



HUST

ĐẠI HỌC BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

ONE LOVE. ONE FUTURE.

Ứng Dụng Machine Learning Trong Chẩn Đoán Hư Hỏng Ô Lăn

Mục tiêu khoá học

Nắm vững quy trình xử lý hoàn chỉnh (Pipeline): từ tín hiệu rung thô đến chẩn đoán lỗi

Tự xây dựng và tối ưu hóa các mô hình SVM và Random Forest

Đánh giá hiệu năng mô hình khách quan, tránh bẫy rò rỉ dữ liệu (data leakage)

Ứng dụng SHAP để giải thích cơ chế đưa ra quyết định của mô hình

 *Tư duy cốt lõi: Khi ổ lăn xuất hiện khuyết tật, mỗi vòng quay kích thích một xung va đập đặc trưng lặp lại tuần hoàn. Mọi công cụ FFT, Envelope, ML, SHAP chỉ là cách phát hiện cái nhip đó.*

Vì sao cần bảo trì dự đoán

Bảo trì phản ứng

Hỏng mới sửa

Chiến lược thụ động: Chỉ tiến hành can thiệp kỹ thuật sau khi thiết bị đã xảy ra sự cố và ngừng hoạt động.

- **Hậu quả:** Gây dừng máy đột ngột, làm tê liệt dây chuyền sản xuất quy mô lớn
- **Thiệt hại:** Chi phí khắc phục khẩn cấp cực kỳ lớn, dễ gây ra hư hỏng cơ khí lan rộng sang các chi tiết liên đới.

Bảo trì phòng ngừa

Theo lịch định kỳ

Chiến lược chủ động dựa trên thời gian vận hành tích lũy hoặc sản lượng sản phẩm

- **Lãng phí:** Buộc phải thay thế các linh kiện cơ khí khi chúng vẫn còn khả năng hoạt động tốt.
- **Rủi ro:** Không thể phát hiện các khuyết tật phát sinh sớm trước kỳ bảo dưỡng, dẫn đến nguy cơ xảy ra sự cố đột xuất.

Bảo trì dự đoán

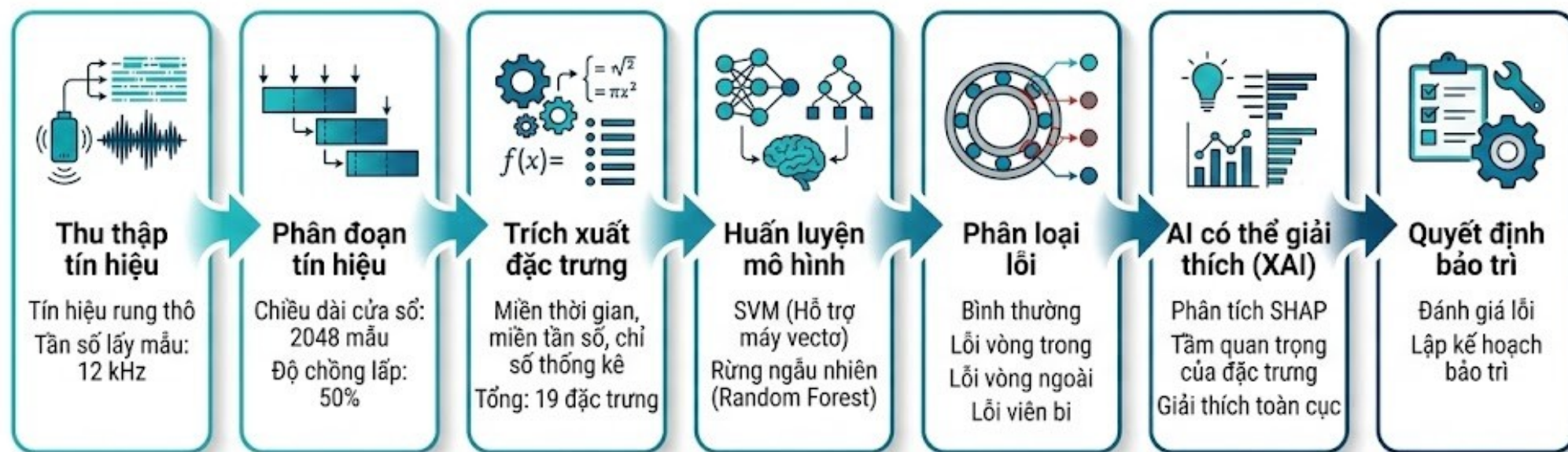
Theo tình trạng thực tế ✓

Chiến lược hiện đại: giám sát liên tục trạng thái sức khỏe của thiết bị bằng cảm biến đo dao động.

- **Mục tiêu:** Đo đặc độ dao động để phát hiện sớm các dấu hiệu hư hỏng của ổ lăn ngay từ giai đoạn khởi phát
- **Hiệu quả:** Chủ động lập kế hoạch bảo dưỡng, tối ưu hóa thời gian dừng máy và khai thác tối đa tuổi thọ hữu ích của linh kiện.

Cảm biến phát hiện sớm lỗi siêu nhỏ trước tai người vài tuần/tháng, dù vậy tín hiệu ban đầu nhỏ và dễ bị che lấp nhiễu và cần xử lý chuyên sâu.

QUY TRÌNH CHẨN ĐOÁN LỖI VÒNG BI DƯA TRÊN HỌC MÁY (MACHINE LEARNING)



Dữ Liệu CWRU

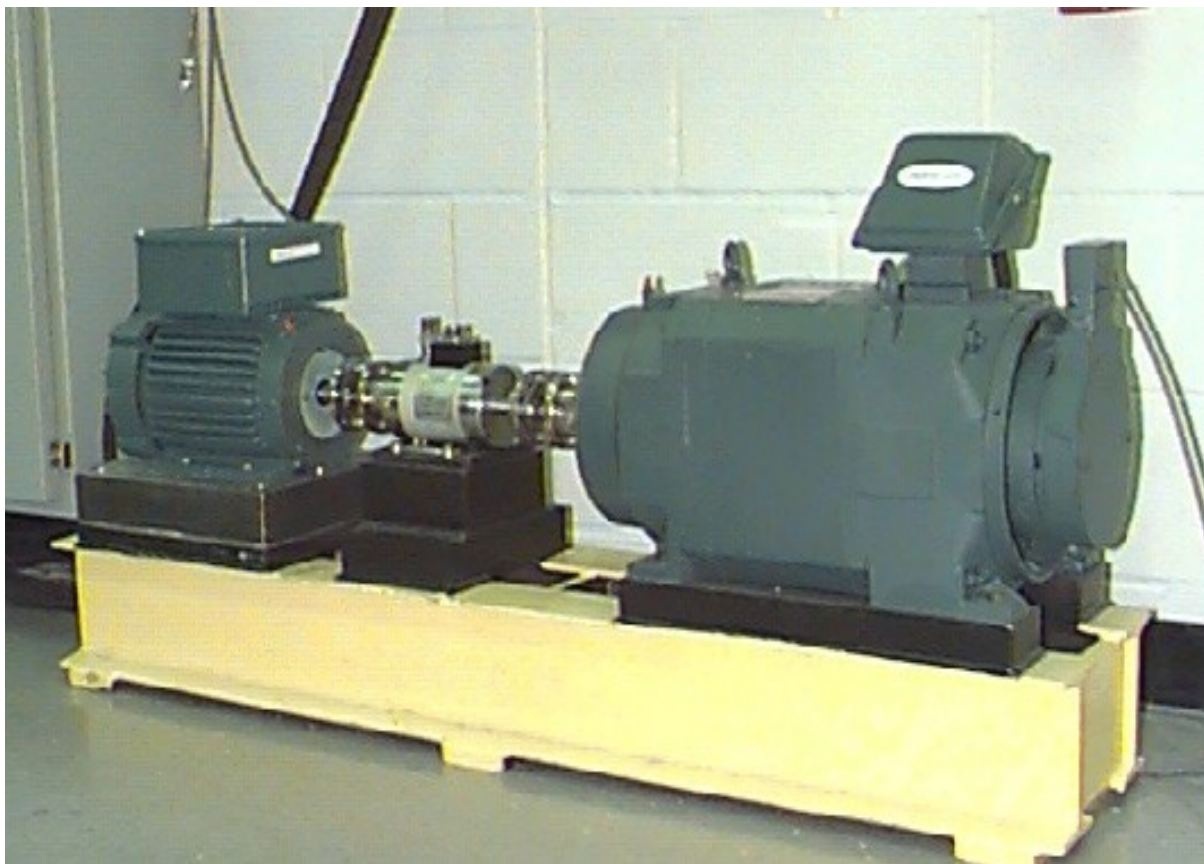
Hệ thí nghiệm, cảm biến, phân loại lỗi và tần số đặc trưng ổ lăn SKF 6205

Cấu tạo CWRU Test Rig & vị trí cảm biến

Phân loại 4 trạng thái kỹ thuật của ổ lăn

Tần số lấy mẫu 12 kHz & mức tải động cơ

❑ Công thức & tần số đặc trưng lỗi SKF 6205



Cấu trúc bộ thử

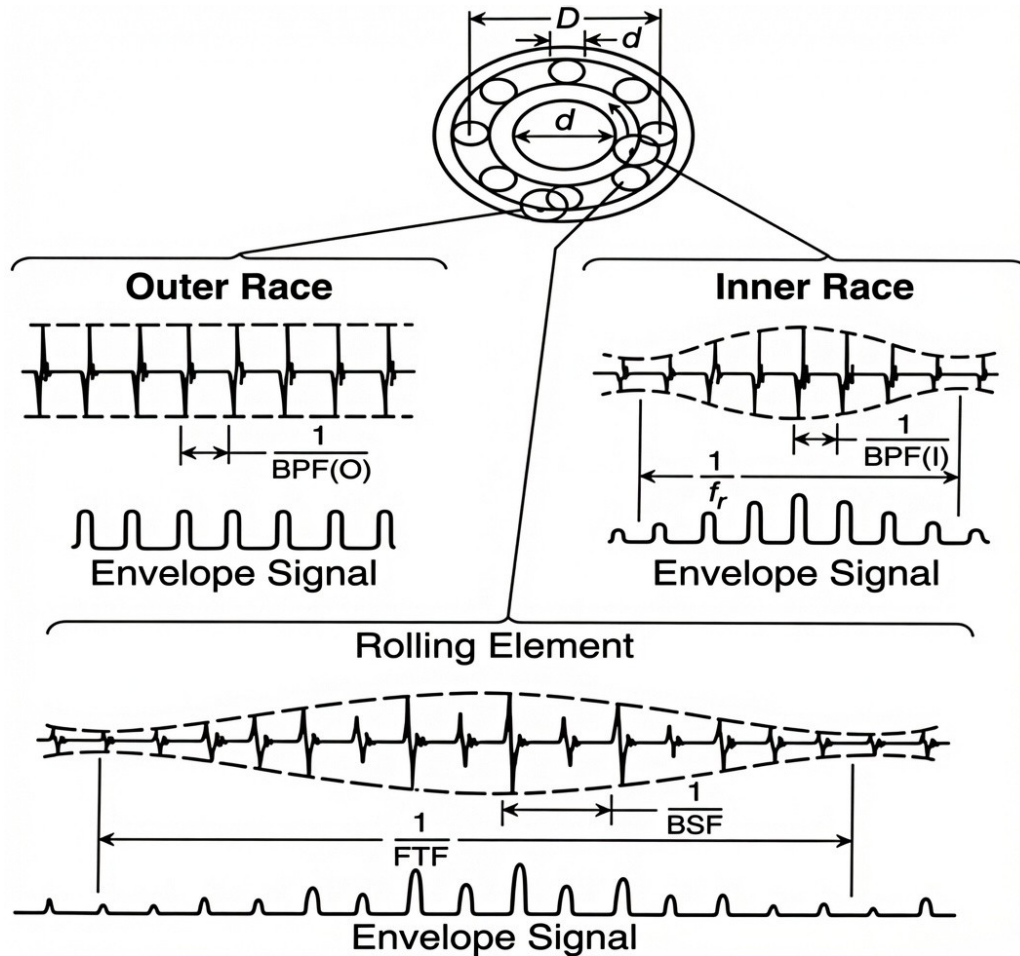
- Động cơ 2 HP, tốc độ 1720–1797 RPM
- Ổ lăn thử SKF 6205-2RS, 6203-2RS
- Khuyết tật nhân tạo EDM: 7, 14, 21 mils
- Tải 0–3 HP qua Dynamometer

Vị trí cảm biến

- DE: Cảm biến gia tốc đặt tại vỏ gối đỡ phía tải
- FE: Cảm biến gia tốc đặt tại vỏ gối đỡ phía quạt
- BA: Cảm biến gia tốc đặt tại đế động cơ

 Hai cảm biến **DE** và **FE** được **bố trí tại hai đầu trục để giám sát độc lập** hai đối tượng khác nhau: Ổ lăn phía tải (**SKF 6205**) và ổ lăn phía quạt (**SKF 6203**).

Trạng Thái Ổ Lăn — Đáp Ứng Rung Động



Bình thường (N)

Biên độ cực nhỏ, tín hiệu phẳng, không có xung bất thường.

Lỗi rãnh ngoài (OR)

Vết nứt nằm trên đường lăn rãnh ngoài cố định. Xung va đập mạnh, chu kỳ ổn định

Lỗi rãnh trong (IR)

Vết nứt nằm trong đường lăn rãnh trong quay cùng trục. Xung bị điều chế biên độ theo chu kỳ quay trục.

Lỗi viên bi (Ball)

Khuyết tật nằm trên bề mặt viên bi. Va đập đa hướng, phức tạp, khó nhận biết nhất.

Tần số hư hỏng đặc trưng

Ý CHÍNH: Mỗi dạng khuyết tật ứng với một tần số kích thích riêng biệt, được tính toán từ thông số hình học của ổ lăn — đây chính là dấu hiệu để nhận diện và xác định loại hư hỏng.

Ký hiệu	Công thức xác định				Hệ số hư hỏng (SKF 6205-2RS)	@1797 RPM (0 HP)
BPFO <i>Tần số va đập khuyết tật rãnh ngoài</i>	Ký hiệu	Công thức xác định	Hệ số hư hỏng (SKF 6205-2RS)	@1797 RPM (0 HP)	3.585	≈ 107 Hz
	BPFO	$\frac{n \cdot f_s}{2} \times (1 - \frac{d}{D} \cos \alpha)$	3.585	≈ 107 Hz		
	BPFI	$\frac{n \cdot f_s}{2} \times (1 + \frac{d}{D} \cos \alpha)$	5.415	≈ 162 Hz		
	BFS	$\frac{d \cdot f_s}{2D} \times (1 - (\frac{d}{D} \cos \alpha)^2)$	2.357	≈ 70.6 Hz		
	FTF	$\frac{f_s}{2} \times (1 - \frac{d}{D} \cos \alpha)$	0.398	≈ 12 Hz		
BPFI <i>Tần số va đập khuyết tật rãnh trong</i>	Ký hiệu	Công thức xác định	Hệ số hư hỏng (SKF 6205-2RS)	@1797 RPM (0 HP)	5.415	≈ 162 Hz
	BPFO	$\frac{n \cdot f_s}{2} \times (1 - \frac{d}{D} \cos \alpha)$	3.585	≈ 107 Hz		
	BPFI	$\frac{n \cdot f_s}{2} \times (1 + \frac{d}{D} \cos \alpha)$	5.415	≈ 162 Hz		
	BFS	$\frac{d \cdot f_s}{2D} \times (1 - (\frac{d}{D} \cos \alpha)^2)$	2.357	≈ 70.6 Hz		
	FTF	$\frac{f_s}{2} \times (1 - \frac{d}{D} \cos \alpha)$	0.398	≈ 12 Hz		
BFS <i>Tần số va đập khuyết tật rãnh ngoài</i>	Ký hiệu	Công thức xác định	Hệ số hư hỏng (SKF 6205-2RS)	@1797 RPM (0 HP)	2.357	≈ 70.6 Hz
	BPFO	$\frac{n \cdot f_s}{2} \times (1 - \frac{d}{D} \cos \alpha)$	3.585	≈ 107 Hz		
	BPFI	$\frac{n \cdot f_s}{2} \times (1 + \frac{d}{D} \cos \alpha)$	5.415	≈ 162 Hz		
	BFS	$\frac{d \cdot f_s}{2D} \times (1 - (\frac{d}{D} \cos \alpha)^2)$	2.357	≈ 70.6 Hz		
	FTF	$\frac{f_s}{2} \times (1 - \frac{d}{D} \cos \alpha)$	0.398	≈ 12 Hz		
FTF <i>Tần số quay vòng cách</i>	Ký hiệu	Công thức xác định	Hệ số hư hỏng (SKF 6205-2RS)	@1797 RPM (0 HP)	0.398	≈ 12 Hz
	BPFO	$\frac{n \cdot f_s}{2} \times (1 - \frac{d}{D} \cos \alpha)$	3.585	≈ 107 Hz		
	BPFI	$\frac{n \cdot f_s}{2} \times (1 + \frac{d}{D} \cos \alpha)$	5.415	≈ 162 Hz		
	BFS	$\frac{d \cdot f_s}{2D} \times (1 - (\frac{d}{D} \cos \alpha)^2)$	2.357	≈ 70.6 Hz		
	FTF	$\frac{f_s}{2} \times (1 - \frac{d}{D} \cos \alpha)$	0.398	≈ 12 Hz		

- n : số viên bi
- d : đường kính viên bi (mm)
- D : đường kính vòng chia (mm)

- α : góc tiếp xúc (°)
- f_s : tần số quay (Hz)

Phân tích dạng sóng

Đánh giá trực quan trong miền thời gian:

❑ Bình thường (N)

Biên độ cực nhỏ (0.2 g), tín hiệu phẳng, không xuất hiện các xung và đập bất thường.

❑ Lỗi vòng trong (IR)

Vết nứt nằm trên đường lăn rãnh trong quay cùng trục. Xung và đập mạnh, không đều và biên độ bị điều biến.

❑ Lỗi vòng ngoài (OR)

Vết nứt nằm trong đường lăn rãnh ngoài cổ đỉnh. Xung và đập mạnh, chu kỳ và biên độ ổn định.

❑ Lỗi viên bi (Ball)

Khuyết tật nằm trên bề mặt viên bi. Va đập đa hướng, phức tạp, khó nhận biết nhất.



Định lượng mức độ rung

RMS (Trị hiệu dụng):

$$RMS = \sqrt{\frac{1}{n} \cdot \sum_{i=1, \dots, n} x_i^2}$$

- Đo **tổng năng lượng dao động** toàn phần của hệ thống.
- Phản ánh **mức độ hủy hoại kết cấu**.

Crest Factor (Hệ số đỉnh):

$$CF = \frac{Peak}{RMS}$$

- Đánh giá **độ sắc nhọn của xung** thông qua tỷ lệ giữa Peak và RMS.
- Nhạy với **giai đoạn chớm lỗi**, nhưng giảm ngược khi lỗi lan rộng (do RMS tăng mạnh).

Peak (Biên độ đỉnh):

$$Peak = \text{Max}|x_i|$$

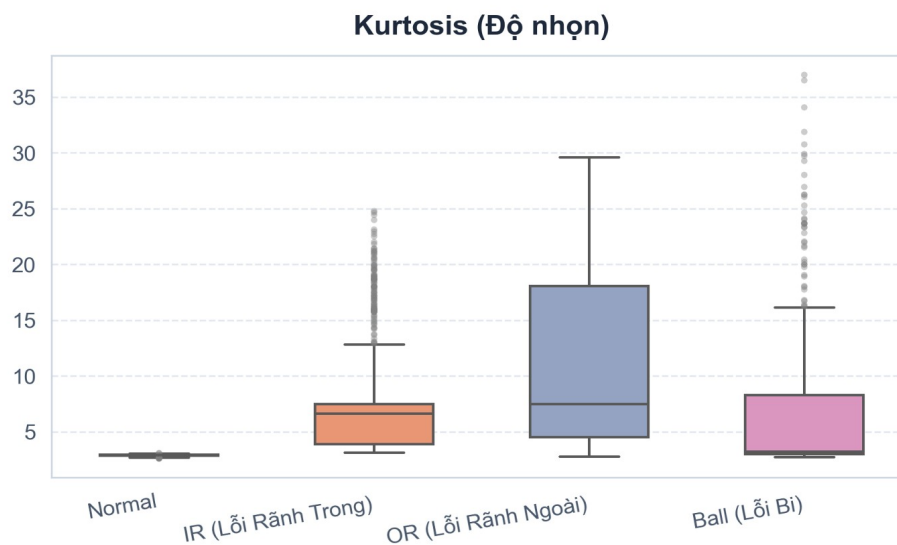
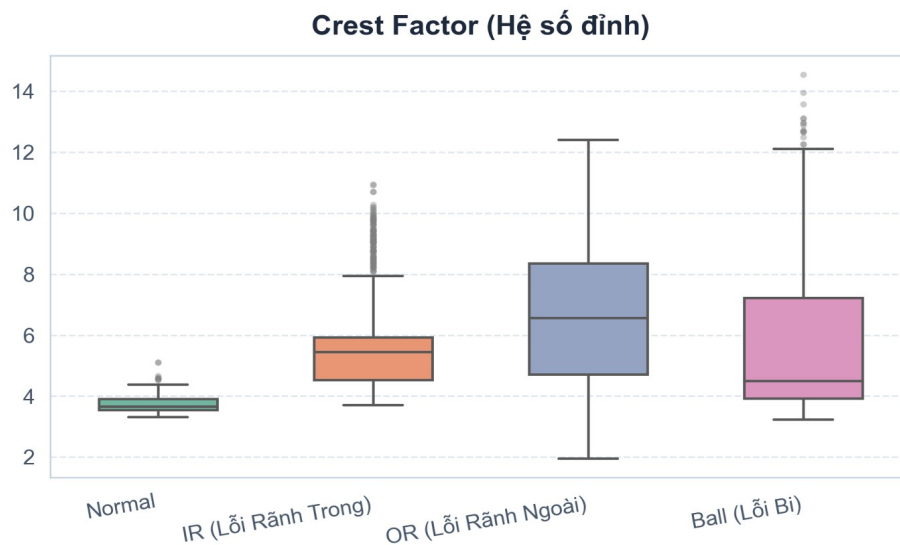
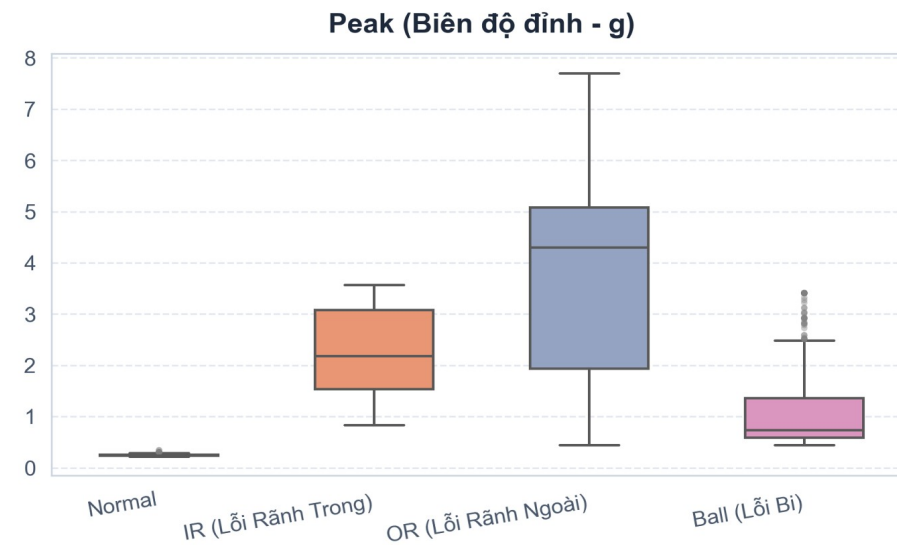
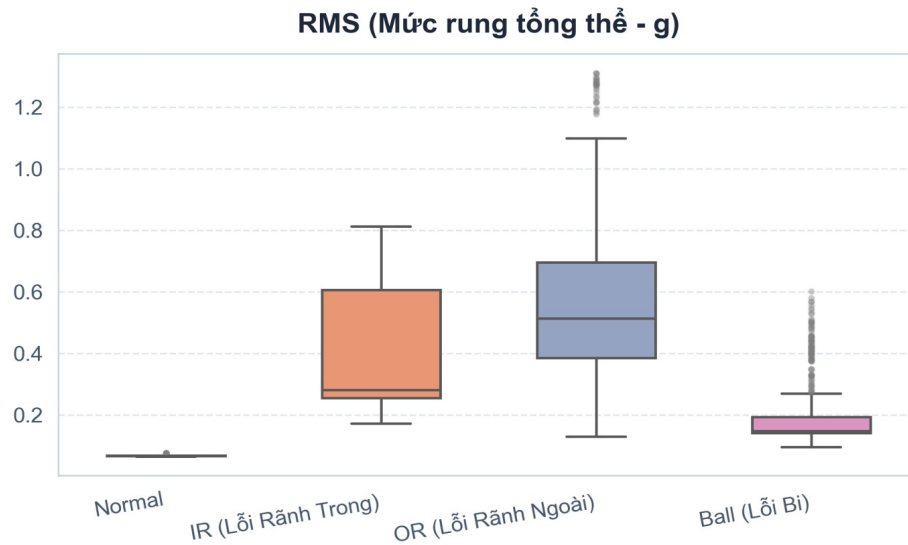
- Ghi nhận **biên độ cực đại tức thời** của các xung và đập động.
- Nhạy với va chạm cục bộ, nhưng dễ bị **sai lệch do nhiễu đỉnh ngẫu nhiên**.

Kurtosis (Độ nhọn):

$$Kurtosis = \frac{\frac{1}{n} \sum_{i=1}^n x_i^4}{RMS^4}$$

- Mô-men bậc 4, đo **độ dày của đuôi phân phối** do xung đột biến gây ra.
- Phát hiện sớm các khuyết tật cục bộ

So sánh mức độ rung động



Phân tích tần số

FFT · Envelope Analysis

Biến đổi Fourier nhanh (FFT)

Hạn chế của FFT thô & Cơ chế điều biên

Kỹ thuật phân tích phổ bao

☐ Nhận diện tần số lỗi đặc trưng

Biến đổi Fourier

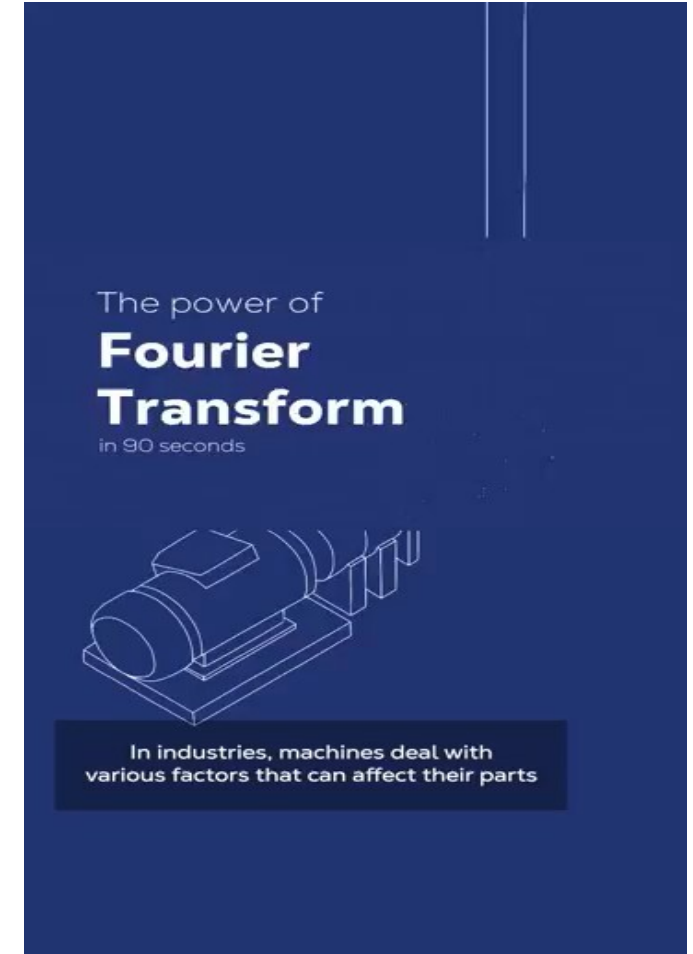
Công cụ toán học nền tảng giúp chuyển đổi tín hiệu từ **miền thời gian** sang **miền tần số** nhằm tách biệt và bóc lộ bản chất các thành phần dao động cơ khí.

$$Y[k] = \sum_{n=0}^{N-1} y[n] \cdot e^{-j\frac{2\pi kn}{N}}$$

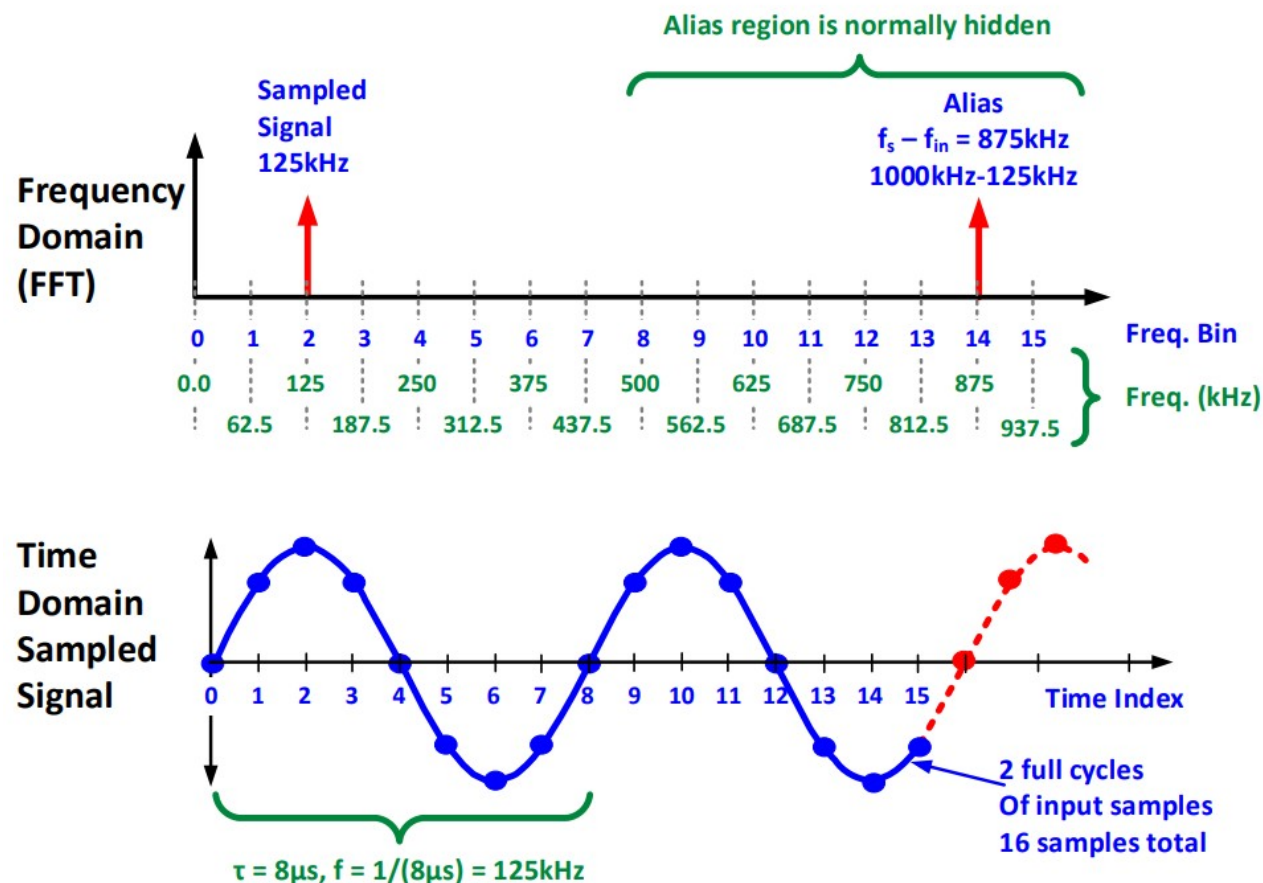
- $y[n]$: Tín hiệu rung động rời rạc trong miền thời gian.
- $Y[k]$: Biên độ phổ tương ứng trong miền tần số.

Giá trị chẩn đoán:

- **Tách phổ:** Mọi tín hiệu phức tạp đều là tổng hợp của các sóng sin đơn lẻ.
- **Giải mã điểm mù:** Nhìn theo miền thời gian tín hiệu rất hỗn loạn và "mù thông tin vị trí", giúp khoanh vùng xác định **vị trí xảy ra hư hỏng** (trục, ổ lăn, động cơ,...).



Ví dụ biến đổi Fourier



Example FFT

$$f_s = 1\text{MHz}$$

$$N_{\text{samp}} = 16$$

$$\Delta f = \frac{f_s}{N_{\text{samp}}} = \frac{1\text{MHz}}{16} = 62.5\text{kHz}$$

$$\Delta t = \frac{1}{f_s} = \frac{1}{1\text{MHz}} = 1\mu\text{s}$$

$$f_{in} = 125\text{kHz}$$

$$k_f = \frac{f_{in}}{\Delta f} = \frac{125\text{kHz}}{62.5\text{kHz}} = 2.0$$

Sampling Rate

Number of Samples

Frequency Resolution

Sampling time

Input signal

Frequency Bin

Note: f_{in} is an exact integer multiple of Δf

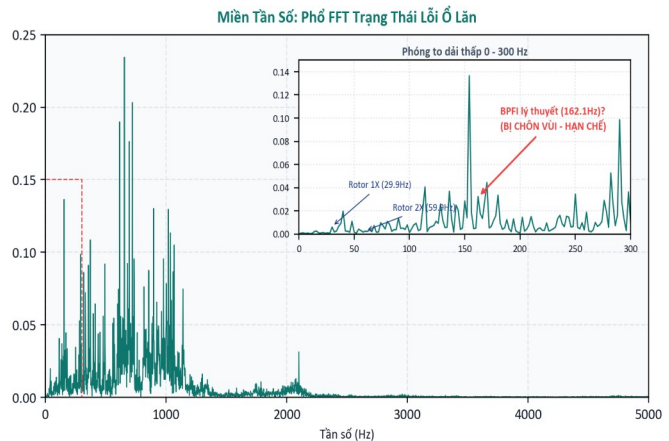
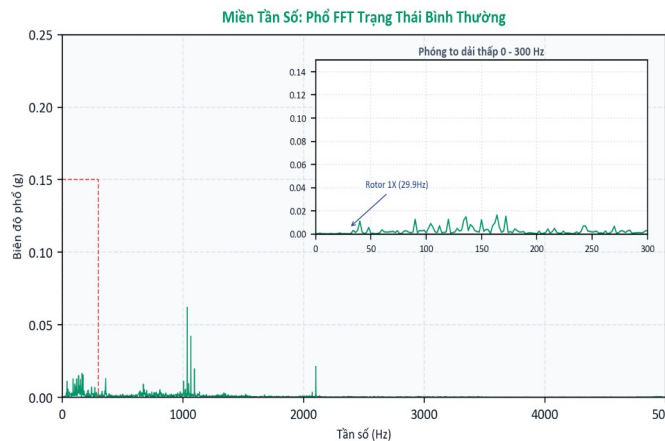
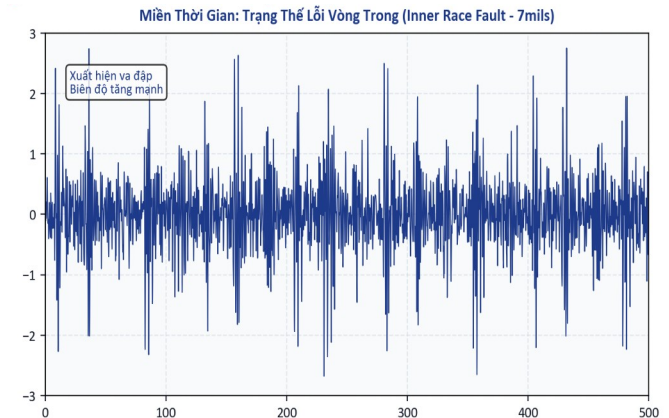
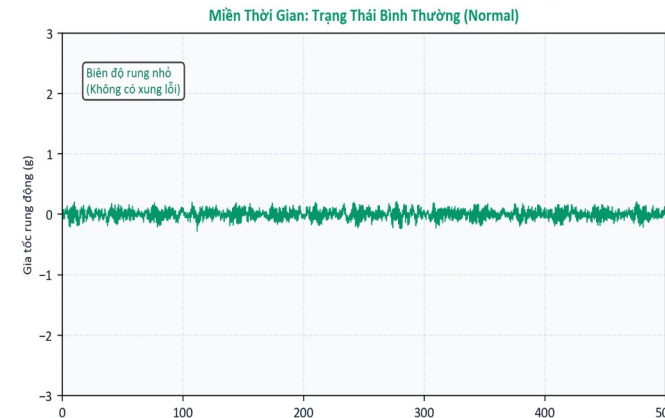
Ưu điểm & Hạn chế của biến đổi Fourier

Ưu điểm:

- **Bộc lộ vùng cộng hưởng:** Chỉ rõ cụm năng lượng tại các dải tần số cộng hưởng tự nhiên do xung lỗi kích thích.
- **Theo dõi nền phổ:** Phản ánh trực quan xu hướng gia tăng năng lượng và hiện tượng **nâng cao nền phổ**.

Nhược điểm:

- **Mất tính thời gian:** FFT không nắm bắt được các thay đổi của tín hiệu theo thời gian.
- **Chưa thể định danh lỗi:** Xung rung động lỗi ổ lăn có năng lượng yếu, dễ bị **chôn vùi dưới nền nhiễu** và các thành phần rung động khác khiến FFT thuần túy chưa phân biệt rõ loại lỗi.



Phân tích phổ đường bao

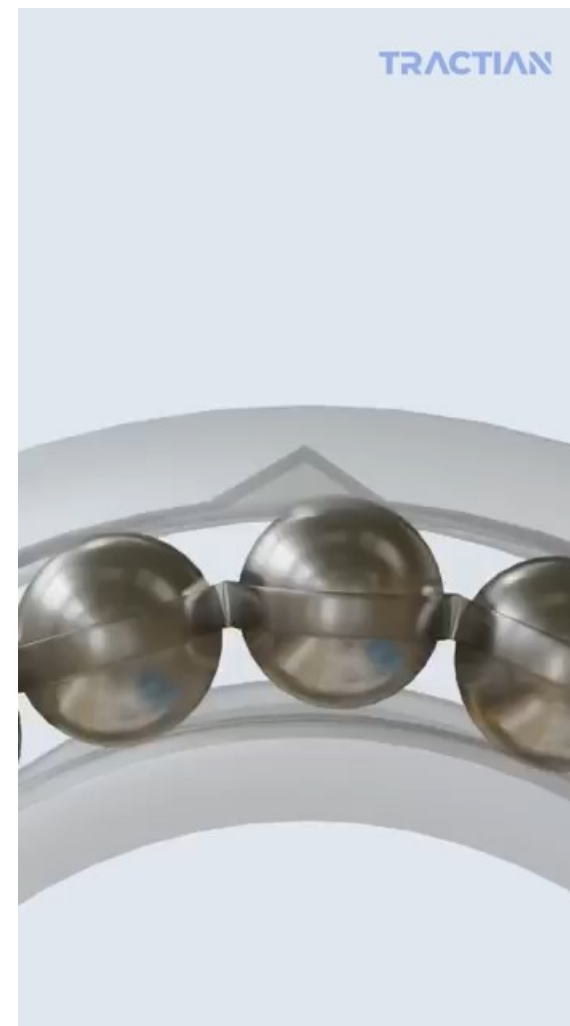
Là kỹ thuật **giải điều chế biên độ (Amplitude Demodulation)** nhằm tách biệt tần số và đập chu kỳ ra khỏi sóng mang là tần số cộng hưởng tự nhiên của kết cấu.

Quy trình thực hiện:

1. **Lọc dải thông:** Cô lập vùng tần số cộng hưởng tần số cao — nơi năng lượng và đập mạnh nhất và sạch nhiễu nền từ dao động quay trục.
2. **Biến đổi Hilbert:** Thiết lập tín hiệu giải tích phức nhằm trích xuất biên độ đường bao và triệt tiêu thành phần sóng mang.
3. **FFT đường bao:** Thực hiện biến đổi Fourier nhanh trên đường bao biên độ vừa thu được.

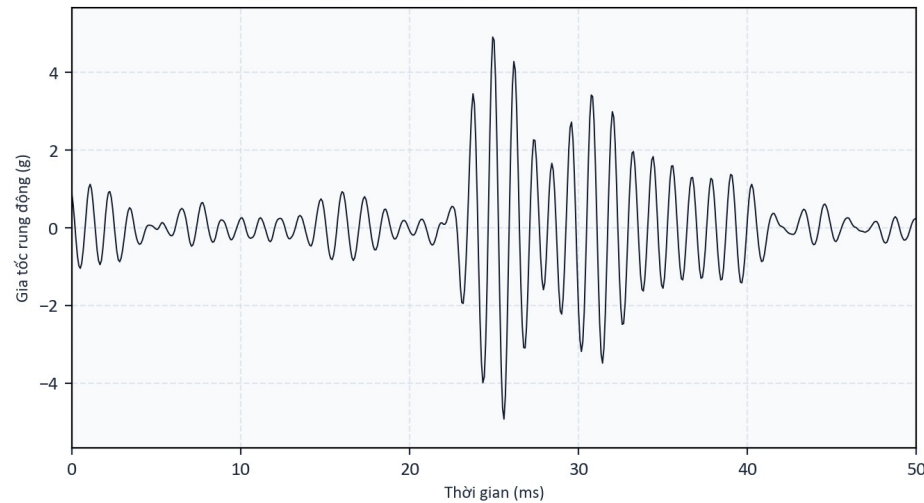
Giá trị chẩn đoán:

- Làm bộc lộ rõ rệt các đỉnh phổ đặc trưng lỗi hình học (BPFI, BPFO, BSF) sắc nét
- Cung cấp vector đặc trưng đầu vào có tính tường minh và phân tách cao, giúp mô hình Học máy hội tụ và phân loại chính xác.

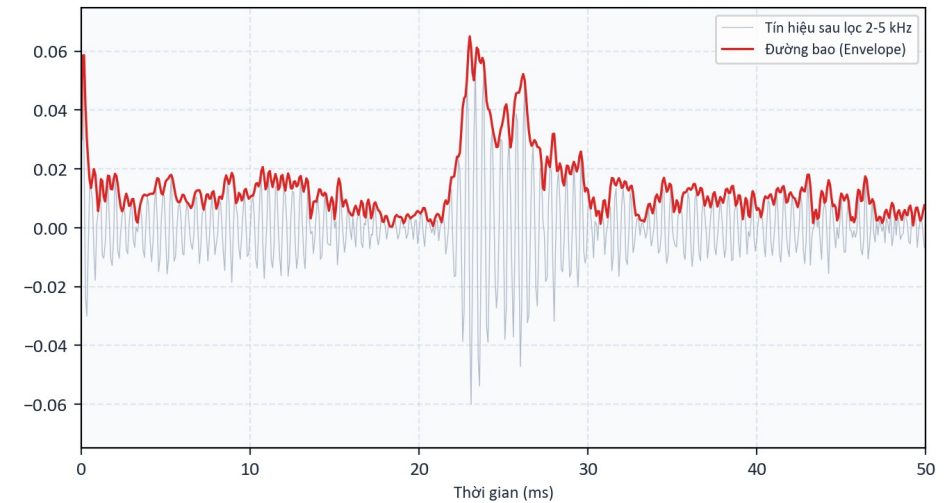


Phân tích phổ đường bao tín hiệu

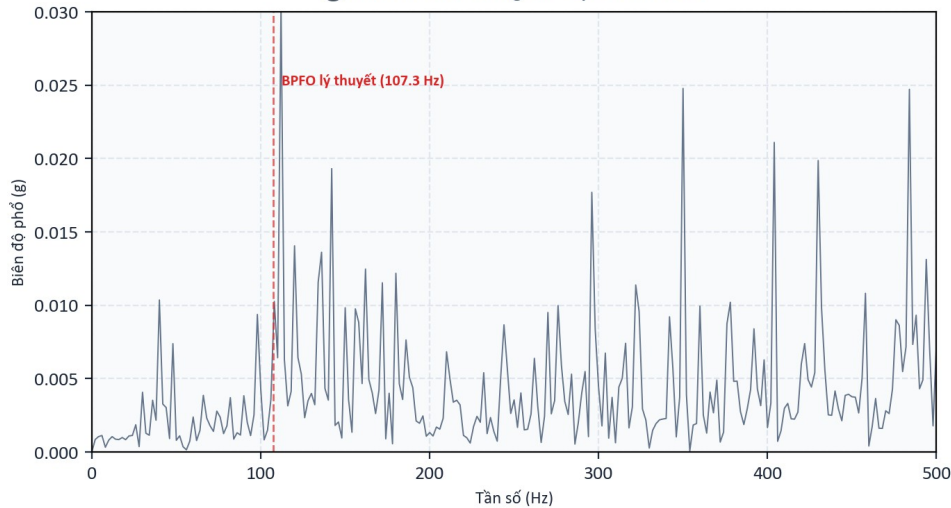
① Tín hiệu rung động thô (Miền thời gian - Zoom 50 ms)



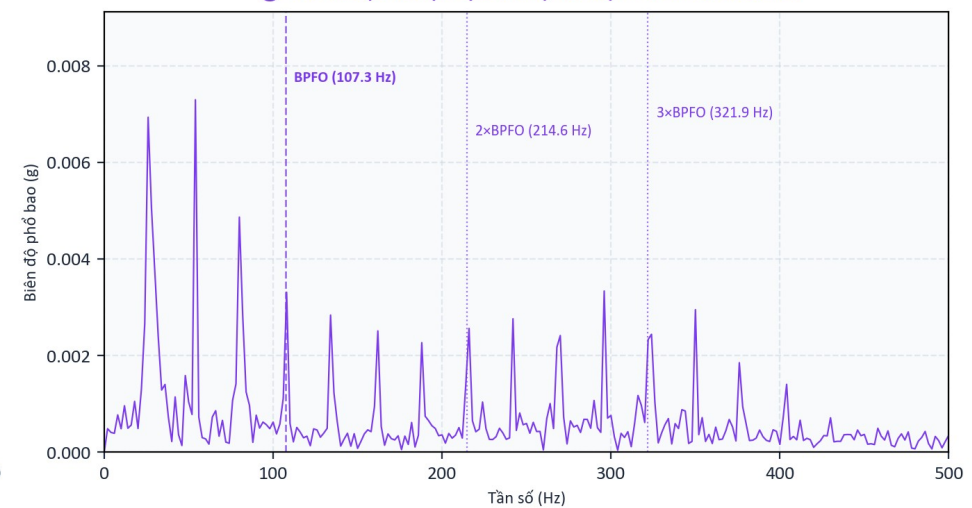
② Giải điều chế: Tín hiệu sau lọc băng thông & Đường bao



③ Phổ FFT thô: Không thể hiện rõ tần số lỗi



④ Phổ bao (Envelope Spectrum): Xác định rõ tần số lỗi BPFO





Phân đoạn dữ liệu

Chuẩn bị đầu vào cho mô hình học máy

Bản chất thống kê của tín hiệu rung & điều kiện tính dừng cục bộ

Chọn kích thước cửa sổ tối ưu.

☐ Chia tập train/test — tránh data leakage



Phân Đoạn Tín Hiệu

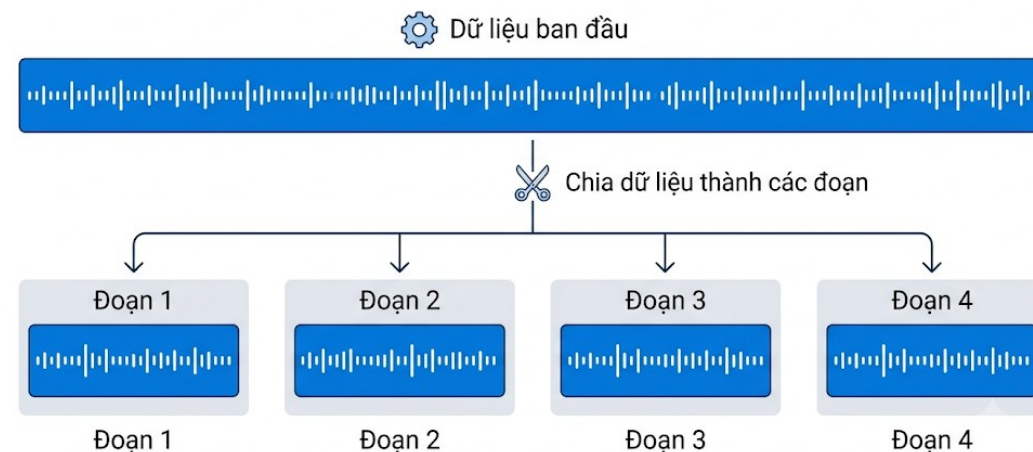
Vấn đề:

- Tín hiệu rung động thô là chuỗi thời gian **không dừng** - phân phối thống kê thay đổi liên tục do nhiễu, tải biến động và xung lồi ngẫu nhiên.
- Mô hình ML về bản chất là bài toán **ước lượng xác suất có điều kiện $P(\text{lỗi} | \text{đặc trưng})$** - nó chỉ học được ánh xạ ổn định khi dữ liệu đầu vào có phân phối nhất quán.

Giải pháp:

Cắt tín hiệu thành các cửa sổ ngắn, **ép về trạng thái gần dừng cục bộ** trong từng đoạn tín hiệu, từ đó:

- Các mô men thống kê (RMS, Kurtosis, Skewness) ổn định và khái quát, đáng tin cậy.
- Vector đặc trưng trích ra phản ánh **đặc tính lỗi thực sự**, không phải biến động ngẫu nhiên của toàn tín hiệu
- Mỗi segment trở thành **1 mẫu huấn luyện độc lập** → tăng kích thước tập dữ liệu từ vài file gốc lên hàng nghìn mẫu



Kích Thước Cửa Sổ & Overlap

Cơ sở lựa chọn kích thước cửa sổ (N=4096)

- ❑ **Tối ưu tốc độ tính toán:** Chọn kích thước cửa sổ là lũy thừa của 2 ($N = 2^k$) giúp thuật toán FFT đạt hiệu năng tính toán cao nhất

- ❑ **Đảm bảo độ phân giải tần số:**

$$\Delta f = \frac{f_s}{N} = \frac{48,000}{4096} = 11,71 \text{ Hz}$$

Độ phân giải này đủ để phân tách các tần số lỗi

- ❑ **Đảm bảo tính dừng:**

- Nếu N quá nhỏ: Phổ bị nhòe, các tần số lỗi bị chồng lấp.
- Nếu N quá lớn: Tín hiệu mất tính dừng cục bộ, giảm số lượng mẫu.

Kỹ thuật Overlap & Chống Data Leakage

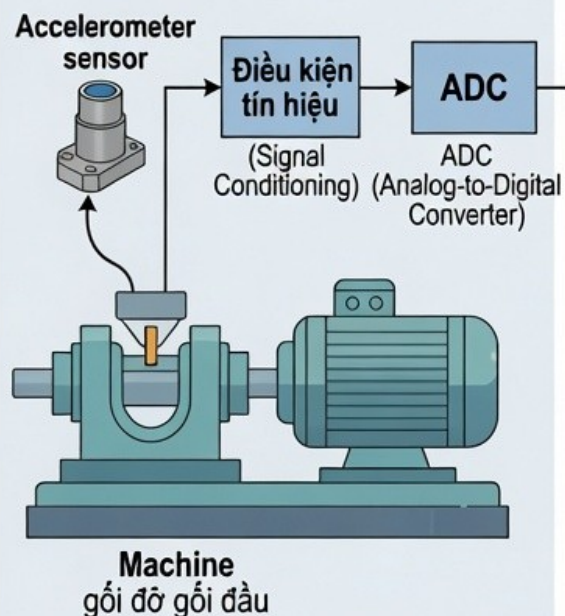
- ❑ **Mục đích :** Tránh mất mát thông tin ở biên. Bước trượt gối chồng (50%) đảm bảo các xung va đập dư thừa vô tình rơi vào ranh giới cắt sẽ nằm ở trung tâm của cửa sổ kế tiếp

- ❑ **Kiểm soát rò rỉ dữ liệu :** Bắt buộc phải phân chia tập Train/Test theo **file dữ liệu thô gốc** hoặc theo **mốc thời gian cô lập** trước khi tiến hành phân đoạn.

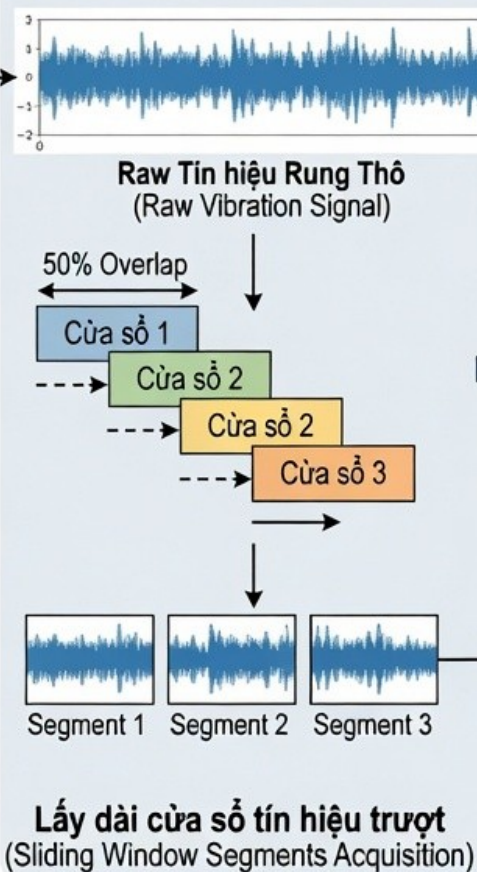
- ❑ Tuyệt đối không trộn lẫn rồi chia ngẫu nhiên, tránh việc mô hình "học thuộc lòng" do các đoạn liên kề chia sẻ chung 50% dữ liệu miền thời gian.

Quy trình chuẩn bị dữ liệu

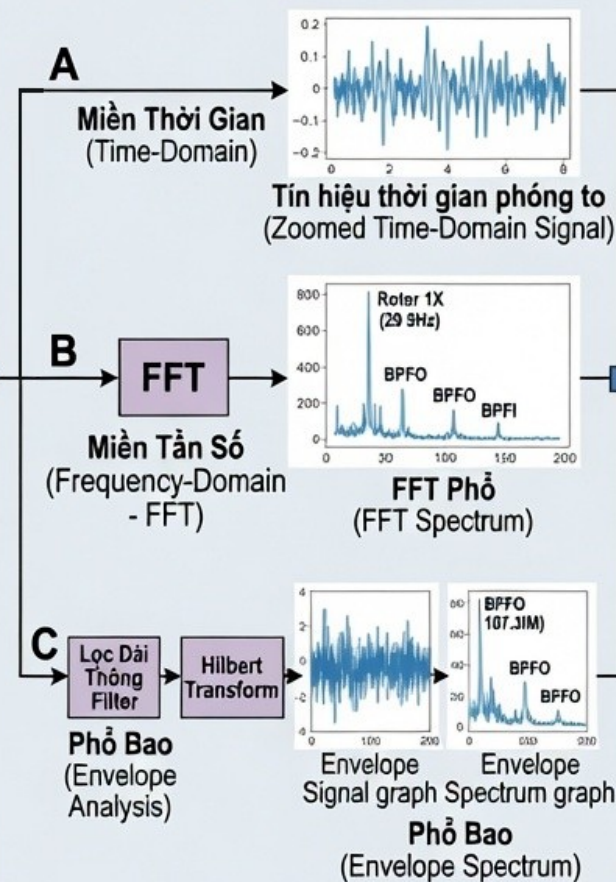
1. Thu thập dữ liệu từ cảm biến (Data Acquisition)



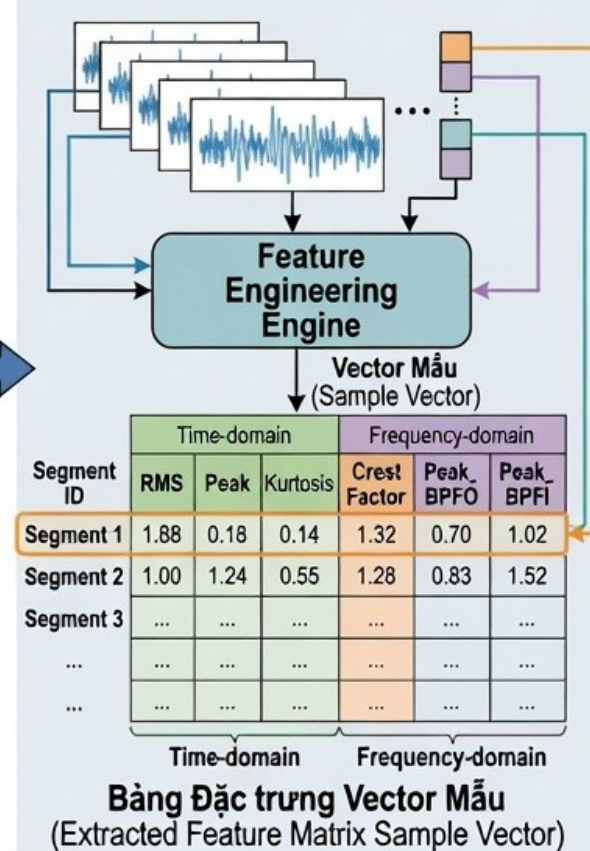
2. Phân đoạn + Overlap (Segmentation + Overlap)



3. Phân tích tín hiệu (Signal Analysis)



4. Trích xuất đặc trưng (Feature Extraction)





Sau khi trích xuất 19 đặc trưng từ hàng nghìn đoạn tin hiệu, chúng ta đối mặt với một **ma trận dữ liệu đa chiều** khổng lồ.



Làm thế nào để xác định chính xác **ranh giới phân lớp** giữa các trạng thái khi dữ liệu chồng lấn phức tạp?

SVM & Random Forest

Nguyên lý cơ bản, cách hoạt động và đánh giá

Giới thiệu trực quan thuật toán SVM

Hàm nhân - nâng chiều không gian để phân tách phi tuyến.

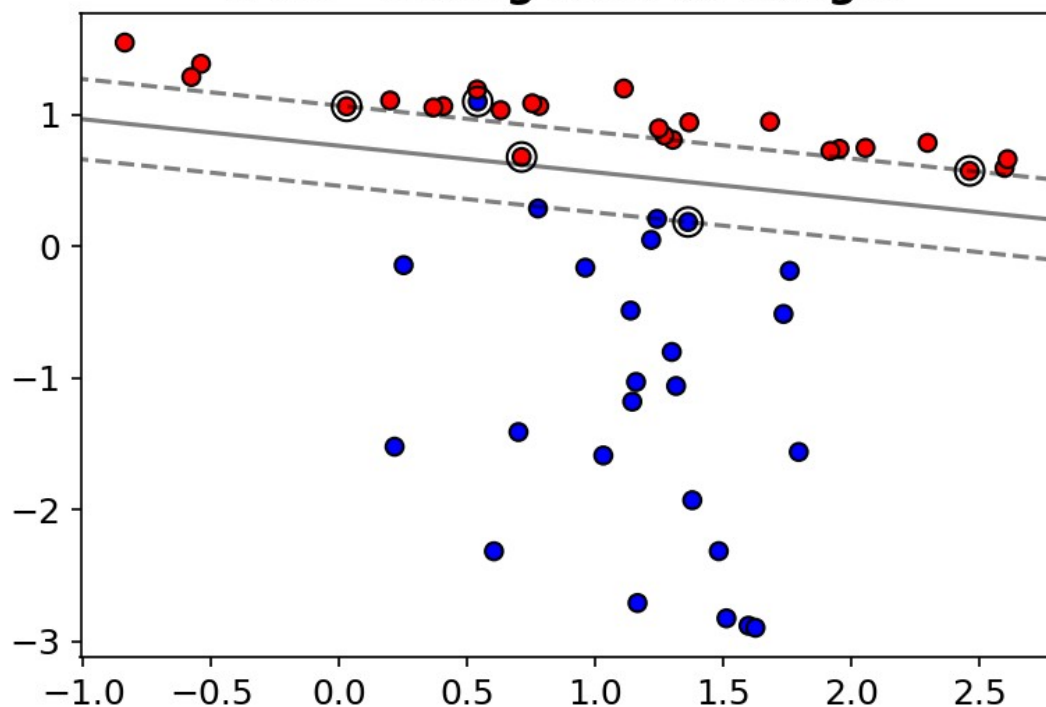
Nguyên lý đồng thuận (Ensemble) và cấu trúc phân nhánh của Rừng ngẫu nhiên.

❑ Phân tích mức độ quan trọng của đặc trưng

Support Vector Machine (SVM)

Bài toán: Phân loại nhị phân hai tập hợp dữ liệu độc lập (Lớp chấm đỏ $y_i = +1$ và Lớp chấm xanh $y_i = -1$) trên mặt phẳng tọa độ 2D.

SVM - Đường biên và Margin



Ý tưởng: Xác định một **siêu phẳng quyết định** phân tách tuyệt đối hai lớp dữ liệu, đồng thời tối đa hóa **biên an toàn** để đạt khả năng khái quát hóa tối ưu.

- Phương trình siêu phẳng phân lớp:

$$w^T x + b = 0$$

- Hàm quyết định:

$$f(x) = \text{sgn}(w^T x + b)$$

Ánh xạ sang bài toán thực tế:

- Đặc trưng đầu vào (x_i): Đặc trưng trích xuất từ tín hiệu
- Nhãn phân loại (y_i): Trạng thái ổ lăn
- Vectơ hỗ trợ: Các điểm giữ liệu gần ranh giới phân loại

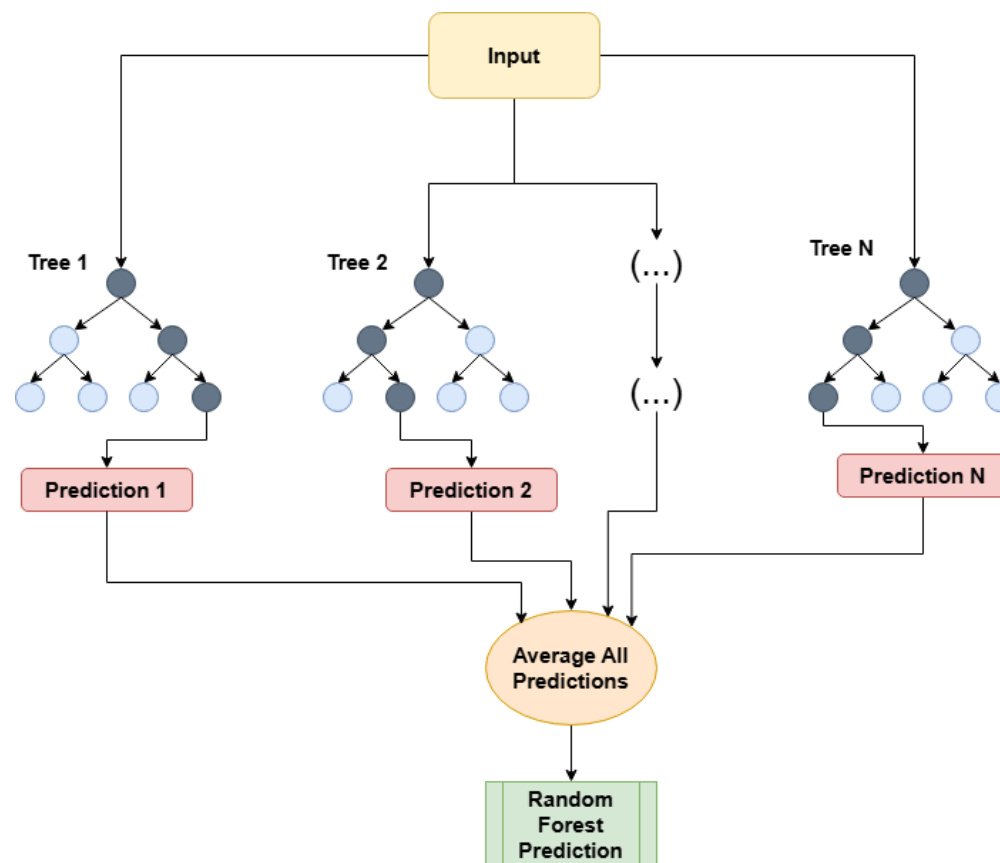
Random Forest

Ý tưởng chính:

- Kết hợp một tập hợp gồm nhiều cây quyết định độc lập được huấn luyện bằng phương pháp **Bagging** (lấy mẫu ngẫu nhiên có lặp lại cả về mẫu dữ liệu và chỉ số đặc trưng)
- Dự đoán cuối cùng được thiết lập dựa trên **cơ chế đồng thuận (bỏ phiếu đa số)** nhằm triệt tiêu sai số của các cây đơn lẻ.

Giá trị chẩn đoán:

- **Chống bẫy quá khớp (Overfitting):** Tính ngẫu nhiên kép giúp mô hình cực kỳ bền vững, không bị phụ thuộc quá mức vào một dải tốc độ hay tải trọng bộ thử cố định.
- **Tường minh hóa đặc trưng (Feature Importance):** Từ công thức Gini, mô hình tự động xếp hạng xem chỉ số rung nào (RMS, Kurtosis hay Crest Factor) nhạy bén nhất với lỗi, giúp kỹ sư tối ưu hóa số lượng cảm biến cần lắp đặt.



Cảm ơn & Hỏi đáp

