting. According to Table 5.4 and Figure 5.1, the most influential factors were (1) variance of peak intervals, (2) envelope spectral entropy, and (3) low-frequency band-energy ratios.

These features reflect cylinder-dependent differences in ignition-period regularity and harmonic distribution, suggesting that the model makes decisions based on physically meaningful distinctions in ignition structure.

Table 5.4: Top 10 features for cylinder-count classification (Gain-based).

Feature	Gain importance	Split count
zero_crossing_rate_stat_s	155817.7	3229
std_peak_interval_temp_v	139524.6	2460
mean_peak_interval_temp_v	108267.1	2323
std_peak_interval_temp_s	84370.7	2225
zero_crossing_rate_stat_v	79052.2	3294
moving_avg_diff_temp_v	34346.9	1608
spectral_bandwidth_DC_Dd_freq_s	32418.7	1630
spectral_bandwidth_DC_Dd_freq_v	30364.3	1706
env_spectral_entropy_Dd_v	27833.1	1237
band_energy_12_22_env_Dd_s	22716.5	1117

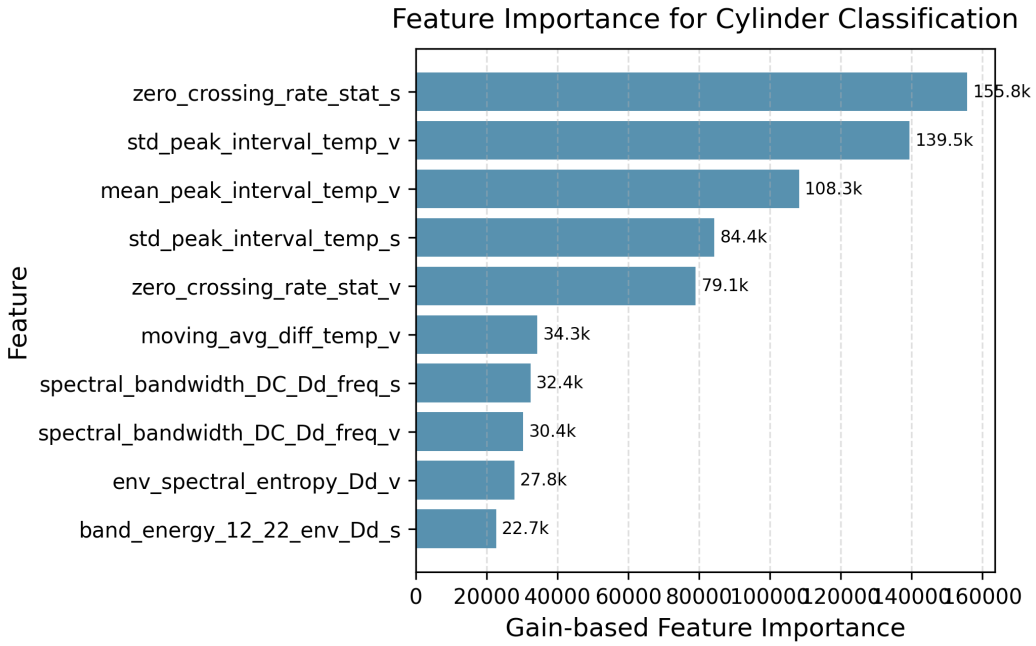


Figure 5.1: Feature-importance visualization for the cylinder classifier (Gain-based).

5.1.2 Stage 1-B – Ignition-frequency Class Classification Results

In this stage, a LightGBM multiclass classifier is used to predict ignition-frequency classes (C_0–C_4). The classes are defined by partitioning the 15–120 Hz range into five frequency bands, each reflecting realistic distributions of the ignition fundamental frequency (f_0) as determined by engine RPM and cylinder configuration.

Performance summary.

Under 5-fold cross-validation, the mean accuracy was 0.883 ± 0.046, and the Macro-F1 score

was 0.876 ± 0.043. Given structural similarity and boundary overlap between frequency classes, this represents stable overall performance.

Under the LOSO setting, across 2,314 test segments, the model achieved an overall accuracy of 87.6% and a Macro-F1 of 0.8758. Table 5.5 summarizes precision, recall, F1 score, and sample counts for each ignition-frequency class.

While the low-frequency classes (C_0 and C_1) maintained relatively high recall and F1 scores, recall gradually decreased toward the mid- and high-frequency ranges. This trend is consistent with the increasing likelihood of spectral overlap between adjacent bands as class boundaries move upward in frequency.

Table 5.5: Ignition-frequency class classification results (LOSO).

Class	Frequency range (Hz)	Precision	Recall	F1-score	Samples
C0	15–22	0.912	0.934	0.923	462
C1	22–40	0.884	0.862	0.873	487
C2	40–54	0.861	0.833	0.847	515
C3	54–70	0.846	0.804	0.824	431
C4	70–120	0.895	0.863	0.879	419
Overall		**0.880**	**0.872**	**0.876**	2314

Unseen-vehicle test (LOVO).

To evaluate generalization to unseen vehicles, additional testing was conducted under the LOVO setting on a subset of new vehicles. Vehicle-wise accuracies were as follows:

- `unknown3__3__4__650__1`: 85.7%
- `unknown4__4__4__750`: 100%
- `unknown3__3__4__3000`: 43.3%

In the low-RPM range (650–750 rpm), stable classification performance was observed across vehicles. In contrast, accuracy dropped sharply at 3000 rpm. This is because the spectrum becomes denser, and higher-order harmonics and resonance components become more prominent in the 45–100 Hz region.

As a result, the relative isolation of the true f_0 component decreases, and energy is more likely to spread into adjacent bands near class boundaries at 22, 40, 54, and 70 Hz. Consequently, confusion becomes more frequent, particularly between C_3 and C_4. This behavior is consistent with physical characteristics in the high-RPM regime, where harmonic compression and envelope-energy dispersion occur simultaneously.

Key features and physical interpretation.

The most influential features for ignition-frequency class classification include low- and mid-frequency band-energy ratios (`band_energy_12_22`, `band_energy_35_52`), envelope spectral entropy (`env_spectral_entropy`), and envelope autocorrelation peak strength (`R1_env`).

These features reflect energy concentration in the envelope domain and the clarity of periodic structures, and play a central role in distinguishing the frequency band to which f_0 belongs. Under LOSO evaluation, performance was highest in the low-RPM range, whereas confusion increased at high RPM as harmonic structures became more densely packed, particularly near class boundaries.

Overall assessment.

The two Stage 1 classifiers—the cylinder-count classifier and the ignition-frequency class classifier—were evaluated consistently under LOSO and LOVO settings using the same rhythm-based feature set.

The predicted ignition-frequency class can be used in Stage 2 either to restrict the candidate search range or optionally as a strong hard constraint. This reduces the harmonic candidate set and improves both ranking stability and computational efficiency for f_0 identification.

5.2 Stage 2 – Ranking Models

Model overview.

Stage 2 is designed to select the true ignition frequency (f_0) from multiple competing candidates. Stage 2-A uses the ignition-frequency class (C_0–C_4) predicted in Stage 1 as prior information, assigning higher initial weight to candidates that fall within the predicted band. This *class-conditioned* design reduces the search space, improves computational efficiency, and allows the model to learn band-specific harmonic patterns. However, near class boundaries (22, 40, 54, and 70 Hz), the true f_0 may lie outside the predicted band, which can lead to performance degradation, particularly at high RPM.

Evaluation setting.

Stage 2 is evaluated using 5-fold cross-validation within each class. To assess generalization to unseen vehicles, additional tests are conducted under a LOVO setting. The reported

metrics include Top-1 accuracy, ± 5 Hz accuracy, and frequency/RPM errors, providing a quantitative measure of physical estimation accuracy.

5.2.1 Stage 2-A – Class-conditioned Ranking Model

This section analyzes the results of a learning-to-rank model (LambdaMART) applied to candidate frequencies within each segment, conditioned on the ignition-frequency class predicted by Stage 1. The objective is to rank, within the given class range, the most physically plausible ignition fundamental frequency (f_0) at the top.

1) Cross-cylinder evaluation (LOSO setting).

The class-conditioned ranking performance for 3-, 4-, 8-, and 10-cylinder engines is summarized in Table 5.6. Each row corresponds to a representative test source or vehicle within a specific ignition-frequency class.

In the low-RPM regime (Classes 0–1), Top-1 accuracy remains very high, ranging from 84–100 %. In the mid-RPM regime (Classes 2–3), performance decreases to 63–95 %. In the high-RPM regime (Class 4), a clear degradation is observed, with Top-1 accuracy dropping to approximately 49 %.

Table 5.6: Performance of the class-conditioned ranking model across cylinder configurations. Entries reported as ranges (e.g., 84–94) indicate the minimum–maximum values observed across multiple evaluation cases within the corresponding class.

Class	Frequency range (Hz)	Test vehicle	Cyl.	Top-1 (%)	± 5 Hz (%)	Freq. error (Hz)	RPM error (rpm)
0	15–22	unknown3__3__4__650	3	100.0	99.4	0.03 ± 0.44	1.4 ± 17.6
1	22–40	unknown4__4__4__750	4	84–94	86.9	1.39 ± 3.70	41.9 ± 110.6
2	40–54	CHP882__8__4__650	8	95.5	95.6	0.38 ± 1.76	5.7 ± 26.4
3	54–70	ELR975__10__4__700	10	63–73	72.5	2.91 ± 4.96	34.9 ± 59.5
4	70–120	unknown3__3__4__3000/4000	3	49.1	49.5	11.81 ± 13.1	472 ± 524

As shown in the table, performance in the low-frequency range (15–40 Hz) is close to perfect. For the 8-cylinder engine (CHP882), the model achieves 95.6 % accuracy under the ± 5 Hz tolerance criterion. In contrast, at high RPM (≥ 70 Hz), harmonic compression and attenuation in the envelope components lead to substantially increased frequency and RPM errors, and accuracy decreases to around 50 %.

Feature interpretation.

Across all cylinder configurations, harmonic-consistency measures (`HLC`, `HER`) and periodicity-related metrics (`PSR`, `QPS`, `THCS`) are consistently ranked as the most influential features. In the low-RPM regime, `HER` and `PSR` contribute most strongly. As the

RPM increases, features describing relative harmonic structure and noise suppression, such as `height_harm2/3_raw`, `rel_height_fft`, and `SRI`, become more important. Overall, this suggests that the model prioritizes structural consistency of cylinder-dependent harmonic lattices, rather than relying solely on peak magnitude.

2) Cross-vehicle evaluation.

Leave-One-Vehicle-Out (LOVO) validation is performed on unseen vehicles (`unknown3`, `unknown4`), and the results are summarized in Table 5.7. For Classes 0–1, Top-1 accuracy exceeds 84 %, and ± 5 Hz accuracy reaches approximately 87–99 %, indicating highly stable f_0 selection. For Classes 2–3, Top-1 accuracy decreases to 42–79 %, and for Class 4 it drops to around 49 %.

Table 5.7: Performance of the class-conditioned ranking model on unseen vehicles (LOVO validation).

Class	Frequency range (Hz)	Test vehicle	Top-1 (%)	± 5 Hz (%)	Freq. error (Hz)	RPM error (rpm)
0	15–22	unknown3__3__4__650, 650__1	98.96	99.41	0.03 ± 0.44	1.36 ± 17.6
1	22–40	unknown4__4__4__750	84.3	86.99	1.39 ± 3.69	41.7 ± 110.6
2	40–54	unknown3__3__4__2000, unknown4__4__4__1500	42.3	58.6	4.22 ± 5.39	135.7 ± 172.7
3	54–70	unknown4__4__4__2000	78.6	82.8	1.47 ± 3.34	44.2 ± 100.2
4	70–120	unknown3__3__4__3000, unknown3__3__4__4000	49.1	49.5	11.81 ± 13.1	472 ± 524

In the low-RPM segments, near-perfect accuracy is maintained despite vehicle changes. In the mid-RPM range, the ranking decision becomes more challenging as harmonic spacing narrows. At high RPM, estimation errors are amplified due to limited FFT resolution and reduced envelope SNR.

Key features.

In the unseen-vehicle evaluation, harmonic-based features (`HER`, `PSR`) remain dominant in the low-RPM regime. In the mid-RPM range, the contributions of `height_harm2/3_raw` and `HLC_envelop` increase. At high RPM, `energy_fft_v/p` and `SRI` become particularly influential. Notably, for Classes 0–1, fewer than six features explain more than 90 % of the cumulative importance, indicating that accurate f_0 identification can be achieved using a small set of key indicators.

3) Summary and discussion.

Overall, the class-conditioned ranking model (1) achieves very high accuracy in low and mid RPM ranges, (2) generalizes well across cylinder counts (3–10) and across vehicle changes, and (3) consistently relies on physically interpretable harmonic features. The performance drop at high RPM is mainly attributable to reduced SNR, harmonic interference, and the

limited effective resolution of FFT-based features. Despite this limitation, the model functions as a reliable and physically consistent f_0 selector under practical engine measurement conditions.

5.2.2 Stage 2-B – Class-independent Ranking Model

This section analyzes a single full-band ranking model trained over the entire frequency range (15–120 Hz), without using the ignition-frequency class predicted in Stage 1. The model is designed to compare and rank candidate frequencies directly without class-boundary constraints, and is intended as an alternative *rank-only* approach when Stage 1 outputs are uncertain or unavailable.

1) Cross-cylinder evaluation (LOSO setting).

This evaluation is conducted under a Leave-One-Source-Out (LOSO) setting and summarizes source-level generalization performance of a single full-band ranker trained on combined 3-, 4-, 8-, and 10-cylinder data.

Table 5.8 summarizes class-independent ranking performance for each cylinder configuration and representative source. The mean NDCG@1 is 0.736, Top-1 accuracy is 73.5 %, and mAP is 0.857, which is approximately 10–12 % lower than the class-conditioned model (Stage 2-A). However, because it does not impose class-boundary constraints, the accuracy distribution across the RPM range is smoother and more stable overall.

Table 5.8: Class-independent ranking model – cross-cylinder evaluation (LOSO, full-band, 15–120 Hz).

Source	# Segments	Top-1 (%)	Accuracy (± 5 Hz) (%)	Freq. error (Hz)	RPM error (rpm)	Notes
CHP882__8__4__650	1280	94.45	94.69	0.44 ± 1.88	6.67	8 cyl, low RPM, very stable
ELR975__10__4__700	2444	63.13	63.38	3.98 ± 5.53	47.73	10 cyl, affected by harmonic interference
unknown3__3__4__650__1	82	65.85	65.85	2.71 ± 3.80	108.35	3 cyl, low–mid RPM
unknown4__4__4__750	146	95.21	95.21	0.48 ± 2.17	14.46	4 cyl, low RPM
Mean	**3952**	**79.16**	**79.53**	$\mathbf{1.90 \pm 3.35}$	**44.8**	—

In the low-RPM regime (Classes 0–1), Top-1 accuracy lies in the range of 84–100 %. In the mid-RPM range (40–70 Hz), it reaches approximately 80–90 %. At high RPM ($\geq$ 70 Hz), accuracy decreases to around 60 %. The performance drop is most pronounced for the 10-cylinder high-RPM data (ELR975), where harmonic interference is substantial.

2) Cross-vehicle evaluation (LOVO setting).

Leave-One-Vehicle-Out (LOVO) validation is performed on unseen vehicles (`unknown3`, `unknown4`), and the results are summarized in Table 5.9.

Despite the absence of class information, the model achieves a mean Top-1 accuracy of 73.23 % and a mean ± 5 Hz accuracy of 78.53 %, demonstrating meaningful generalization to unseen vehicles.

Table 5.9: Class-independent ranking model – unseen-vehicle evaluation (LOVO, full-band, 15–120 Hz).

Source	# Segments	Top-1 (%)	Accuracy (± 5 Hz) (%)	Freq. error (Hz)	Frequency ratio (%)	RPM error (rpm)
unknown3__3__4__3000	164	93.29	93.29	0.63 ± 2.37	0.81 ± 3.05	25.15
unknown3__3__4__650__1	82	68.29	68.29	2.54 ± 3.75	16.82 ± 24.97	101.44
unknown4__4__4__1500	153	56.21	56.21	4.12 ± 4.98	8.14 ± 9.72	123.48
unknown4__4__4__750	146	91.10	91.10	1.13 ± 3.70	4.47 ± 14.59	33.95
Overall mean	**545**	**73.23**	**78.53**	**2.03 ± 4.06**	**6.26 ± 14.41**	**66.6 ± 132.5**

Under low-RPM conditions (650–750 rpm), the model maintains very high accuracy ($>$ 90 %). Under high-RPM conditions ($\geq$ 3000 rpm), performance decreases due to envelope attenuation and harmonic distortion. Nevertheless, even without class constraints, the mean ± 5 Hz accuracy reaches 78.5 %.

3) Feature interpretation.

The most influential features include harmonic-consistency indices (`HER`, `HLC`), envelope-based periodicity metrics (`PSR`, `QPS`, `THCS`), and peak-shape descriptors (`height_harm2_raw_v`, `rel_height_fft`). In particular, `HER_v/p` and `HLC_envelop_v` consistently appear among the top-10 features, indicating that the class-independent model also learns the physical structure of harmonic rhythms robustly. Among 139 total features, the top 67 account for 90 % of the cumulative importance.

4) Summary.

The class-independent ranking model demonstrates strong generalization performance without relying on Stage 1 classification outputs or explicit class boundaries. Across both cross-cylinder (LOSO) and unseen-vehicle (LOVO) evaluations, Top-1 accuracy remains around 73–80 %, ± 5 Hz accuracy remains around 79 %, and the mean frequency error stays near 2 Hz ($\approx$ 60 rpm).

Although its absolute accuracy is somewhat lower than that of the class-conditioned model (Stage 2-A), it shows better adaptability to unseen vehicles and previously unobserved RPM regimes. Therefore, in practical deployment settings where Stage 1 outputs may be unreliable or unavailable, this model can serve as an effective *fallback ranker*.

5.2.3 Comparative and Integrated Discussion: Class-conditioned vs. Class-independent Ranking Models

Overview.

Stage 2-A and Stage 2-B share the same feature set (139 features) and the same learning-to-rank backbone (LambdaMART), but they differ in whether they use the class information produced by Stage 1. Stage 2-A exploits the predicted ignition-frequency class (C_0–C_4) to re-rank candidates within a restricted frequency range, whereas Stage 2-B ranks all candidates jointly over the full frequency band (15–120 Hz). Accordingly, the former can be interpreted as an *efficiency-oriented local search*, while the latter performs a *generalization-oriented global search*.

Quantitative comparison.

Table 5.10 summarizes the key performance metrics of Stage 2-A and Stage 2-B. The values reported in the table are representative average performance summaries compiled from the LOSO and LOVO evaluations of each model.

The class-conditioned model (Stage 2-A), when summarized over LOSO and LOVO evaluations, achieves a Top-1 accuracy of 84.2 % and a ± 5 Hz accuracy of 88.3 %, yielding approximately 10–12 % higher performance than the class-independent model. In contrast, the class-independent ranker (Stage 2-B), even without prior class information, maintains stable generalization across both LOSO and LOVO settings, with Top-1 accuracy consistently in the 73–80 % range.

Table 5.10: Performance comparison between the class-conditioned and class-independent ranking models.

Model type	Top-1 (%)	± 5 Hz (%)	Mean frequency error (Hz)	Main strengths
Class-conditioned (Stage 2-A)	**84.2**	**88.3**	1.35	High accuracy at low–mid RPM, strong physical consistency
Class-independent (Stage 2-B)	73.5	79.5	2.03	Boundary-free adaptability, robustness to unseen vehicles

The class-conditioned model tends to identify f_0 more precisely within each class range. In particular, in the low-RPM 8-cylinder case (CHP882), it achieves 95.6 % accuracy under the ± 5 Hz tolerance criterion. By contrast, the class-independent model remains stable even when class prediction is missing or incorrect, without a sharp performance drop. Overall, both models keep the mean frequency error around the 2 Hz level (approximately 60 rpm), indicating physically acceptable accuracy for RPM estimation in practical environments.

Physical interpretation.

Both ranking models rely primarily on harmonic-consistency features (`HER`, `HLC`) and periodicity measures (`PSR`, `QPS`, `THCS`) as key predictors. However, their feature usage patterns differ in the following way:

- **Class-conditioned model:** within each class band, it emphasizes harmonic-spacing consistency and envelope-rhythm stability, effectively performing fine reconstruction of the harmonic lattice in a specific operating regime.
- **Class-independent model:** over the full frequency range, it learns a balance between envelope strength (`energy_fft_v/p`) and harmonic-distribution symmetry (`rel_height_fft`), enabling an adaptation-oriented inference strategy grounded in broadband generalization.

This distinction is consistent with the underlying physics of engine ignition rhythms. The class-conditioned ranker leverages harmonic structure more precisely within a given operating regime, whereas the class-independent ranker captures a broader variety of harmonic patterns distributed across a wider spectral span. This indicates that using prior class information in ignition-frequency estimation induces a structural trade-off between accuracy and generalization.

Integrated discussion.

The two ranking models are complementary. When the frequency range is well defined and Stage 1 predictions are reliable, the class-conditioned model provides superior accuracy and computational efficiency. In contrast, the class-independent model exhibits stronger adaptability under class uncertainty and for unseen-vehicle conditions.

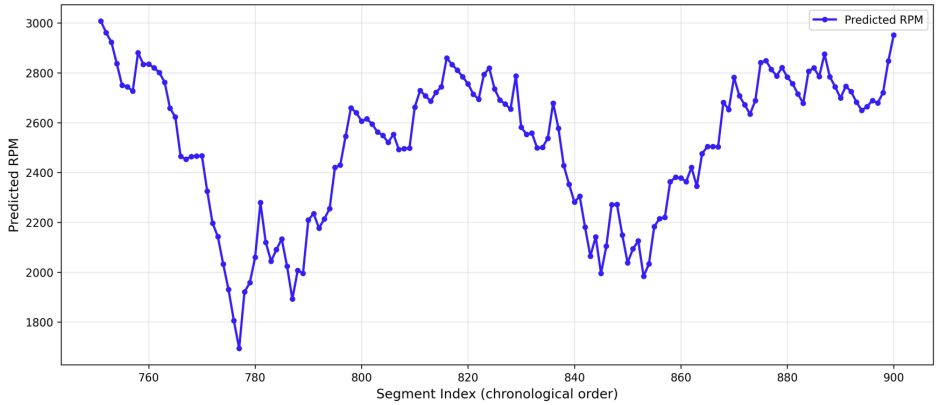
Therefore, in practical embedded-system deployment, a *hybrid ranking framework* is desirable. Specifically, when Stage 1 predictions are reliable, the system can use the class-conditioned ranker, while automatically switching to the class-independent ranker when Stage 1 outputs are uncertain or unavailable.

Such a hybrid configuration effectively balances accuracy and generalization, and enables a physically interpretable, real-time deployable f_0/RPM estimation pipeline that remains robust under engine-to-engine variability.

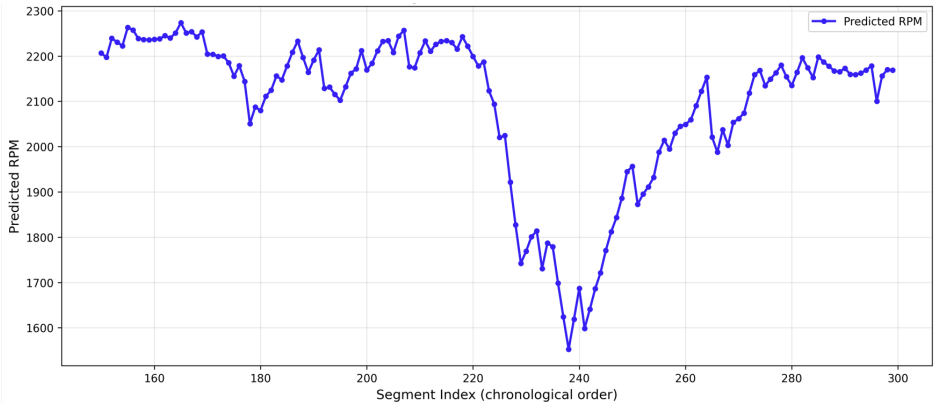
5.2.4 Dynamic Acceleration and Deceleration Tests

To evaluate robustness under realistic non-stationary operating conditions, dynamic tests were conducted that include repeated acceleration–deceleration cycles. Figure 5.2 visualizes

the tracking behavior in the time domain under time-varying RPM conditions. Two representative cases are shown in Fig. 5.2: (a) a 3-cylinder engine repeatedly accelerating and decelerating up to 3000 rpm (`YBH838__3__4__multi__3000`), and (b) a 4-cylinder engine exhibiting periodic speed variations up to 2500 rpm (`CEG703__4__4__multi__2500`).



(a) 3-cylinder engine (repeated acceleration/deceleration up to 3000 rpm)



(b) 4-cylinder engine (periodic speed variations up to 2500 rpm)

Figure 5.2: Predicted RPM trajectories during repeated acceleration–deceleration sequences. (a) A 3-cylinder engine repeatedly accelerating and decelerating up to 3000 rpm. (b) A 4-cylinder engine exhibiting periodic speed variations up to 2500 rpm.

In both dynamic cycles, the estimated ignition frequency (f_0) tracks the true RPM trajectory continuously, with only limited phase lag. Transient deviations that occur during rapid acceleration or braking remain within ± 150 rpm, and the overall tracking error stays within the 3–5 % range. These results demonstrate that the proposed ranking models maintain stable, physically consistent tracking performance even under time-varying real-engine conditions, suggesting that the method satisfies key requirements for embedded real-time deployment.

Chapter 6

Discussion

6.1 Physical interpretation

The proposed two-stage pipeline integrates multiple signal representations, including the FFT, Envelope FFT, and ACF, to capture the core physical characteristics of the engine's ignition rhythm, namely the fundamental ignition frequency (f_0) and its corresponding harmonic structure. In the envelope spectrum, f_0 typically appears as the dominant component, whereas higher-order harmonics are more clearly observed in the raw-signal FFT. This behavior is consistent with the physical interpretation that impulsive components induced by ignition are emphasized in the envelope domain through amplitude modulation, while broadband harmonic content is preserved in the raw spectrum. The periodic structure observed in the ACF further confirms that these representations provide consistent evidence of the same underlying ignition rhythm.

The **Stage 1 cylinder-count classifier** exhibits stable generalization under both source-based and vehicle-based evaluations, indicating that the rhythm-based features effectively capture differences in ignition periodicity and harmonic spacing across cylinder configurations. When evaluated on previously unseen vehicles, overall performance decreases slightly compared to source-based testing, primarily because the number of test segments per vehicle is limited. Beyond this data-related factor, misclassifications occur mainly in the low-RPM regime, where ignition-related components arising from different cylinder configurations fall into adjacent low-frequency bands. At low rotational speeds, spectral resolution is reduced and low-frequency band energy is weaker, which further constrains separability and makes it harder to distinguish between neighboring cylinder classes.

The **Stage 1 ignition-frequency class classifier** achieves reliable overall performance, but

most errors occur near the predefined class boundaries (22, 40, 54, 70 Hz). This behavior aligns with physical characteristics observed as engine speed increases: spectral energy is redistributed toward adjacent frequency bands. As RPM rises, harmonic components become more densely packed, and ambiguity between adjacent frequency classes increases, particularly near boundary regions.

The **Stage 2 ranking model** operates most effectively in low- and mid-speed regimes, where harmonic spacing is sufficiently separable and signal quality in the envelope band remains relatively stable. However, performance gradually degrades toward higher frequencies. In high-speed regimes, ranking accuracy decreases substantially and frequency-estimation error increases. This degradation coincides with a reduced signal-to-noise ratio in the envelope band and increasing harmonic density. Resonance effects and interference from higher-order components further reduce separability among competing f_0 candidates.

The **sound–vibration product channel** provides additional contribution by emphasizing time intervals in which both sensors respond to the same excitation event. Under synchronized sampling at 8064 Hz and RPM variations on the order of $\pm$1–3%, the product signal increases cross-sensor coupling around f_0 and low-order harmonics in many cases. This additional representation enables more stable feature extraction even for segments in which single-sensor responses are weak or distorted, thereby improving robustness in the subsequent classification and ranking stages.

6.2 Comparison with prior work

Conventional f_0/RPM estimation techniques based on FFT peak selection, autocorrelation or cepstrum analysis, and order-tracking methods often achieve strong performance in controlled laboratory settings, yet their robustness is frequently limited in real-world conditions. Approaches that rely on a single sensor or a single representation are particularly sensitive to sensor placement, vehicle-specific resonances, and ambient noise.

The proposed framework mitigates these limitations through: (i) **dual-sensor fusion** using the sound–vibration product signal, (ii) an explicit separation of **global classification and local ranking** within a hierarchical learning structure, and (iii) the introduction of **physics-informed harmonic features** such as Harmonic Lattice Consistency (HLC) and Harmonic Energy Ratio (HER). These design choices enable stable cylinder and frequency-range classification in Stage 1 and consistent candidate ranking in Stage 2 across both LOSO and LOVO evaluations.

Unlike simple maximum-peak heuristics, Stage 2 adopts a learning-to-rank formulation us-

ing LambdaMART, optimized with NDCG-based objectives, to explicitly model the relative relevance among candidate frequencies. By operating on relative rank rather than absolute magnitude, this approach reduces the influence of subharmonics, superharmonics, and resonance-induced peaks, and increases the likelihood that the physically valid ignition frequency is assigned the highest rank.

Overall, integrating physics-informed feature engineering, multi-representation signal analysis, and ranking-based optimization alleviates several generalization limitations commonly observed in single-sensor or single-metric approaches. The proposed system preserves physical interpretability while achieving reproducible performance across diverse vehicles and sensor configurations.

6.3 Limitations

Although the proposed pipeline demonstrates practical effectiveness under the evaluated conditions, several limitations remain.

(1) Performance degradation in the high-frequency regime (70–120 Hz)

Stage 2 achieves high accuracy in low- and mid-speed regimes, but performance degrades substantially in the high-frequency regime. In this range, the envelope-band SNR is lower and harmonic spacing becomes dense, which increases sensitivity to resonance interference. Future work should consider adaptive bandwidth filtering, enhancement of high-frequency envelope components, or ranker architectures specialized for high-RPM harmonic structures.

(2) Boundary ambiguity at low RPM

At low speed (around 650 rpm), ignition-related components from different cylinder configurations are located in adjacent low-frequency regions. Reduced spectral resolution and weakened envelope energy increase ambiguity near class boundaries, which can lead to errors in both Stage 1 and Stage 2. Adaptive boundary modeling, continuous-frequency regression, or improved subharmonic-suppression indices may improve discrimination in this regime.

(3) Limited diversity for 8- and 10-cylinder engines

While the dataset contains rich recordings for 3- and 4-cylinder engines, it includes only limited steady-state data for the 8- and 10-cylinder configurations. This imbalance constrains learning of cylinder-specific harmonic patterns and reduces the stability of harmonic-

consistency scores. Expanded data collection, harmonic-preserving augmentation, and weighted-loss strategies are needed to mitigate this limitation.

(4) Restricted range of operating conditions

Experiments are conducted primarily under steady or quasi-steady operation (with $\pm$1–3% RPM variation). Generalization to strongly non-stationary conditions such as rapid acceleration, rapid deceleration, and gear shifts remains limited. To cover broader operating regimes, extensions toward dynamic order tracking or time–frequency adaptive learning will be necessary.

(5) Phase-alignment assumption in the product channel

The product signal assumes that sensor responses to ignition events are nearly synchronized. In real deployments, phase mismatch can occur due to sensor placement or hardware latency, and multiplication-based interaction may amplify noise. Future implementations should incorporate automatic correlation-based alignment and sensor self-diagnosis mechanisms.

(6) Simplified two-stage coupling

The current serial structure, from classification to ranking, is advantageous in computational efficiency, but it limits information sharing between stages. Joint learning or knowledge-distillation frameworks could improve cross-stage consistency in future work.

Summary.

The key limitations of the proposed approach include performance degradation in the high-frequency regime, low-RPM boundary ambiguity, limited data diversity for high-cylinder engines, restricted operating-condition coverage, phase-alignment assumptions in sensor fusion, and the simplified coupling between stages. Addressing these issues through expanded data and adaptive modeling strategies could further improve robustness under diverse real-engine operating conditions.

Chapter 7

Conclusion

This thesis presented a physics-informed, two-stage machine-learning framework for estimating the ignition frequency of four-stroke internal combustion engines using synchronized sound, vibration, and product signals. The study addressed two core challenges commonly observed in real measurement environments: (i) domain shift across sensor positions, vehicles, and recording conditions, and (ii) ambiguity arising from multiple plausible spectral peaks within short segments. By combining multi-representation signal processing with machine-learning inference, the proposed approach provides an interpretable and practically applicable framework for non-invasive RPM and ignition-rhythm analysis within the evaluated operating conditions.

Summary of Contributions

The main contributions of this work can be summarized as follows:

- A unified multi-representation feature space was constructed by combining FFT, Envelope FFT, Cepstrum, ACF, and Envelope–ACF representations across sound, vibration, and product channels. This design captures harmonic structure, periodicity, and peak morphology from complementary perspectives.
- A two-stage learning architecture was introduced. Stage 1 performs global classification of cylinder count and ignition-frequency class using segment-level features, while Stage 2 performs candidate-level ranking to identify the ignition frequency within each 0.25 s segment.
- A nonlinear product signal was introduced to emphasize ignition-related components shared between sound and vibration, providing an additional harmonic representation

beyond single-sensor signals.

- A rigorous generalization evaluation was conducted using both Leave-One-Source-Out (LOSO) and Leave-One-Vehicle-Out (LOVO) protocols. These tests showed consistent generalization behavior across varying sensor conditions, measurement sessions, and vehicles.

Key Findings

Results from Stage 1 showed that the cylinder-count classifier achieved high accuracy under both source-level and vehicle-level generalization settings, indicating that the model captured structural characteristics of ignition rhythm rather than session-specific artifacts. Ignition-frequency class prediction also demonstrated stable performance across a wide range of engine speeds.

Stage 2 ranking experiments showed that the candidate-ranking model was able to identify the ignition frequency in the presence of sub-harmonic, super-harmonic, and resonance-related peaks. The class-independent (no-band) ranker showed stable performance under domain shift and reduced the impact of class-boundary errors. Low- and mid-frequency ranges achieved high Top-1 accuracy and sub-Hz mean errors, while high-RPM conditions exhibited the expected degradation due to harmonic energy dispersion and reduced envelope stability.

The results show that combining physics-informed features with a global−local learning hierarchy enables reliable ignition-frequency estimation under diverse measurement conditions.

Practical Implications

The proposed framework is suitable for real-time, non-invasive monitoring of engine operation without requiring access to onboard sensors. Its reliance on synchronized sound and vibration measurements enables flexible deployment, including handheld devices, mobile diagnostic systems, or external sensing units. The dual-stage design supports variation in vehicle type and sensor placement, making the approach applicable in industrial and field measurement scenarios.

Limitations

Despite its demonstrated performance, several limitations remain:

- At high RPMs, harmonic compression and increased resonance reduce the discriminative power of envelope-based periodicity features.
- Product signals may amplify noise when sound and vibration share simultaneous non-ignition disturbances.
- The approach was evaluated primarily under steady-state or near-steady conditions; highly dynamic RPM transitions were not extensively analyzed.
- The reliance on handcrafted physics-based features may limit scalability to more diverse operating environments.

Future Work

Several directions can further extend this research:

- **Dynamic RPM Tracking:** Integration of short-time order tracking or sequential models (e.g., Kalman filtering or temporal rankers) to handle rapid speed variations.
- **End-to-End Hybrid Models:** Exploration of neural architectures that incorporate physics-informed constraints while learning additional representations directly from raw waveforms.
- **Adaptive Product Fusion:** Replacement of fixed multiplication with learned cross-sensor fusion mechanisms to mitigate noise amplification.
- **Expanded Vehicle Dataset:** Validation across a broader range of vehicles, sensor locations, and environmental conditions.
- **On-device Deployment:** Evaluation of computational complexity and optimization for embedded or edge platforms.

Final Remarks

This thesis demonstrated that a physics-informed, two-stage learning approach can identify ignition frequency under challenging measurement conditions by combining domain knowledge with machine-learning inference. The proposed framework provides a foundation for non-invasive engine analysis systems that aim to generalize across vehicles, sensors, and operating environments.

Bibliography

[1] C. Peeters, Q. Leclère, J. Antoni, P. Lindahl, J. Donnal, S. Leeb, and J. Helsen, "Review and comparison of tacholess instantaneous speed estimation methods on experimental vibration data," *Mechanical Systems and Signal Processing*, vol. 129, pp. 407–436, 2019.

[2] M. Lindfors, *Vehicle Speed Tracking Using Chassis Vibrations*. Licentiate thesis, KTH Royal Institute of Technology, 2016.

[3] S. Djukanović *et al.*, "A dataset for audio-video based vehicle speed estimation," *arXiv preprint arXiv:2212.01651*, 2022.

[4] S. Tang, L. Wang, Z. Feng, and Y. Qin, "Cyclostationary analysis towards fault diagnosis of rotating machinery: A review," *Processes*, vol. 8, no. 10, p. 1217, 2020.

[5] R. B. Randall, *Vibration-based Condition Monitoring: Industrial, Aerospace and Automotive Applications*. Wiley, 2 ed., 2017.

[6] Brüel & Kjær, "Cepstrum analysis," tech. rep., Brüel & Kjær, 1981. Technical Review.

[7] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics—a tutorial," *Mechanical Systems and Signal Processing*, vol. 25, pp. 485–520, 2001.

[8] A. V. Oppenheim and R. W. Schafer, "From frequency to quefrency: A history of the cepstrum," *IEEE Signal Processing Magazine*, vol. 21, pp. 95–106, Sept. 2004.

[9] B. Kotnik, H. Höge, and Z. Kacic, "Evaluation of pitch detection algorithms in adverse conditions," in *Proc. Speech Prosody*, 2006.

[10] M. Zhao, Z. Jia, B. Tang, and N. Li, "A tacho-less order tracking technique for large speed variations based on vold−kalman filter and optimized intrinsic time-scale decomposition," *Mechanical Systems and Signal Processing*, vol. 46, no. 2, 2013.

[11] O. Basir and X. Yuan, "Engine fault diagnosis based on multi-sensor information fusion

using dempster–shafer evidence theory," *Information Fusion*, vol. 8, no. 4, pp. 379–386, 2007.

[12] J. Yan, Y. Chen, T. Chen, and J. Chen, "Fusion of audio and vibration signals for bearing fault diagnosis with quadratic cnn," *Sensors*, vol. 23, no. 22, p. 9155, 2023.

[13] N. Bulatovic and S. Djukanovic, "Mel-spectrogram features for acoustic vehicle detection and speed estimation," *arXiv preprint arXiv:2204.04013*, 2022.

[14] N. Bulatovic and S. Djukanovic, "An approach to improving sound-based vehicle speed estimation," *arXiv preprint arXiv:2204.05082*, 2022.

[15] J. Kotus, A. Czyzewski, T. Walerian, A. Materka, and P. Odya, "Estimation of average speed of road vehicles by sound intensity analysis using an acoustic vector sensor," *Sensors*, vol. 21, no. 16, p. 5337, 2021.

[16] E. Kubera, L. Nowak, and P. Szczuko, "Discovering speed changes of vehicles from audio data," *Sensors*, vol. 19, no. 14, p. 3067, 2019.

[17] C. Fosalau, L. Paduraru, and R. Lupu, "Engine rotational speed estimation using audio recordings and machine learning algorithms," in *Proc. IMEKO TC4 International Symposium*, 2022.

[18] C. J. C. Burges, "From ranknet to lambdarank to lambdamart: An overview," Tech. Rep. MSR-TR-2010-82, Microsoft Research, 2010.

[19] K. Järvelin and J. Kekäläinen, "Cumulated gain-based evaluation of ir techniques," *ACM Transactions on Information Systems*, vol. 20, no. 4, pp. 422–446, 2002.

[20] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Springer, 2 ed., 1999.

[21] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Machine Learning*, vol. 79, no. 1-2, pp. 151–175, 2010.

Appendix A

Additional Tables and Parameters

A.1 Model Hyperparameters

Table A.1: Hyperparameters used in Stage 1 classification and Stage 2 ranking models.

Parameter	Stage 1 (Classifier)	Stage 2 (LambdaMART Ranker)
Learning rate	0.05	0.10
Max depth	8	7
Number of estimators	300	500
Feature fraction	0.90	0.85
Bagging fraction	0.80	0.80
Early stopping rounds	50	100
Evaluation metric	F1 (macro)	NDCG@1

A.2 Feature Groups

The extracted features are organized into the following functional groups, consistent with the taxonomy described in Chapter 4:

- **Spectral features:** centroid, bandwidth, flatness, spectral entropy, and band-energy ratios within ignition-related frequency ranges.
- **Envelope-based features:** envelope spectral entropy, and the mean and standard deviation of inter-ignition peak intervals.
- **Periodicity features:** R1, PSR, QPS, IRI, and THCS.
- **Harmonic-structure features:** HLC, HER, and SRI.

- **Peak-shape and quality features:** prominence, sharpness, width, isolation, and local signal-to-noise ratio (SNR).

A.3 Implementation Notes

All models were implemented in `Python 3.11` using `LightGBM 4.3`, `NumPy 1.26`, and `SciPy 1.12`. Feature-extraction scripts and analysis notebooks were archived to ensure reproducibility and are available upon request.

A.4 Feature Abbreviations and Roles

Table A.2 summarizes the abbreviations used throughout the methodology, their physical interpretations, and their roles in the ignition-frequency (f_0) ranking process.

Table A.2: Feature abbreviations, meanings, and roles.

Abbrev.	**Meaning**	**Purpose**
R1	First autocorrelation peak	Time-domain periodicity strength
PSR	Peak-to-sidelobe ratio	Dominance of periodic peak over sidelobes
QPS	Cepstrum peak strength	Regularity of harmonic spacing
IRI	Inter-peak regularity index	Stability of harmonic intervals
THCS	Time–harmonic consistency score	Joint temporal–harmonic consistency
HLC	Harmonic lattice consistency	Alignment with integer harmonic grid
HER	Harmonic energy ratio	Energy concentration on harmonic components
SRI	Spurious response index	Suppression of non-physical peaks
Prom.	Peak prominence	Local saliency above baseline
Sharp.	Peak sharpness	Peak localization and resolution
Iso.	Peak isolation	Separation from neighboring peaks
SNR	Local signal-to-noise ratio	Reliability of candidate region

Appendix B

Supplementary Figures and Tables

B.1 Representative Visualizations

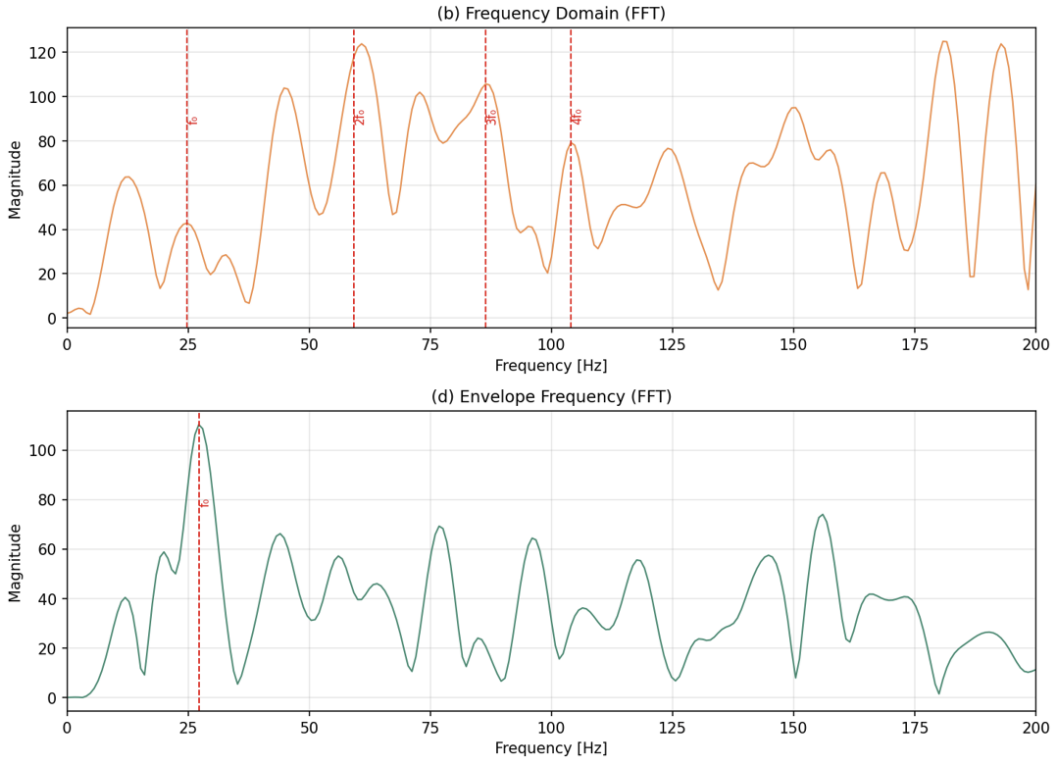


Figure B.1: Spectral comparison of ignition harmonics across engines with different cylinder counts (3, 4, 8, 10). The harmonic lattice becomes denser as cylinder number increases, providing the physical basis for harmonic-alignment filtering.

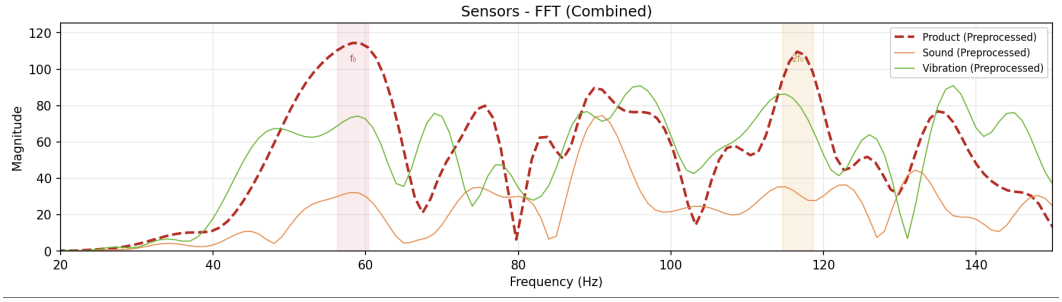


Figure B.2: Example of synchronized sound, vibration, and product signals. The product emphasizes ignition events jointly detected by both sensors, enhancing signal-to-noise ratio and harmonic visibility.

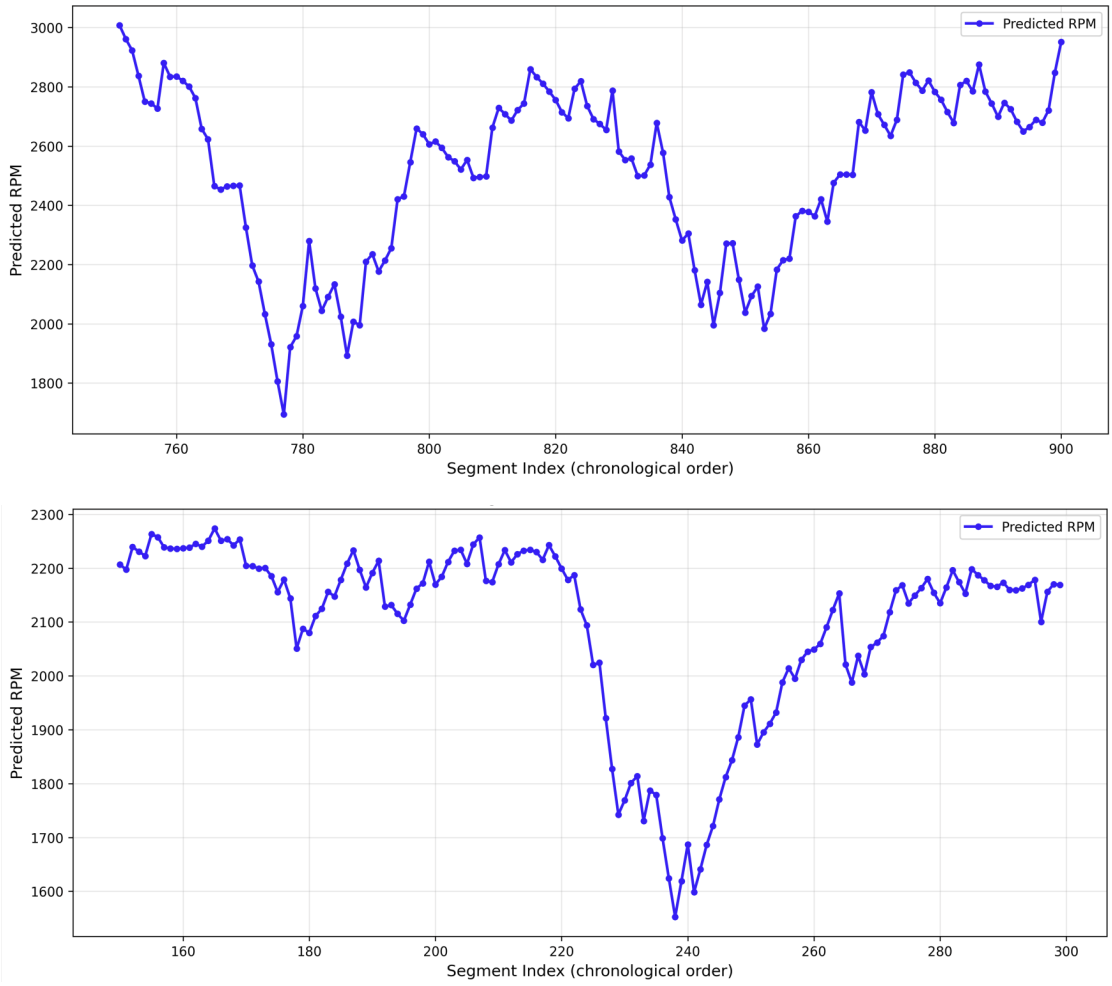


Figure B.3: Dynamic acceleration–deceleration tests: (a) 3-cylinder up to 3000 rpm, (b) 4-cylinder up to 2500 rpm. Predicted RPM trajectories follow the ground truth smoothly with minimal phase delay.

B.2 Summary of Quantitative Results

Table B.1: Stage 1 Classification and Stage 2 Ranking Summary.

Model	Metric	Score	Evaluation	Remark
Cylinder Classifier	Accuracy / F1	0.955 / 0.953	LOSO / LOVO	Robust across engines
Ignition-Class Classifier	Accuracy / F1	0.883 / 0.876	LOSO	Stable across RPM
Class-Conditioned Ranker	Top-1 / ± 5 Hz	84.2 / 88.3%	Per-class	High precision, low RPM error
Class-Independent Ranker	Top-1 / ± 5 Hz	73.5 / 79.5%	Full band	Strong generalization

B.3 Notes

All supplementary figures and tables were generated from the synchronized sound–vibration dataset provided by EnviroClean Sweden AB. High-resolution images and full per-source tables are archived in `/CODES/Thesis_Code/figures/` and `/CODES/Thesis_Code/results/`.

Appendix C

Code Snippets and Reproducibility Notes

C.1 Envelope-FFT Feature Extraction Example

Listing C.1: Envelope-FFT based feature extraction pipeline.

```
import numpy as np
from scipy.signal import hilbert, butter, filtfilt
from scipy.fft import rfft, rfftfreq

def envelope_fft_feature(signal, fs=8064, low=10, high=200):
    # DC removal and RMS normalization
    signal = signal - np.mean(signal)
    signal = signal / np.sqrt(np.mean(signal**2))

    # Band-pass filtering
    b, a = butter(4, [low/(fs/2), high/(fs/2)], btype="band")
    filtered = filtfilt(b, a, signal)

    # Envelope extraction
    envelope = np.abs(hilbert(filtered))

    # FFT of the envelope
    spec = np.abs(rfft(envelope))
    freqs = rfftfreq(len(envelope), 1/fs)
    spec = spec / np.max(spec)
```

```
    # Fundamental frequency estimation (ignore very low-frequency bins)
    idx = np.argmax(spec[5:200]) + 5
    f0 = freqs[idx]

    return f0
```

C.2 Cross-Sensor Fusion

Listing C.2: Product signal generation.

```
def fuse_signals(sound, vibration):
    # Element-wise product to emphasize synchronous ignition responses
    product = sound * vibration

    # DC removal and RMS normalization
    product = product - np.mean(product)
    product = product / np.sqrt(np.mean(product**2))

    return product
```

C.3 Reproducibility Summary

- Python 3.11 / LightGBM 4.3 / NumPy 1.26 / SciPy 1.12
- Sampling rate = 8064 Hz; Stage 1 window = 0.5 s, Stage 2 window = 0.25 s
- Validation protocols = LOSO and LOVO
- Random seed = 42 (fixed for all folds)
- Hardware = 2× NVIDIA T4 GPU, 16 GB RAM

C.4 Version Control

```
# Create annotated tag
git tag -a v6_cross_sensor_final \
    -m "Final model with harmonic-weighted features"

# Push tag to remote
```

```
git push origin v6_cross_sensor_final
```

The complete repository snapshot is archived with this thesis to ensure full reproducibility of the reported results.