**Đề 8: Xác định nhịp sinh học từ dữ liệu nhiệt độ cơ thể**

1. **Mục đích**

Báo cáo này nhằm phân tích dữ liệu nhiệt độ cơ thể đo mỗi 10 phút trong 7 ngày liên tiếp, tập trung vào:

Nhận diện chu kỳ 24h

Phát hiện các dao động utralian (ngắn hơn 24h)

So sánh nhiệt độ ban ngày và đêm

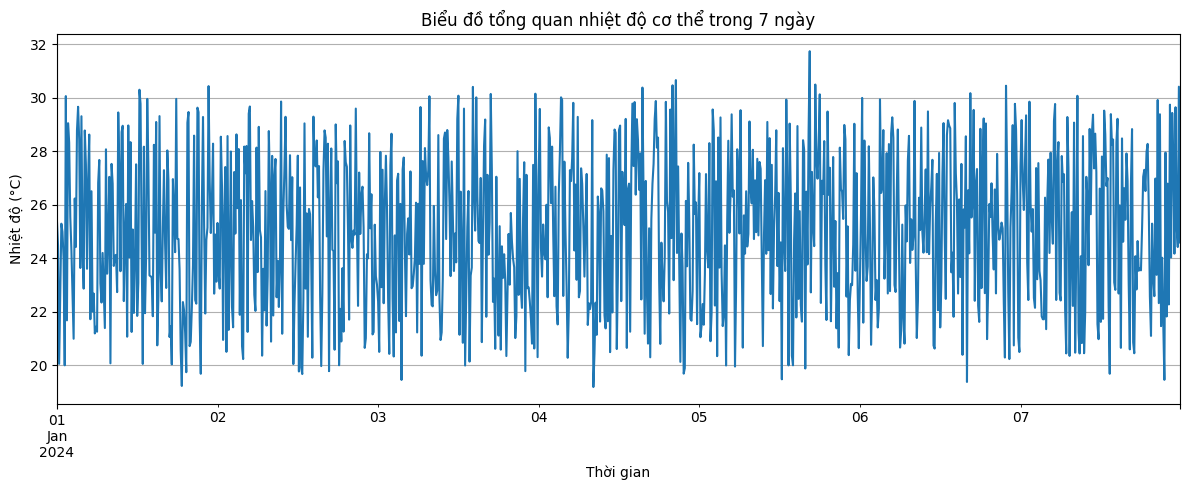
Đánh giá dấu hiệu rối loạn nhịp sinh học

1. **Dữ liệu và tiền xử lý dữ liệu**

Nguồn dữ liệu: <https://www.kaggle.com/datasets/aakashjoshi123/tempsense-temperature-and-sensor-data>

Xử lý thời gian: Tạo trục thời gian đều với khoảng cách 10 phút cho mỗi dòng

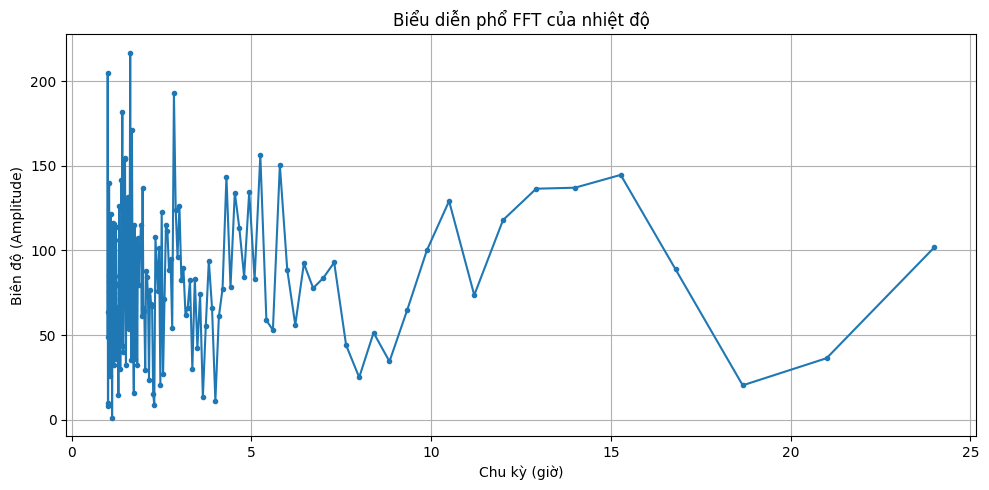
Biểu đồ tổng quát:

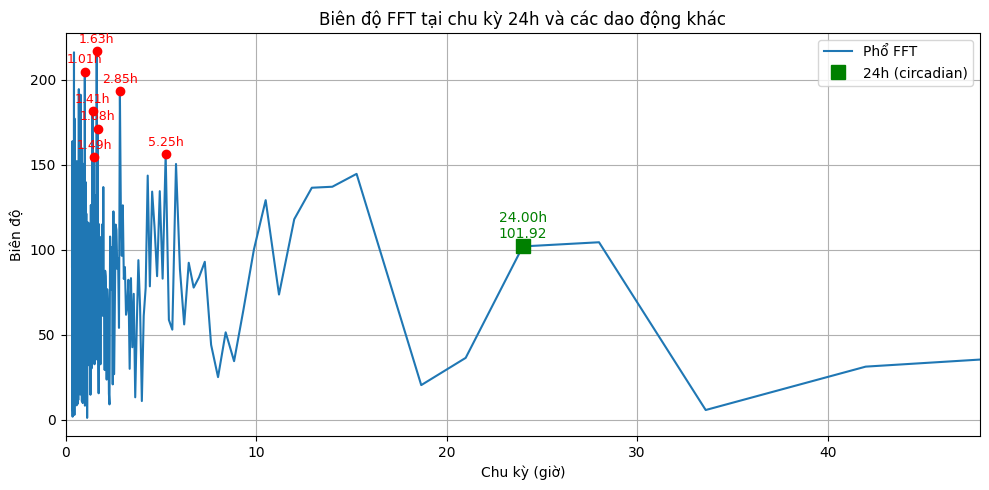


Biểu đồ tổng quan cho thấy nhiệt độ cơ thể biến động liên tục trong 7 ngày. Dao động từ 20-32 độ C, tuy nhiên đa số các giá trị tập trung trong khoảng 22-28 độ C. Nhìn vào biểu đồ, ta thấy xuất hiện đỉnh và đáy khá dày, xen kẽ, thể hiện sự biến động ngắn hạn nhiều hơn là xu hướng nhịp ngày-đêm rõ ràng. Biểu đồ tổng quan cho thấy tín hiệu nhiệt độ có tính biến động cao, chưa phản ánh rõ nhịp sinh học ngày-đêm, nhiều khả năng bị chi phối bởi dao động ngắn hạn (ultradian).

1. **Phân tích chu kỳ 24h và dao động utradian**

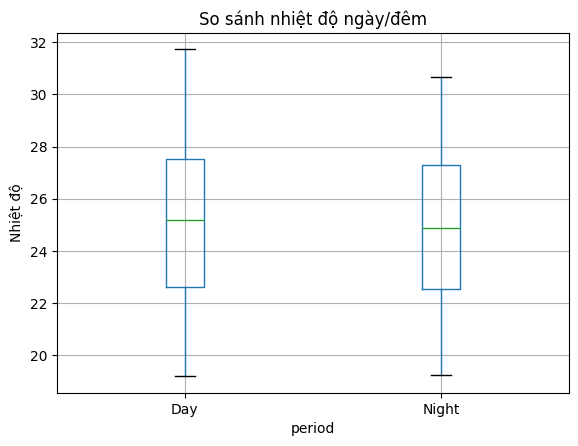
Sử dụng biến đổi Fourier rời rạc (FFT) để chuyển đổi tín hiệu nhiệt độ từ miền thời gian sang miền tần số





Nhìn vào biểu đồ, ta có thể thấy chu kỳ mạnh nhất ~1.63 giờ (ultradian). Biên độ 24h (circadian) ~101.92, yếu hơn Ultradian nhiều lần. Biên độ tại chu kỳ đó = 216.73, suy ra đây là một dao động Ultradian. Dữ liệu nhiệt độ của bộ mẫu không phản ánh được rõ rệt đặc trưng của nhịp sinh học ngày–đêm. Sự thiếu vắng của một đỉnh lớn tại chu kỳ 24h chứng tỏ rằng nhịp sinh học ngày–đêm trong bộ dữ liệu này không phải là dao động chi phối, hoặc thậm chí có thể bị che lấp bởi các dao động ngắn hạn.Thay vào đó, các dao động ultradian chiếm ưu thế, cho thấy tín hiệu bị ảnh hưởng mạnh bởi các nhịp điệu ngắn hơn 24h hoặc các yếu tố ngẫu nhiên từ quá trình mô phỏng dữ liệu. Điều này làm giảm khả năng sử dụng dữ liệu này để phân tích hoặc nghiên cứu chuyên sâu về đồng hồ sinh học circadian thực sự của con người.

1. **So sánh nhiệt độ ngày và đêm**



Biểu đồ boxplot cho thấy dải phân bố nhiệt độ giữa hai nhóm có mức độ giao thoa cao, tức là giá trị nhiệt độ ngày và đêm khá giống nhau, không hình thành hai quần thể rõ rệt. Khoảng cách giữa giá trị trung vị của hai nhóm chỉ khoảng 0.25°C – một mức chênh lệch thấp hơn nhiều so với đặc điểm nhịp sinh học cơ bản (thông thường nhiệt độ ban ngày sẽ cao hơn đêm khoảng 0.5–1°C).

Điều này cho thấy, tín hiệu nhiệt độ trong dữ liệu không thể hiện được rõ rệt đặc trưng sinh lý bình thường về sự khác biệt giữa ngày và đêm. Việc này cũng đồng thuận với kết quả phân tích FFT, khi biên độ chu kỳ 24h thấp hơn nhiều so với các nhịp ngắn hơn 24h (ultradian). Khả năng lớn là dữ liệu bị chi phối bởi nhiễu hoặc đặc tính mô phỏng không phản ánh đúng sinh lý học tự nhiên.

1. **Đánh giá nhịp sinh học**

Đánh giá rối loạn nhịp sinh học dựa trên hai tiêu chí chính:

Độ nổi bật của nhịp 24h trên phổ FFT (tức là xem có đỉnh lớn tại chu kỳ 24h hay không).

Chênh lệch rõ ràng giữa nhiệt độ ban ngày và ban đêm.

Trong phân tích này, cả hai tiêu chí đều không đạt:

* Biên độ nhịp 24h (circadian) thấp hơn nhiều so với các dao động ultradian. Trên phổ FFT, đỉnh tại chu kỳ 24h chỉ ở mức 101.92, trong khi các đỉnh ultradian vượt mức này gấp đôi.
* Nhiệt độ trung bình ngày và đêm gần như không khác biệt, không tạo ra sự “lên xuống” chu kỳ ngày–đêm đặc trưng.

Những kết quả này phản ánh hai khả năng:

* Nhịp sinh học ngày–đêm thực sự không tồn tại rõ rệt trong tín hiệu (có thể do nguồn dữ liệu mô phỏng, hoặc người/thiết bị đo không có nhịp circadian).
* Hoặc dữ liệu bị chi phối quá mạnh bởi các dao động ngắn hạn, dẫn tới việc đỉnh 24h bị che lấp.

1. **Kết luận**

**Dữ liệu không thể hiện rõ nhịp sinh học ngày–đêm** cả về mặt phổ tần số lẫn về giá trị nhiệt độ trung bình ngày – đêm. Điều này thể hiện ở việc biên độ chu kỳ 24h thấp và sự khác biệt nhiệt độ giữa ngày – đêm nhỏ.

**Các dao động ultradian (ngắn hơn 24h) chiếm ưu thế,** cho thấy tín hiệu bị ảnh hưởng chủ yếu bởi các nhịp điệu ngắn hạn hoặc có thể do nhiễu sinh ra từ quá trình mô phỏng dữ liệu.

**Báo cáo này không khuyến nghị dùng bộ dữ liệu trên cho mục tiêu nghiên cứu sâu về nhịp sinh học circadian ở người**, do đặc điểm tín hiệu không phù hợp. Nếu muốn đánh giá hoặc phát hiện rối loạn nhịp sinh học, nên thu thập dữ liệu thực từ người thật với điều kiện sinh hoạt ngày–đêm ổn định.

Tuy nhiên, kết quả cũng cho thấy giá trị ứng dụng của các phương pháp phân tích chu kỳ thời gian và phổ FFT/Wavelet trong đánh giá tín hiệu sinh học. Những công cụ này rất hữu ích để nhận diện các dao động ẩn trong dữ liệu time series, kể cả khi đặc tính tín hiệu không chuẩn như lý thuyết.