TEMPLATE HƯỚNG DẪN TRÌNH BÀY ĐỒ ÁN MÔN HỌC

(Lưu ý: Sinh viên cần kiểm tra và tuân thủ các quy định cụ thể về định dạng và cấu trúc của Khoa/Bộ môn nếu có)

(Trang Bìa Chính)

[TÊN TRƯỜNG ĐẠI HỌC/CAO ĐẲNG]

[TÊN KHOA/VIỆN]

[TÊN BỘ MÔN (Nếu có)]

--- --- ---

[BIỂU TRƯNG CỦA TRƯỜNG/KHOA (Logo)]

ĐỒ ÁN MÔN HỌC

[TÊN MÔN HỌC]

ĐỀ TÀI:

[TÊN ĐẦY ĐỦ CỦA ĐỀ TÀI ĐỒ ÁN]

Sinh viên thực hiện: [Họ và Tên]

Mã số sinh viên: [Mã số SV]

Lớp: [Tên lớp]

Khóa: [Khóa học]

Giảng viên hướng dẫn: [Học hàm, Học vị. Họ và Tên]

[Thành phố], [Tháng] năm [Năm]

(Trang Bìa Phụ - Tương tự bìa chính, thường không có logo)

(Trang Nhiệm vụ đồ án - Nếu có yêu cầu)

(Thường là mẫu do Khoa/Bộ môn cung cấp, có chữ ký xác nhận của GVHD và Trưởng Bộ môn/Khoa)

LỜI CẢM ƠN

* Nội dung: Trình bày ngắn gọn, chân thành lời cảm ơn đến những cá nhân, tập thể đã hỗ trợ bạn hoàn thành đồ án.
  + Giảng viên hướng dẫn (cảm ơn sự chỉ bảo, định hướng...).
  + Các thầy cô khác trong Khoa/Bộ môn (nếu có hỗ trợ kiến thức, góp ý...).
  + Gia đình, bạn bè (nguồn động viên tinh thần...).
  + Cơ quan, tổ chức (nếu có hỗ trợ cung cấp số liệu, tạo điều kiện thực tập/khảo sát...).
* Định dạng: Trang riêng, không đánh số trang hoặc đánh số i.

TÓM TẮT / TRÍCH YẾU (ABSTRACT)

* Nội dung: Trình bày cực kỳ cô đọng về toàn bộ đồ án (thường giới hạn khoảng 200-300 từ). Bao gồm:
  + Bối cảnh/Lý do thực hiện đề tài (1-2 câu).
  + Mục tiêu chính của đề tài.
  + Phương pháp/Cách tiếp cận chính đã sử dụng.
  + Kết quả nổi bật nhất đạt được.
  + Kết luận quan trọng nhất hoặc ý nghĩa chính của kết quả.
* Yêu cầu: Viết rõ ràng, súc tích, đủ ý. Đọc xong phần này người đọc phải hình dung được tổng quan đề tài và kết quả chính.
* Lưu ý: Nhiều nơi yêu cầu có cả bản Tóm tắt tiếng Việt và tiếng Anh (Abstract).
* Định dạng: Trang riêng, đánh số trang i, ii,...

(ABSTRACT - Bản tiếng Anh của Tóm tắt)

* Nội dung: Dịch chính xác và tự nhiên nội dung của phần Tóm tắt tiếng Việt.
* Định dạng: Trang riêng, tiếp tục đánh số trang i, ii,...

MỤC LỤC

* Nội dung: Liệt kê tất cả các đề mục lớn, nhỏ trong đồ án kèm theo số trang tương ứng. Bao gồm cả các phần phụ lục (nếu có).
* Cách tạo: Sử dụng chức năng tạo Mục lục tự động của Google Docs (Insert -> Table of contents). Để làm được điều này, bạn *phải* sử dụng các định dạng Heading (Tiêu đề 1, Tiêu đề 2, Tiêu đề 3...) cho các chương, mục trong nội dung chính.
* Định dạng: Trang riêng, tiếp tục đánh số trang i, ii,...

DANH MỤC HÌNH ẢNH (Hoặc DANH MỤC HÌNH)

* Nội dung: Liệt kê tên tất cả các hình ảnh, biểu đồ, sơ đồ... có trong đồ án theo thứ tự xuất hiện, kèm số trang.
* Cách tạo: Sử dụng chức năng "Insert caption" (Chèn chú thích) cho hình ảnh và sau đó tạo danh mục tự động (Insert -> Table of contents -> Table of figures).
* Định dạng: Trang riêng, tiếp tục đánh số trang i, ii,... (Nếu Mục lục quá dài thì phần này có thể bắt đầu trang mới). Tên hình cần rõ ràng, phản ánh nội dung hình.

DANH MỤC BẢNG BIỂU (Hoặc DANH MỤC BẢNG)

* Nội dung: Liệt kê tên tất cả các bảng biểu có trong đồ án theo thứ tự xuất hiện, kèm số trang.
* Cách tạo: Tương tự như Danh mục hình ảnh, sử dụng "Insert caption" cho bảng và tạo danh mục tự động.
* Định dạng: Trang riêng, tiếp tục đánh số trang i, ii,... Tên bảng cần rõ ràng, phản ánh nội dung bảng.

DANH MỤC TỪ VIẾT TẮT VÀ KÝ HIỆU (Nếu cần)

* Nội dung: Liệt kê các từ viết tắt, các ký hiệu chuyên ngành được sử dụng trong đồ án kèm theo diễn giải đầy đủ. Sắp xếp theo thứ tự Alphabet.
* Mục đích: Giúp người đọc dễ dàng tra cứu khi gặp các từ viết tắt hoặc ký hiệu lạ.
* Ví dụ:
  + CNTT: Công nghệ Thông tin
  + CSDL: Cơ sở dữ liệu
  + GVHD: Giảng viên hướng dẫn
  + TP.HCM: Thành phố Hồ Chí Minh
* Định dạng: Trang riêng, tiếp tục đánh số trang i, ii,...

(Bắt đầu nội dung chính - Đánh số trang 1, 2, 3...)

MỞ ĐẦU (Hoặc CHƯƠNG 1: GIỚI THIỆU)

* 1.1. Lý do chọn đề tài / Tính cấp thiết của đề tài:
  + Nêu bối cảnh chung (thực trạng, vấn đề...).
  + Chỉ ra tại sao vấn đề này/lĩnh vực này cần được nghiên cứu/giải quyết.
  + Khẳng định tầm quan trọng và ý nghĩa (khoa học, thực tiễn) của việc thực hiện đề tài.
* 1.2. Mục tiêu nghiên cứu:
  + Liệt kê các mục tiêu cụ thể mà đồ án cần đạt được (rõ ràng, đo lường được). Thường bắt đầu bằng động từ: *Nghiên cứu..., Phân tích..., Thiết kế..., Xây dựng..., Đánh giá...*
* 1.3. Đối tượng và phạm vi nghiên cứu:
  + Đối tượng: Cái gì được nghiên cứu/khảo sát/thiết kế? (Ví dụ: quy trình X, hệ thống Y, hành vi Z của nhóm đối tượng A...).
  + Phạm vi: Giới hạn về không gian, thời gian, nội dung nghiên cứu (Ví dụ: Nghiên cứu tại công ty A, trong giai đoạn 2023-2024, tập trung vào khía cạnh B...).
* 1.4. Phương pháp nghiên cứu:
  + Nêu tóm tắt các phương pháp chính sẽ được sử dụng để đạt mục tiêu (Ví dụ: Thu thập dữ liệu, phân tích tài liệu, khảo sát, phỏng vấn, thực nghiệm, mô hình hóa, thiết kế hệ thống, lập trình kiểm thử...). Sẽ trình bày chi tiết hơn ở chương sau.
* 1.5. Ý nghĩa khoa học và thực tiễn của đề tài (Có thể gộp vào Lý do chọn đề tài):
  + Đóng góp về mặt lý thuyết (nếu có).
  + Khả năng ứng dụng vào thực tế, giải quyết vấn đề gì.
* 1.6. Cấu trúc của đồ án:
  + Nêu rõ đồ án gồm mấy chương, tóm tắt ngắn gọn nội dung chính của từng chương. (Ví dụ: Ngoài phần Mở đầu, Kết luận, Tài liệu tham khảo, đồ án gồm 3 chương chính: Chương 2 trình bày cơ sở lý thuyết..., Chương 3 trình bày phương pháp/thiết kế..., Chương 4 trình bày kết quả và thảo luận...).

CHƯƠNG 2: TỔNG QUAN TÌNH HÌNH NGHIÊN CỨU / CƠ SỞ LÝ THUYẾT

* 2.1. Tổng quan về vấn đề nghiên cứu:
  + Trình bày các khái niệm cốt lõi liên quan đến đề tài.
  + Phân tích các công trình nghiên cứu, các giải pháp đã có liên quan đến đề tài (trong và ngoài nước).
  + Nêu ưu điểm, nhược điểm của các giải pháp đó. (MQTT) ,
  + Chỉ ra "khoảng trống" nghiên cứu hoặc những điểm mà đề tài này sẽ tập trung cải tiến, phát triển.
  + QUAN TRỌNG: Phải trích dẫn nguồn đầy đủ cho mọi thông tin tham khảo [1], [2]...
* 2.2. Cơ sở lý thuyết liên quan:
  + Trình bày các lý thuyết, nguyên lý, công thức, mô hình... làm nền tảng cho việc thực hiện đề tài.
  + Ví dụ: Nếu làm về web, trình bày về HTML, CSS, JS, Framework... Nếu làm về kinh tế, trình bày các mô hình kinh tế lượng, lý thuyết marketing...
* 2.3. Cơ sở pháp lý (Nếu có):
  + Liệt kê các văn bản quy phạm pháp luật, quy định, tiêu chuẩn... liên quan (nếu đề tài thuộc lĩnh vực cần tuân thủ pháp luật).

CHƯƠNG 3: PHƯƠNG PHÁP NGHIÊN CỨU / THIẾT KẾ VÀ TRIỂN KHAI HỆ THỐNG

*(Tên chương và nội dung cụ thể sẽ thay đổi tùy thuộc vào bản chất đề tài: nghiên cứu lý thuyết, nghiên cứu thực nghiệm, khảo sát, hay thiết kế, xây dựng sản phẩm/hệ thống)*

* Nếu là đề tài nghiên cứu/khảo sát:
  + 3.1. Thiết kế nghiên cứu: Mô tả quy trình thực hiện nghiên cứu (các bước).
  + 3.2. Phương pháp thu thập dữ liệu:
    - Chọn mẫu (cách chọn, kích thước mẫu).
    - Công cụ thu thập (bảng hỏi khảo sát, phiếu phỏng vấn, thiết bị đo lường...).
    - Quy trình thu thập.
  + 3.3. Phương pháp phân tích dữ liệu:
    - Các kỹ thuật, công cụ thống kê, phần mềm sẽ sử dụng để xử lý và phân tích dữ liệu thu thập được.
* Nếu là đề tài thiết kế/xây dựng hệ thống/sản phẩm:
  + 3.1. Phân tích yêu cầu:
    - Yêu cầu chức năng (Hệ thống làm được gì?).
    - Yêu cầu phi chức năng (Hiệu năng, bảo mật, dễ sử dụng...).
  + 3.2. Thiết kế hệ thống:
    - Thiết kế kiến trúc (Mô hình tổng thể, các thành phần chính và mối liên hệ).
    - Thiết kế cơ sở dữ liệu (Sơ đồ ERD, thiết kế các bảng...).
    - Thiết kế giao diện (Wireframe, Mockup...). (NOTE RED)
    - Thiết kế thuật toán (Lưu đồ, mã giả...).
  + 3.3. Công nghệ và công cụ sử dụng:
    - Liệt kê ngôn ngữ lập trình, framework, thư viện, CSDL, phần mềm... đã dùng.
  + 3.4. Triển khai (Implementation):
    - Mô tả quá trình xây dựng, lập trình các module chính. Có thể có một vài đoạn code minh họa (nhưng không đưa toàn bộ code vào đây).
  + 3.5. Kiểm thử (Testing):
    - Mô tả cách thức kiểm thử các chức năng, hiệu năng... (Unit test, Integration test, User acceptance test...).

CHƯƠNG 4: KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN

* 4.1. Trình bày kết quả:
  + Trình bày các kết quả thu được một cách rõ ràng, logic (từ khảo sát, thực nghiệm, sản phẩm chạy thử...).
  + Sử dụng bảng biểu, hình ảnh, đồ thị để minh họa. Các bảng biểu, hình ảnh phải được đánh số thứ tự, có tên/chú thích rõ ràng và được nhắc đến trong phần lời văn.
  + Ví dụ: "Kết quả khảo sát được thể hiện ở Bảng 4.1", "Hình 4.2 mô tả giao diện đăng nhập của hệ thống"...
  + Chỉ trình bày kết quả, *chưa* phân tích sâu hoặc bình luận ở mục này.
* 4.2. Thảo luận kết quả:
  + Phân tích: Giải thích ý nghĩa của các kết quả. Chúng nói lên điều gì?
  + So sánh: Đối chiếu kết quả đạt được với mục tiêu đề ra ban đầu. So sánh với các nghiên cứu trước đó (đã nêu ở Chương 2). Có điểm gì tương đồng, khác biệt? Tại sao?
  + Bàn luận: Nêu lên những điểm nổi bật, đóng góp mới của đề tài (nếu có). Chỉ ra các hạn chế của nghiên cứu/sản phẩm và giải thích nguyên nhân (khách quan, chủ quan).
  + Ý nghĩa: Khẳng định lại ý nghĩa thực tiễn hoặc hàm ý chính sách (nếu có) từ kết quả.

KẾT LUẬN VÀ KIẾN NGHỊ (Hoặc HƯỚNG PHÁT TRIỂN)

* 5.1. Kết luận:
  + Tóm tắt lại những kết quả chính, quan trọng nhất đã đạt được của đồ án.
  + Đối chiếu với các mục tiêu đã đề ra ở phần Mở đầu để khẳng định mức độ hoàn thành.
  + Nhấn mạnh lại đóng góp chính của đề tài (nếu có).
* 5.2. Hạn chế của đề tài (Có thể đưa vào phần Thảo luận):
  + Tóm tắt ngắn gọn lại những hạn chế còn tồn tại.
* 5.3. Kiến nghị / Hướng phát triển tương lai:
  + Đề xuất các giải pháp, kiến nghị (nếu đề tài mang tính ứng dụng, giải quyết vấn đề thực tế).
  + Nêu ra các hướng nghiên cứu, hướng phát triển tiếp theo cho đề tài (cải tiến sản phẩm, mở rộng phạm vi nghiên cứu, nghiên cứu sâu hơn về một khía cạnh...).

TÀI LIỆU THAM KHẢO

* Nội dung: Liệt kê TẤT CẢ các nguồn tài liệu (sách, bài báo khoa học, luận văn, website,...) mà bạn đã đọc, trích dẫn hoặc sử dụng ý tưởng trong quá trình làm đồ án.
* Định dạng:
  + Phải nhất quán theo MỘT chuẩn trích dẫn quy định (ví dụ: APA, IEEE, MLA...). *Hỏi GVHD hoặc Khoa/Bộ môn về chuẩn bắt buộc.*
  + Sắp xếp danh mục theo thứ tự Alphabet của Tên tác giả hoặc theo thứ tự xuất hiện trong bài [1], [2]... (tùy theo chuẩn).
  + Thông tin mỗi tài liệu phải đầy đủ (Tác giả, Năm xuất bản, Tên tài liệu, Nơi xuất bản/Tên tạp chí/URL...).
* Ví dụ (Theo chuẩn APA):
  + Sách: Nguyễn Văn A. (2023). *Tên cuốn sách*. Nhà xuất bản XYZ.
  + Bài báo tạp chí: Trần Thị B., & Lê Văn C. (2022). Tên bài báo. *Tên Tạp chí Khoa học*, *Số Tập*(Số), trang đầu-trang cuối. DOI (nếu có).
  + Website: Tên Tổ Chức hoặc Tác giả (Nếu có). (Năm công bố hoặc cập nhật gần nhất). *Tiêu đề trang hoặc bài viết*. Truy cập ngày [Ngày tháng năm bạn truy cập] từ<https://en.wiktionary.org/wiki/%C4%91%E1%BA%A7y_%C4%91%E1%BB%A7>.
* Công cụ hỗ trợ: Sử dụng các công cụ quản lý trích dẫn như Zotero, Mendeley, EndNote hoặc chức năng Citations & Bibliography của Word/Google Docs để quản lý và định dạng tự động, tránh sai sót.

PHỤ LỤC (Nếu có)

* Nội dung: Bao gồm các thông tin bổ sung, minh họa chi tiết nhưng không tiện đưa vào nội dung chính vì làm loãng mạch văn hoặc quá dài.
  + Bảng số liệu thô.
  + Phiếu khảo sát/phỏng vấn mẫu.
  + Các sơ đồ, bản vẽ chi tiết.
  + Một số đoạn code quan trọng (nhưng không phải toàn bộ source code).
  + Kết quả kiểm thử chi tiết.
  + Hình ảnh bổ sung.
* Định dạng: Đánh số Phụ lục A, Phụ lục B... Mỗi phụ lục nên có tiêu đề rõ ràng. Số trang có thể tiếp tục theo nội dung chính hoặc đánh số lại (PL1, PL2...).

HƯỚNG DẪN ĐỊNH DẠNG CHUNG (Tham khảo - Cần kiểm tra quy định cụ thể)

1. Font chữ: Thường là Times New Roman. Cỡ chữ 12 hoặc 13 cho nội dung chính.
2. Dãn dòng (Line spacing): 1.5 lines.
3. Căn lề (Margins):
   * Lề trên (Top): 2 - 2.5 cm
   * Lề dưới (Bottom): 2 - 2.5 cm
   * Lề trái (Left): 3 - 3.5 cm (để đóng gáy)
   * Lề phải (Right): 2 cm
4. Đánh số trang:
   * Các trang phụ (Lời cảm ơn, Mục lục, Danh mục...) đánh số La Mã thường (i, ii, iii...).
   * Nội dung chính (từ Mở đầu đến hết Tài liệu tham khảo/Phụ lục) đánh số Ả Rập (1, 2, 3...).
   * Vị trí: Thường ở giữa, phía dưới trang (Footer).
5. Định dạng Tiêu đề (Headings):
   * Sử dụng nhất quán các cấp độ Tiêu đề (Heading 1 cho Tên Chương, Heading 2 cho Mục 1.1, 1.2..., Heading 3 cho Mục 1.1.1, 1.1.2...).
   * Ví dụ:
     + CHƯƠNG 1: GIỚI THIỆU (Heading 1, Cỡ chữ 14-16, IN HOA, Đậm)
     + 1.1. Lý do chọn đề tài (Heading 2, Cỡ chữ 13-14, Đậm)
     + *1.1.1. Bối cảnh thực tế* (Heading 3, Cỡ chữ 12-13, Đậm, Nghiêng)
6. Hình ảnh và Bảng biểu:
   * Phải được đánh số thứ tự theo Chương (Ví dụ: Hình 2.1, Bảng 3.2).
   * Tên hình đặt ở *dưới* hình, căn giữa.
   * Tên bảng đặt ở *trên* bảng, căn trái hoặc giữa.
   * Phải có nguồn trích dẫn rõ ràng nếu lấy từ tài liệu khác.
   * Phải được nhắc đến (tham chiếu) trong nội dung bài viết.
7. Trích dẫn trong bài:
   * Khi tham khảo ý tưởng, số liệu, nhận định... của người khác, phải có trích dẫn ngay tại vị trí đó.
   * Định dạng trích dẫn trong bài phải tương ứng với chuẩn đã chọn ở phần Tài liệu tham khảo (Ví dụ: [1], [2] hoặc (Tác giả, Năm)).
8. Ngôn ngữ: Sử dụng văn phong khoa học, trang trọng, rõ ràng, mạch lạc, không mắc lỗi chính tả, ngữ pháp. Hạn chế dùng từ ngữ nói, từ địa phương, viết tắt tùy tiện (chỉ dùng các từ viết tắt đã được liệt kê trong Danh mục).
9. Kiểm tra đạo văn: Đảm bảo tính trung thực khoa học, tuyệt đối không sao chép nếu không có trích dẫn phù hợp. Sử dụng các công cụ kiểm tra đạo văn nếu cần thiết hoặc được yêu cầu.

LƯU Ý QUAN TRỌNG:

* Luôn tham khảo quy định cụ thể của Khoa/Bộ môn: Mỗi nơi có thể có những yêu cầu riêng về cấu trúc, định dạng, chuẩn trích dẫn. Template này chỉ mang tính gợi ý chung.
* Trao đổi thường xuyên với GVHD: Giảng viên hướng dẫn là người nắm rõ nhất yêu cầu và sẽ đưa ra những góp ý quan trọng.
* Lập dàn ý chi tiết trước khi viết: Giúp cấu trúc bài logic, tránh lan man, lạc đề.
* Viết nháp và chỉnh sửa nhiều lần: Đọc lại, sửa lỗi chính tả, ngữ pháp, diễn đạt, logic. Có thể nhờ bạn bè hoặc người có kinh nghiệm đọc góp ý.
* Sử dụng hiệu quả Google Docs: Tận dụng tính năng Styles (Định dạng Tiêu đề), Table of Contents (Mục lục tự động), Comments (Nhận xét), Revision History (Lịch sử chỉnh sửa).
* Sao lưu thường xuyên: Tránh mất dữ liệu do sự cố.

Tuyệt vời! Dưới đây là một dàn ý chi tiết cho đề tài "Phân loại Giai đoạn Giấc ngủ sử dụng Mạng Cảm biến Quang phổ Đồ (PPG)", phù hợp cho một đề cương nghiên cứu hoặc luận văn:

**MẪU ĐỀ TÀI NGHIÊN CỨU**

**1. Tên đề tài:** Phân loại Giai đoạn Giấc ngủ sử dụng Mạng Cảm biến Quang phổ Đồ (PPG)

**(Tên tiếng Anh:** Sleep Stage Classification using Photoplethysmography (PPG) Sensor Network)

**2. Đặt vấn đề/Tổng quan (Introduction/Background)**

* **Tầm quan trọng của giấc ngủ:** Giấc ngủ đóng vai trò thiết yếu đối với sức khỏe thể chất và tinh thần, ảnh hưởng đến trí nhớ, khả năng học tập, tâm trạng, và chức năng miễn dịch. Các giai đoạn giấc ngủ khác nhau (Thức - Wake, Ngủ nông - Light NREM, Ngủ sâu - Deep NREM, Ngủ REM) có chức năng sinh lý riêng biệt.
* **Phương pháp tiêu chuẩn vàng:** Đa ký giấc ngủ (Polysomnography - PSG) là tiêu chuẩn vàng để phân loại giai đoạn giấc ngủ, nhưng đòi hỏi thiết bị phức tạp, đắt tiền, cần thực hiện trong môi trường phòng thí nghiệm/bệnh viện, gây bất tiện và không phù hợp cho việc theo dõi liên tục tại nhà.
* **Nhu cầu theo dõi tại nhà:** Ngày càng có nhu cầu lớn về các giải pháp theo dõi giấc ngủ tiện lợi, không xâm lấn, chi phí thấp và có thể sử dụng tại nhà để phát hiện sớm các rối loạn giấc ngủ và theo dõi chất lượng giấc ngủ hàng ngày.
* **Tiềm năng của cảm biến PPG:** Cảm biến Quang phổ đồ (PPG) là công nghệ phổ biến trong các thiết bị đeo (wearables), đo sự thay đổi thể tích máu qua da. Tín hiệu PPG chứa thông tin về nhịp tim (HR), biến thiên nhịp tim (HRV), và các đặc điểm huyết động khác phản ánh hoạt động của hệ thần kinh tự chủ (ANS) – vốn thay đổi rõ rệt qua các giai đoạn giấc ngủ.
* **Ưu điểm của mạng cảm biến:** Việc sử dụng *mạng* cảm biến PPG (có thể là nhiều cảm biến trên một người hoặc tích hợp dữ liệu từ nhiều thiết bị đơn giản) có tiềm năng cải thiện chất lượng tín hiệu (giảm nhiễu do chuyển động), tăng độ tin cậy và có thể thu thập thông tin phong phú hơn so với một cảm biến đơn lẻ.

**3. Lý do chọn đề tài/Tính cấp thiết (Problem Statement/Motivation)**

* Hạn chế của PSG (chi phí, sự bất tiện, khả năng tiếp cận).
* Các thiết bị theo dõi giấc ngủ thương mại hiện tại dựa trên cảm biến đeo (như actigraphy hoặc PPG đơn giản) thường chỉ phân loại được Thức/Ngủ hoặc đưa ra các ước tính giai đoạn giấc ngủ với độ chính xác chưa cao, đặc biệt là phân biệt các giai đoạn NREM và REM.
* Tín hiệu PPG nhạy cảm với nhiễu do chuyển động và chất lượng tín hiệu có thể thay đổi tùy vị trí đo và người dùng.
* Việc trích xuất các đặc trưng (features) đáng tin cậy từ PPG và xây dựng mô hình học máy hiệu quả để phân loại chính xác 4 hoặc 5 giai đoạn giấc ngủ (Wake, N1, N2, N3, REM) là một thách thức khoa học và kỹ thuật.
* Nghiên cứu khả năng kết hợp dữ liệu từ nhiều điểm đo PPG (mạng cảm biến) để tăng độ chính xác và độ bền vững (robustness) của hệ thống phân loại là một hướng đi có tiềm năng.

**4. Câu hỏi nghiên cứu/Mục tiêu đề tài (Research Questions/Objectives)**

* **Câu hỏi nghiên cứu:**
  + Tín hiệu PPG thu thập từ mạng cảm biến (hoặc một cảm biến trong ngữ cảnh mạng) có đủ thông tin để phân biệt các giai đoạn giấc ngủ chính (Wake, Light NREM, Deep NREM, REM) với độ chính xác chấp nhận được không?
  + Những đặc trưng nào (miền thời gian, miền tần số, phi tuyến) trích xuất từ tín hiệu PPG là quan trọng nhất cho việc phân loại giai đoạn giấc ngủ?
  + Mô hình học máy/học sâu nào (ví dụ: SVM, Random Forest, CNN, LSTM) phù hợp và hiệu quả nhất cho bài toán này?
  + Làm thế nào để kết hợp/tổng hợp dữ liệu từ nhiều cảm biến PPG (nếu áp dụng) để cải thiện hiệu suất phân loại?
  + Hiệu suất của hệ thống đề xuất so với các phương pháp hiện có (dựa trên actigraphy hoặc PPG khác) và/hoặc so với tiêu chuẩn PSG (nếu có dữ liệu đối chứng) như thế nào?
* **Mục tiêu đề tài:**
  + Xây dựng quy trình tiền xử lý tín hiệu PPG từ mạng cảm biến để loại bỏ nhiễu và chuẩn hóa dữ liệu.
  + Nghiên cứu và trích xuất các bộ đặc trưng tiềm năng từ tín hiệu PPG liên quan đến các giai đoạn giấc ngủ.
  + Thiết kế, huấn luyện và đánh giá các mô hình học máy/học sâu khác nhau để thực hiện phân loại tự động 4 giai đoạn giấc ngủ (Wake, Light NREM, Deep NREM, REM) dựa trên đặc trưng PPG.
  + (Tùy chọn/Nâng cao) Khảo sát các phương pháp tổng hợp dữ liệu từ nhiều cảm biến để nâng cao độ chính xác.
  + Đánh giá độ chính xác và hiệu quả của mô hình đề xuất bằng các độ đo chuẩn (accuracy, F1-score, Cohen's Kappa) trên tập dữ liệu chuẩn (nếu có) hoặc so sánh với dữ liệu được gán nhãn bởi chuyên gia/PSG.

**5. Tổng quan tình hình nghiên cứu (Literature Review)**

* Sinh lý học giấc ngủ và các giai đoạn giấc ngủ theo tiêu chuẩn AASM (American Academy of Sleep Medicine).
* Các phương pháp phân loại giai đoạn giấc ngủ truyền thống (PSG và các tín hiệu EEG, EOG, EMG).
* Các phương pháp thay thế/bổ sung: Actigraphy, Đo nhiệt độ cơ thể, Âm thanh (tiếng ngáy),...
* Các nghiên cứu sử dụng PPG để theo dõi giấc ngủ:
  + Phân tích HRV trong giấc ngủ.
  + Phát hiện chứng ngưng thở khi ngủ (Sleep Apnea).
  + Các nghiên cứu *sơ bộ* về phân loại giai đoạn giấc ngủ bằng PPG (thường chỉ phân biệt Wake/Sleep hoặc 2-3 giai đoạn).
* Các kỹ thuật trích xuất đặc trưng từ tín hiệu sinh lý (time-domain, frequency-domain, entropy, fractal analysis,...).
* Các mô hình học máy và học sâu ứng dụng trong phân loại chuỗi thời gian và tín hiệu sinh lý (SVM, RF, CNN, RNN, LSTM, Transformer,...).
* Các nghiên cứu về mạng cảm biến cơ thể (Body Sensor Networks - BSN) và kỹ thuật tổng hợp dữ liệu (data fusion).

**6. Phương pháp nghiên cứu (Methodology)**

* **Thu thập dữ liệu (Data Acquisition):**
  + Sử dụng bộ dữ liệu công khai có sẵn chứa tín hiệu PPG và nhãn giai đoạn giấc ngủ đồng bộ từ PSG (ví dụ: MESA, SHHS trên PhysioNet, hoặc các bộ dữ liệu khác). Nêu rõ bộ dữ liệu được chọn.
  + *Hoặc:* Thiết kế thực nghiệm để tự thu thập dữ liệu PPG từ các tình nguyện viên bằng các thiết bị đeo cụ thể (nêu rõ thiết bị, vị trí đo, số lượng người tham gia), đồng thời ghi lại tín hiệu tham chiếu (ví dụ: PSG hoặc thiết bị đo giấc ngủ đáng tin cậy khác). Mô tả quy trình đồng bộ hóa dữ liệu.
* **Tiền xử lý (Preprocessing):**
  + Lọc tín hiệu PPG (ví dụ: lọc thông dải để giữ tần số liên quan đến nhịp tim).
  + Phân đoạn tín hiệu thành các epoch (ví dụ: 30 giây) tương ứng với tiêu chuẩn ghi PSG.
  + Phát hiện và xử lý nhiễu (artifact detection and removal/interpolation) do chuyển động hoặc tín hiệu yếu.
* **Trích xuất đặc trưng (Feature Extraction):**
  + **Từ HRV (nếu có thể trích xuất R-R intervals):**
    - Miền thời gian: Mean HR, SDNN, RMSSD, pNN50,...
    - Miền tần số: Power in LF, HF bands, LF/HF ratio.
  + **Trực tiếp từ dạng sóng PPG:**
    - Đặc trưng thống kê: Mean, variance, skewness, kurtosis của tín hiệu trong epoch.
    - Đặc trưng hình thái học: Biên độ, độ rộng xung, thời gian dốc lên/dốc xuống,...
  + **Đặc trưng phi tuyến:** Sample Entropy (SampEn), Approximate Entropy (ApEn), Detrended Fluctuation Analysis (DFA).
* **Xây dựng mô hình (Model Development):**
  + Lựa chọn các mô hình ứng viên:
    - Học máy cổ điển: Support Vector Machine (SVM), Random Forest (RF), Gradient Boosting (XGBoost, LightGBM).
    - Học sâu: Convolutional Neural Networks (1D-CNN), Recurrent Neural Networks (LSTM, GRU), CNN-LSTM hybrid, Transformer-based models.
  + Lựa chọn đặc trưng (Feature selection) nếu cần thiết.
  + Xử lý mất cân bằng lớp (nếu các giai đoạn ngủ có thời lượng rất khác nhau).
* **Huấn luyện và Đánh giá (Training and Evaluation):**
  + Phân chia dữ liệu: Tập huấn luyện (Training), tập kiểm định (Validation), tập kiểm thử (Testing).
  + Chiến lược kiểm định chéo (Cross-validation): Ví dụ: K-fold cross-validation, Leave-One-Subject-Out (LOSO) cross-validation để đánh giá khả năng tổng quát hóa trên người dùng mới.
  + Các độ đo đánh giá: Accuracy tổng thể, Precision, Recall, F1-score cho từng lớp (giai đoạn ngủ), Ma trận nhầm lẫn (Confusion Matrix), Cohen's Kappa score để đo lường sự phù hợp với nhãn tham chiếu.

**7. Kết quả dự kiến/Đóng góp của đề tài (Expected Results/Contributions)**

* Một thuật toán/hệ thống hoàn chỉnh có khả năng phân loại 4 giai đoạn giấc ngủ chính từ tín hiệu PPG với độ chính xác được lượng hóa và chấp nhận được.
* Xác định được tập hợp các đặc trưng PPG hiệu quả nhất cho bài toán phân loại giai đoạn giấc ngủ.
* So sánh hiệu năng của các mô hình học máy/học sâu khác nhau cho nhiệm vụ này.
* Đóng góp vào việc phát triển công nghệ theo dõi giấc ngủ không xâm lấn, chi phí thấp, tiện lợi tại nhà.
* Mở ra hướng nghiên cứu ứng dụng mạng cảm biến PPG để cải thiện chất lượng theo dõi sức khỏe.

Chắc chắn rồi! Dưới đây là một dàn ý mẫu cho đề tài nghiên cứu kết hợp ESP32 với mạng cảm biến tín hiệu PPG, tập trung vào khía cạnh xây dựng hệ thống và ứng dụng thời gian thực.

**MẪU ĐỀ TÀI NGHIÊN CỨU**

**1. Tên đề tài:** Xây dựng Mạng Cảm biến PPG Không dây Công suất thấp dựa trên ESP32 để Theo dõi Sinh lý Thời gian thực

**(Tên tiếng Anh:** Building a Low-Power Wireless PPG Sensor Network based on ESP32 for Real-time Physiological Monitoring)

**2. Đặt vấn đề/Tổng quan (Introduction/Background)**

* **Nhu cầu theo dõi sinh lý liên tục:** Việc giám sát các chỉ số sinh lý như nhịp tim (HR), biến thiên nhịp tim (HRV), độ bão hòa oxy trong máu (SpO2) một cách liên tục và thời gian thực ngày càng quan trọng trong y tế, thể thao, chăm sóc sức khỏe cá nhân và an toàn lao động.
* **Hạn chế của hệ thống hiện tại:** Các hệ thống theo dõi có dây gây bất tiện, hạn chế di chuyển. Các thiết bị đeo đơn lẻ có thể không cung cấp đủ thông tin hoặc độ tin cậy trong mọi tình huống (ví dụ, nhiễu do vận động tại một điểm đo).
* **Tiềm năng của mạng cảm biến không dây (WSN):** WSN cho phép thu thập dữ liệu từ nhiều điểm trên cơ thể hoặc từ nhiều người, tăng khả năng bao phủ, cải thiện độ tin cậy thông qua dự phòng hoặc tổng hợp dữ liệu, và cho phép các ứng dụng giám sát phân tán.
* **Vai trò của PPG:** Cảm biến PPG cung cấp nguồn dữ liệu sinh lý phong phú, không xâm lấn, dễ tích hợp vào thiết bị nhỏ gọn.
* **ESP32 là nền tảng lý tưởng:** Vi điều khiển ESP32 cung cấp giải pháp chi phí thấp, hiệu năng đủ mạnh, tích hợp sẵn Wi-Fi và Bluetooth (BLE), hỗ trợ các giao thức truyền thông công suất thấp (như ESP-NOW), và có các chế độ tiết kiệm năng lượng, rất phù hợp để xây dựng các nút cảm biến (sensor nodes) và cổng nối (gateway) cho WSN trong lĩnh vực y tế/sức khỏe.

**3. Lý do chọn đề tài/Tính cấp thiết (Problem Statement/Motivation)**

* **Thách thức về năng lượng:** Việc duy trì hoạt động liên tục của mạng cảm biến đeo trên người đòi hỏi tối ưu hóa nghiêm ngặt về mặt năng lượng cho từng nút cảm biến để kéo dài tuổi thọ pin.
* **Đồng bộ hóa dữ liệu:** Đảm bảo các nút cảm biến trong mạng hoạt động đồng bộ về mặt thời gian là rất quan trọng đối với một số phân tích (ví dụ: tính toán thời gian truyền sóng mạch - Pulse Transit Time từ nhiều điểm PPG) nhưng là thách thức trong môi trường không dây.
* **Quản lý dữ liệu và băng thông:** Truyền liên tục dữ liệu PPG thô từ nhiều nút có thể gây tắc nghẽn băng thông và tiêu tốn năng lượng. Cần có cơ chế xử lý/nén dữ liệu tại nút (edge processing) trên ESP32.
* **Độ tin cậy và chất lượng tín hiệu:** Đảm bảo việc truyền dữ liệu đáng tin cậy trong môi trường không dây và xử lý các vấn đề về chất lượng tín hiệu PPG (nhiễu vận động) trong một hệ thống phân tán.
* **Tích hợp hệ thống:** Việc thiết kế và tích hợp cả phần cứng (cảm biến, ESP32, nguồn), phần mềm (firmware cho nút và gateway), và giao thức truyền thông thành một hệ thống hoạt động ổn định là một bài toán kỹ thuật phức tạp.

**4. Câu hỏi nghiên cứu/Mục tiêu đề tài (Research Questions/Objectives)**

* **Câu hỏi nghiên cứu:**
  + Làm thế nào để thiết kế một giao thức truyền thông không dây (ví dụ: dựa trên ESP-NOW hoặc BLE) hiệu quả về năng lượng và độ trễ cho mạng cảm biến PPG sử dụng ESP32?
  + Làm thế nào để đạt được sự đồng bộ thời gian giữa các nút ESP32 với độ chính xác đủ cho các phân tích sinh lý dựa trên đa điểm PPG?
  + Những thuật toán tiền xử lý và trích xuất đặc trưng PPG cơ bản nào (ví dụ: phát hiện nhịp, ước tính HR) có thể được thực thi hiệu quả *trên tài nguyên hạn chế của ESP32 tại nút*?
  + Dữ liệu từ mạng cảm biến PPG có thể được tổng hợp, xử lý và hiển thị/phân tích gần thời gian thực như thế nào?
  + Hiệu suất tổng thể của hệ thống (mức tiêu thụ năng lượng, độ trễ dữ liệu, tỷ lệ mất gói, độ chính xác của các tham số sinh lý ước tính) là bao nhiêu?
* **Mục tiêu đề tài:**
  + Thiết kế và chế tạo phần cứng cho các nút cảm biến PPG dựa trên ESP32 và cảm biến PPG phù hợp (ví dụ: MAX30102, MAX86150).
  + Phát triển và triển khai một giao thức truyền thông không dây công suất thấp (ví dụ: ESP-NOW, BLE custom profile) cho việc trao đổi dữ liệu giữa các nút và/hoặc một gateway trung tâm (cũng có thể dùng ESP32).
  + Triển khai cơ chế đồng bộ hóa thời gian giữa các nút.
  + Xây dựng firmware cho các nút ESP32 để: thu thập tín hiệu PPG, thực hiện tiền xử lý cơ bản (lọc nhiễu), trích xuất các đặc trưng quan trọng tại chỗ (nếu có thể), quản lý năng lượng (sử dụng chế độ ngủ sâu/ngủ nhẹ), và truyền dữ liệu.
  + Xây dựng một ứng dụng trung tâm (gateway/máy tính/cloud) để nhận, tổng hợp, lưu trữ, phân tích sâu hơn và hiển thị dữ liệu từ mạng lưới.
  + Đánh giá hiệu năng của hệ thống thông qua các thực nghiệm đo lường (năng lượng tiêu thụ, độ trễ, độ tin cậy truyền) và so sánh các tham số sinh lý ước tính (HR, SpO2 nếu có) với thiết bị đo tham chiếu.

**5. Tổng quan tình hình nghiên cứu (Literature Review)**

* Mạng cảm biến không dây (WSNs): Kiến trúc (star, mesh), giao thức (Zigbee, BLE, LoRa, Wi-Fi HaLow, ESP-NOW), các thách thức (năng lượng, đồng bộ, định tuyến, bảo mật).
* Mạng cảm biến cơ thể không dây (WBSNs - Wireless Body Sensor Networks): Ứng dụng trong y tế, thể thao; các tiêu chuẩn liên quan (ví dụ: IEEE 802.15.6).
* ESP32: Kiến trúc, các ngoại vi, khả năng kết nối không dây (Wi-Fi, Bluetooth Classic/BLE), giao thức ESP-NOW, các chế độ tiết kiệm năng lượng, các framework lập trình (ESP-IDF, Arduino).
* Công nghệ cảm biến PPG: Nguyên lý hoạt động, các loại cảm biến, tín hiệu thu được và các tham số sinh lý có thể trích xuất.
* Kỹ thuật xử lý tín hiệu PPG: Lọc nhiễu (nhiễu vận động), phát hiện đỉnh sóng, tính toán HR, HRV, ước tính SpO2.
* Kỹ thuật đồng bộ hóa thời gian trong WSNs (ví dụ: FTSP, Flooding Time Synchronization Protocol; RBS, Reference Broadcast Synchronization).
* Điện toán biên (Edge Computing) trong WSNs/IoT: Lợi ích của việc xử lý dữ liệu tại nguồn.
* Các hệ thống theo dõi sinh lý dựa trên ESP32 đã có (thường là các hệ thống đơn nút hoặc mạng đơn giản).

**6. Phương pháp nghiên cứu (Methodology)**

* **Thiết kế kiến trúc hệ thống:**
  + Xác định cấu trúc mạng (ví dụ: star - các nút gửi về gateway, mesh - các nút có thể chuyển tiếp cho nhau).
  + Lựa chọn công nghệ truyền thông chính (ví dụ: ESP-NOW cho giao tiếp giữa các ESP32, BLE để gateway gửi dữ liệu lên điện thoại).
  + Xác định vai trò các thành phần: Nút cảm biến (ESP32 + PPG sensor), Gateway (ESP32 nhận dữ liệu, có thể kết nối Wi-Fi/Ethernet/BLE).
* **Thiết kế phần cứng:**
  + Lựa chọn module ESP32 (WROOM, WROVER, S3, C3,... tùy yêu cầu).
  + Lựa chọn cảm biến PPG (độ nhạy, giao tiếp I2C/SPI, mức tiêu thụ năng lượng).
  + Thiết kế mạch nguồn và quản lý pin (Battery Management System - BMS).
  + (Tùy chọn) Thiết kế mạch in (PCB) để tích hợp gọn nhẹ.
* **Phát triển Firmware (Nút cảm biến):**
  + Lập trình driver giao tiếp với cảm biến PPG (sử dụng thư viện có sẵn hoặc tự viết).
  + Thuật toán thu thập và đệm dữ liệu PPG.
  + Triển khai các bộ lọc số (ví dụ: Butterworth, FIR) để loại bỏ nhiễu tần số thấp/cao.
  + (Tùy chọn) Thuật toán phát hiện và giảm thiểu nhiễu vận động.
  + (Tùy chọn) Thuật toán phát hiện đỉnh sóng PPG và tính HR, HRV cơ bản.
  + Triển khai giao thức truyền thông đã chọn (ESP-NOW/BLE).
  + Triển khai thuật toán đồng bộ thời gian.
  + Quản lý trạng thái hoạt động và chế độ ngủ để tiết kiệm năng lượng.
* **Phát triển Firmware/Software (Gateway/Trung tâm):**
  + Nhận và giải mã dữ liệu từ các nút.
  + Kiểm tra tính toàn vẹn, xử lý gói tin lỗi/mất.
  + Đồng bộ và tổng hợp dữ liệu từ các nút khác nhau.
  + Lưu trữ dữ liệu (thẻ nhớ SD, cơ sở dữ liệu thời gian thực như InfluxDB).
  + Thực hiện các phân tích phức tạp hơn (nếu cần).
  + Cung cấp giao diện hiển thị (web server trên ESP32, gửi qua MQTT lên cloud dashboard, ứng dụng di động).
* **Kế hoạch đánh giá:**
  + Đo lường dòng tiêu thụ của nút cảm biến ở các chế độ (active, sleep, transmit) bằng thiết bị chuyên dụng (ví dụ: Power Profiler Kit).
  + Đo độ trễ từ lúc thu thập dữ liệu tại nút đến lúc nhận được tại gateway.
  + Đánh giá tỷ lệ mất gói tin trong các điều kiện khác nhau (khoảng cách, vật cản).
  + Đánh giá độ chính xác của cơ chế đồng bộ thời gian.
  + So sánh giá trị HR, SpO2 (nếu có) ước tính từ hệ thống với thiết bị y tế chuẩn.
  + Kiểm tra độ ổn định hoạt động của toàn hệ thống trong thời gian dài.

**7. Kết quả dự kiến/Đóng góp của đề tài (Expected Results/Contributions)**

* Một nguyên mẫu (prototype) hoạt động của mạng cảm biến PPG không dây, công suất thấp dựa trên ESP32.
* Thiết kế và mã nguồn mở (nếu có thể) cho phần cứng nút cảm biến và gateway.
* Firmware được tối ưu hóa cho nút ESP32 thực hiện thu thập, tiền xử lý và truyền dữ liệu PPG hiệu quả.
* Giao thức truyền thông và đồng bộ hóa phù hợp cho ứng dụng theo dõi sinh lý trên nền tảng ESP32.
* Báo cáo chi tiết về hiệu năng của hệ thống (năng lượng, độ trễ, độ tin cậy, độ chính xác).
* Nền tảng phần cứng/phần mềm có thể mở rộng và tùy biến cho các ứng dụng theo dõi sức khỏe và thể thao khác nhau sử dụng mạng cảm biến.
* Đóng góp vào lĩnh vực thiết kế hệ thống WBSN thực tế, chi phí thấp.

**MẪU ĐỀ TÀI NGHIÊN CỨU**

**1. Tên đề tài:** Ứng dụng Học tăng cường (Reinforcement Learning) dựa trên Tín hiệu PPG để Đề xuất Can thiệp Giảm căng thẳng Thích ứng

**(Tên tiếng Anh:** Applying Reinforcement Learning based on PPG Signals for Adaptive Stress Reduction Intervention Recommendation)

**2. Đặt vấn đề/Tổng quan (Introduction/Background)**

* **Tác động của căng thẳng (stress):** Căng thẳng mãn tính là yếu tố nguy cơ chính cho nhiều vấn đề sức khỏe thể chất và tinh thần (bệnh tim mạch, suy giảm miễn dịch, rối loạn lo âu, trầm cảm).
* **Nhu cầu can thiệp cá nhân hóa:** Các phương pháp quản lý căng thẳng (ví dụ: thở sâu, thiền, vận động nhẹ) cần được áp dụng đúng thời điểm và phù hợp với trạng thái sinh lý cá nhân để đạt hiệu quả tối ưu.
* **PPG như một chỉ dấu sinh học:** Tín hiệu PPG phản ánh hoạt động của hệ thần kinh tự chủ (ANS) thông qua các chỉ số như nhịp tim (HR) và đặc biệt là biến thiên nhịp tim (HRV - Heart Rate Variability). Các tham số HRV (ví dụ: SDNN, RMSSD, tỷ lệ LF/HF) được công nhận là có tương quan với mức độ căng thẳng và trạng thái thư giãn.
* **Hạn chế của phương pháp truyền thống:** Các lịch trình can thiệp cố định hoặc hệ thống dựa trên luật (rule-based) thường không linh hoạt và kém hiệu quả do không thích ứng được với sự biến động tức thời và tính cá thể hóa của phản ứng căng thẳng.
* **Tiềm năng của Học tăng cường (RL):** RL là một lĩnh vực của học máy, nơi một "tác nhân" (agent) học cách đưa ra các quyết định (hành động - actions) tuần tự trong một "môi trường" (environment) để tối đa hóa "phần thưởng" (reward) tích lũy dài hạn. Trong bối cảnh này, RL có thể học một "chính sách" (policy) tối ưu để đề xuất các *chuỗi* can thiệp (actions) dựa trên trạng thái sinh lý hiện tại (state - trích xuất từ PPG) nhằm đạt được mục tiêu dài hạn là giảm căng thẳng hoặc duy trì trạng thái cân bằng (homeostasis).

**3. Lý do chọn đề tài/Tính cấp thiết (Problem Statement/Motivation)**

* **Tính cá thể và động của phản ứng căng thẳng:** Mỗi người phản ứng với căng thẳng và các biện pháp can thiệp một cách khác nhau và thay đổi theo thời gian.
* **Thách thức trong việc tối ưu hóa can thiệp:** Xác định *thời điểm* và *loại hình* can thiệp tối ưu dựa trên trạng thái sinh lý thời gian thực là rất khó khăn.
* **Giới hạn của học có giám sát:** Học có giám sát (Supervised Learning) có thể dự đoán mức độ căng thẳng từ PPG, nhưng không trực tiếp học được một chính sách (policy) về việc *khi nào* và *làm thế nào* để can thiệp.
* **Thách thức khi áp dụng RL:**
  + **Biểu diễn trạng thái (State Representation):** Làm thế nào để biểu diễn hiệu quả trạng thái sinh lý (liên quan đến căng thẳng) từ tín hiệu PPG nhiễu và có độ biến thiên cao?
  + **Thiết kế hàm thưởng (Reward Function Design):** Làm thế nào để định nghĩa một hàm thưởng có ý nghĩa, phản ánh mục tiêu giảm căng thẳng hoặc đạt trạng thái thư giãn, và có thể tính toán được từ dữ liệu PPG?
  + **Độ trễ của tác động:** Hiệu quả của một can thiệp có thể không tức thì, gây khó khăn cho việc quy kết phần thưởng (credit assignment) cho hành động trước đó.
  + **An toàn và ổn định:** Cần đảm bảo chính sách học được không dẫn đến các vòng lặp phản hồi tiêu cực hoặc các đề xuất không an toàn.
  + **Dữ liệu và môi trường:** Việc thu thập đủ dữ liệu tương tác (trạng thái, hành động, phần thưởng, trạng thái tiếp theo) trong thế giới thực là khó khăn. Có thể cần môi trường mô phỏng (simulation) hoặc các kỹ thuật học ngoại tuyến (offline RL).

**4. Câu hỏi nghiên cứu/Mục tiêu đề tài (Research Questions/Objectives)**

* **Câu hỏi nghiên cứu:**
  + Làm thế nào để biểu diễn các trạng thái sinh lý liên quan đến căng thẳng (ví dụ: căng thẳng cao, đang phục hồi, bình tĩnh) thành không gian trạng thái (S) phù hợp cho tác nhân RL, sử dụng các đặc trưng trích xuất từ PPG/HRV?
  + Những hành động (A) nào đại diện cho các can thiệp khả thi (ví dụ: gợi ý thở sâu, nghỉ ngơi ngắn, nghe nhạc thư giãn, không làm gì) cần được xem xét?
  + Hàm thưởng (R) nào có thể được thiết kế để khuyến khích tác nhân RL học các chính sách dẫn đến giảm căng thẳng có thể đo lường được (ví dụ: dựa trên sự thay đổi của các chỉ số HRV)?
  + Thuật toán RL nào (ví dụ: Q-learning, DQN, Policy Gradients, Actor-Critic) phù hợp để học chính sách can thiệp từ dữ liệu PPG?
  + Chính sách học được có khả năng thích ứng đề xuất can thiệp dựa trên phản hồi PPG thời gian thực không? Hiệu suất so với các chiến lược cơ bản (ngẫu nhiên, cố định) như thế nào?
* **Mục tiêu đề tài:**
  + Phát triển phương pháp trích xuất đặc trưng PPG/HRV và ánh xạ chúng thành các biểu diễn trạng thái (rời rạc hoặc liên tục) cho RL.
  + Xác định tập hợp các hành động can thiệp tiềm năng.
  + Thiết kế và triển khai các hàm thưởng phù hợp dựa trên sự thay đổi sinh lý phản ánh qua PPG.
  + Lựa chọn, triển khai và huấn luyện các thuật toán RL phù hợp (có thể bắt đầu bằng dữ liệu ngoại tuyến hoặc môi trường mô phỏng).
  + Đánh giá hiệu quả của chính sách đã học trong môi trường mô phỏng hoặc thông qua nghiên cứu người dùng có kiểm soát (đo lường thay đổi sinh lý và phản hồi người dùng).
  + Phân tích chính sách đã học để hiểu các chiến lược thích ứng mà tác nhân RL khám phá được.

**5. Tổng quan tình hình nghiên cứu (Literature Review)**

* Sinh lý học căng thẳng, hệ thần kinh tự chủ (ANS) và biểu hiện qua HR, HRV.
* Kỹ thuật xử lý tín hiệu PPG để phân tích HRV và đánh giá/phát hiện căng thẳng.
* Các kỹ thuật can thiệp quản lý căng thẳng (mindfulness, biofeedback, kỹ thuật thở,...).
* Nguyên lý cơ bản của Học tăng cường: Markov Decision Processes (MDPs), Value-based methods (Q-learning, DQN), Policy-based methods (REINFORCE, A2C, A3C, PPO), Actor-Critic methods.
* Ứng dụng của RL trong y tế, chăm sóc sức khỏe cá nhân hóa (tối ưu hóa điều trị, hệ thống gợi ý, điều khiển vòng kín).
* Các nghiên cứu (nếu có) sử dụng tín hiệu sinh lý (EEG, ECG, EDA, PPG) kết hợp với RL cho các bài toán tương tự (có thể đây là hướng đi còn mới).
* Thách thức khi áp dụng RL vào các ứng dụng y sinh thực tế (an toàn, thiết kế phần thưởng, tính không dừng của môi trường - non-stationarity, yêu cầu dữ liệu).

**6. Phương pháp nghiên cứu (Methodology)**

* **Nguồn dữ liệu:**
  + **Tùy chọn 1 (Dữ liệu ngoại tuyến):** Sử dụng các bộ dữ liệu hiện có chứa bản ghi PPG/HRV trong các điều kiện căng thẳng khác nhau (ví dụ: WESAD, DEAP nếu có PPG). Có thể cần mô phỏng tác động của can thiệp lên dữ liệu. Ưu điểm: Có sẵn dữ liệu. Nhược điểm: Không có dữ liệu tương tác thực sự với can thiệp. Cần dùng kỹ thuật Offline RL.
  + **Tùy chọn 2 (Thu thập dữ liệu có kiểm soát):** Thiết kế nghiên cứu người dùng nơi tình nguyện viên thực hiện các tác vụ gây căng thẳng, nhận các can thiệp ngẫu nhiên hoặc theo lịch trình, trong khi tín hiệu PPG được ghi lại. Ưu điểm: Có dữ liệu tương tác. Nhược điểm: Tốn kém, phức tạp, vấn đề đạo đức.
  + **Tùy chọn 3 (Môi trường mô phỏng):** Xây dựng một mô hình mô phỏng phản ứng sinh lý (dựa trên PPG/HRV) với căng thẳng và can thiệp, dựa trên các mô hình toán học hoặc học từ dữ liệu thực. Ưu điểm: An toàn, cho phép khám phá nhanh. Nhược điểm: Độ chính xác của mô phỏng là yếu tố then chốt.
* **Biểu diễn Trạng thái (S):**
  + Trích xuất các đặc trưng PPG/HRV (ví dụ: HR, RMSSD, SDNN, LF/HF, SampEn) theo cửa sổ thời gian.
  + Rời rạc hóa không gian đặc trưng (ví dụ: dùng ngưỡng, clustering như K-Means) để tạo trạng thái rời rạc (ví dụ: "Bình tĩnh", "Căng thẳng nhẹ", "Căng thẳng cao").
  + Hoặc sử dụng trực tiếp vector đặc trưng làm trạng thái liên tục (đòi hỏi thuật toán RL phù hợp như DQN, DDPG, PPO).
* **Không gian Hành động (A):** Xác định tập hợp các hành động rời rạc mà tác nhân có thể chọn, ví dụ: A = {Đề\_xuất\_Thở\_sâu, Đề\_xuất\_Nghỉ\_ngơi, Đề\_xuất\_Nghe\_nhạc, Không\_làm\_gì}.
* **Thiết kế Hàm thưởng (R):** Đây là bước quan trọng và cần thử nghiệm. Ví dụ:
  + R = +1 nếu trạng thái chuyển từ "Căng thẳng cao" sang "Căng thẳng nhẹ".
  + R = +2 nếu trạng thái chuyển sang "Bình tĩnh".
  + R = -1 nếu trạng thái chuyển sang "Căng thẳng cao" hơn.
  + R = -0.1 cho mỗi hành động can thiệp được thực hiện (chi phí).
  + Hoặc R = change\_in\_RMSSD (phần thưởng liên tục dựa trên thay đổi HRV).
  + Cần xem xét phần thưởng trễ (delayed rewards).
* **Lựa chọn và Triển khai Thuật toán RL:**
  + Nếu trạng thái/hành động rời rạc và nhỏ: Q-learning, SARSA.
  + Nếu trạng thái liên tục hoặc lớn: DQN, Double DQN, Dueling DQN (value-based); REINFORCE, A2C, PPO (policy-based/actor-critic).
  + Sử dụng các thư viện: Stable Baselines3, TensorFlow Agents, RLlib, PyTorch RL.
* **Quá trình Huấn luyện:**
  + Nếu dùng dữ liệu ngoại tuyến: Áp dụng các kỹ thuật Offline RL (ví dụ: Batch Constrained Q-learning - BCQ, Conservative Q-Learning - CQL).
  + Nếu dùng mô phỏng: Huấn luyện tác nhân tương tác với môi trường mô phỏng.
  + Xác định các siêu tham số: Tỷ lệ học (learning rate), hệ số chiết khấu (discount factor γ), chiến lược thăm dò (exploration, ví dụ: ε-greedy, noise addition).
* **Đánh giá:**
  + **Đánh giá ngoại tuyến (Offline Evaluation):** Sử dụng một tập dữ liệu kiểm thử riêng biệt để đánh giá chính sách đã học (ví dụ: đo lường phần thưởng tích lũy dự kiến, xem hành động được chọn có hợp lý không).
  + **Đánh giá trong mô phỏng:** Chạy tác nhân đã huấn luyện trong môi trường mô phỏng, đo các chỉ số hiệu suất (thời gian đạt trạng thái mục tiêu, mức độ căng thẳng trung bình, tần suất can thiệp).
  + **(Nâng cao/Tương lai) Đánh giá trực tuyến (Online Evaluation):** Triển khai chính sách trong một nghiên cứu người dùng có kiểm soát, so sánh hiệu quả (thay đổi sinh lý thực tế, phản hồi chủ quan của người dùng) với nhóm đối chứng (không can thiệp, can thiệp ngẫu nhiên/cố định). Cần có sự phê duyệt về đạo đức nghiên cứu.

**7. Kết quả dự kiến/Đóng góp của đề tài (Expected Results/Contributions)**

* Một khung làm việc (framework) mới lạ áp dụng RL vào tín hiệu PPG để đưa ra đề xuất can thiệp thích ứng.
* Các phương pháp biểu diễn trạng thái và thiết kế hàm thưởng hiệu quả cho việc hướng dẫn RL trong bối cảnh sinh lý học.
* Một tác nhân/chính sách RL đã được huấn luyện, có khả năng đề xuất can thiệp giảm căng thẳng cá nhân hóa dựa trên PPG.
* Kết quả đánh giá hiệu năng cho thấy tiềm năng/ưu điểm của phương pháp RL so với các phương pháp không thích ứng.
* Những hiểu biết sâu sắc về thách thức và kinh nghiệm thực tiễn khi áp dụng RL vào dữ liệu tín hiệu y sinh thời gian thực.
* Đóng góp vào lĩnh vực y tế số cá nhân hóa và các hệ thống huấn luyện/hỗ trợ sức khỏe tự động.

**MẪU ĐỀ TÀI NGHIÊN CỨU**

**1. Tên đề tài:** Ước tính Huyết áp Liên tục từ Tín hiệu PPG sử dụng Mạng Transformer

**(Tên tiếng Anh:** Continuous Blood Pressure Estimation from PPG Signals using Transformer Networks)

**2. Đặt vấn đề/Tổng quan (Introduction/Background)**

* **Tầm quan trọng của theo dõi huyết áp (HA):** HA là một chỉ số sinh tồn quan trọng. Việc theo dõi HA liên tục, không xâm lấn (cuffless) có ý nghĩa lớn trong việc quản lý tăng huyết áp, đánh giá nguy cơ tim mạch, và theo dõi bệnh nhân tại nhà hoặc trong bệnh viện.
* **Hạn chế của phương pháp đo HA truyền thống:** Đo HA bằng bao hơi (cuff-based) chỉ cung cấp các phép đo rời rạc, gây khó chịu và không phù hợp cho việc theo dõi liên tục.
* **Tiềm năng của PPG:** Tín hiệu PPG, dễ dàng thu thập từ các thiết bị đeo, chứa thông tin phong phú về sóng mạch. Hình thái (morphology) và thời gian lan truyền của sóng mạch (Pulse Wave Velocity - PWV, có thể gián tiếp suy ra từ PPG) có mối tương quan với giá trị HA.
* **Các phương pháp ước tính HA từ PPG hiện có:**
  + Dựa trên đặc trưng thủ công (handcrafted features): Thường dựa vào Thời gian truyền sóng mạch (Pulse Transit Time - PTT, yêu cầu thêm tín hiệu ECG), hoặc các đặc trưng hình thái học của PPG. Các mô hình học máy cổ điển (Linear Regression, SVM, Random Forest) thường được sử dụng.
  + Dựa trên học sâu (Deep Learning): Mạng Nơ-ron Tích chập (CNN) để trích xuất đặc trưng hình thái, Mạng Nơ-ron Hồi quy (RNN/LSTM/GRU) để mô hình hóa sự phụ thuộc thời gian. Tuy nhiên, các mô hình này có thể gặp khó khăn trong việc nắm bắt các mối quan hệ phức tạp, phi tuyến và các phụ thuộc tầm xa (long-range dependencies) trong tín hiệu PPG.
* **Tại sao dùng Transformer?** Kiến trúc Transformer, với cơ chế tự chú ý (self-attention), đã cách mạng hóa lĩnh vực xử lý ngôn ngữ tự nhiên bằng cách nắm bắt hiệu quả các mối quan hệ ngữ cảnh và phụ thuộc tầm xa trong dữ liệu tuần tự. **Giả thuyết:** Transformer có thể mô hình hóa tốt hơn các tương tác phức tạp giữa các điểm/đặc trưng khác nhau *trong cùng một chu kỳ PPG* và *giữa các chu kỳ PPG liên tiếp*, từ đó nắm bắt những thay đổi hình thái tinh tế liên quan đến biến động HA hiệu quả hơn so với CNN (trường nhìn hạn chế) hoặc RNN (vấn đề vanishing/exploding gradients).

**3. Lý do chọn đề tài/Tính cấp thiết (Problem Statement/Motivation)**

* **Thách thức cố hữu:** Ước tính HA chính xác chỉ từ PPG là cực kỳ khó khăn do:
  + Sự khác biệt lớn giữa các cá nhân (yêu cầu hiệu chuẩn - calibration).
  + Sự biến thiên HA trong cùng một cá nhân (do tư thế, hoạt động, cảm xúc).
  + Tín hiệu PPG nhạy cảm với vị trí đặt cảm biến, lực ép, và đặc biệt là nhiễu do chuyển động (motion artifacts).
  + Mối quan hệ giữa PPG và HA là phức tạp, phi tuyến và bị ảnh hưởng bởi nhiều yếu tố sinh lý khác.
* **Hạn chế của mô hình hiện tại:** Các mô hình hiện có có thể chưa khai thác hết thông tin trong chuỗi thời gian PPG hoặc gặp khó khăn trong việc mô hình hóa các phụ thuộc dài hạn.
* **Thách thức khi áp dụng Transformer:**
  + **Biểu diễn đầu vào:** Làm thế nào để biểu diễn chuỗi tín hiệu PPG (thô, đặc trưng, "patch" hóa) một cách tối ưu cho đầu vào của Transformer?
  + **Chi phí tính toán:** Cơ chế tự chú ý có độ phức tạp tính toán bậc hai theo độ dài chuỗi, đòi hỏi tài nguyên tính toán đáng kể, đặc biệt với chuỗi PPG dài.
  + **Yêu cầu dữ liệu:** Transformer thường cần lượng lớn dữ liệu để huấn luyện hiệu quả.
  + **Tính diễn giải (Interpretability):** Hiểu được cách cơ chế chú ý hoạt động và đóng góp vào dự đoán HA như thế nào.

**4. Câu hỏi nghiên cứu/Mục tiêu đề tài (Research Questions/Objectives)**

* **Câu hỏi nghiên cứu:**
  + Kiến trúc Transformer có thể mô hình hóa hiệu quả mối quan hệ giữa chuỗi tín hiệu PPG và giá trị HA liên tục (Tâm thu - SBP, Tâm trương - DBP) không?
  + Phương pháp biểu diễn tín hiệu PPG nào (chuỗi thô, chuỗi đặc trưng, chuỗi các "patch" tín hiệu) là tối ưu cho mô hình Transformer trong bài toán ước tính HA?
  + Mô hình dựa trên Transformer có độ chính xác (ví dụ: Lỗi tuyệt đối trung bình - MAE, Độ lệch chuẩn của lỗi - SDE theo tiêu chuẩn BHS/AAMI) như thế nào so với các mô hình ML cổ điển và DL khác (CNN, LSTM) trên cùng bộ dữ liệu?
  + Cơ chế tự chú ý trong Transformer có cung cấp thông tin hữu ích về việc mô hình tập trung vào những phần nào của tín hiệu PPG để ước tính HA không?
  + Mô hình Transformer có khả năng chống chịu (robustness) với nhiễu tín hiệu và sự khác biệt giữa các đối tượng như thế nào?
* **Mục tiêu đề tài:**
  + Phát triển quy trình tiền xử lý tín hiệu PPG và định dạng dữ liệu phù hợp cho đầu vào của mạng Transformer.
  + Thiết kế và triển khai các kiến trúc dựa trên Transformer (ví dụ: Transformer Encoder chuẩn, các biến thể như Vision Transformer được điều chỉnh cho chuỗi thời gian) cho bài toán hồi quy ước tính HA.
  + Huấn luyện và đánh giá các mô hình đề xuất trên các bộ dữ liệu chuẩn chứa tín hiệu PPG và HA động mạch xâm lấn (ABP) đồng bộ (ví dụ: MIMIC-III Waveform Database).
  + So sánh hiệu năng của mô hình đề xuất với các phương pháp cơ sở (baseline) như hồi quy tuyến tính, SVM, RF, CNN, LSTM sử dụng các độ đo đánh giá chuẩn (MAE, SDE, hệ số tương quan, phân tích Bland-Altman).
  + Phân tích bản đồ chú ý (attention maps) để diễn giải sự tập trung của mô hình trên các phần khác nhau của chuỗi PPG.
  + Khảo sát khả năng chống chịu nhiễu và nghiên cứu các chiến lược hiệu chuẩn tiềm năng.

**5. Tổng quan tình hình nghiên cứu (Literature Review)**

* Các phương pháp theo dõi HA (xâm lấn, không xâm lấn, liên tục, rời rạc).
* Các phương pháp ước tính HA không dùng bao hơi (cuffless BP estimation): Dựa trên PTT, dựa trên hình thái PPG.
* Ứng dụng của học máy và học sâu trong ước tính HA từ PPG: Kỹ thuật trích xuất đặc trưng, mô hình hồi quy, CNN, RNN/LSTM/GRU.
* Kiến trúc Transformer: Cơ chế tự chú ý (self-attention, multi-head attention), mã hóa vị trí (positional encoding), cấu trúc encoder-decoder.
* Ứng dụng của Transformer ngoài NLP: Thị giác máy tính (Vision Transformer - ViT), phân tích chuỗi thời gian, dự báo, xử lý âm thanh.
* Các nghiên cứu (nếu có) đã áp dụng Transformer vào các tín hiệu sinh lý khác (EEG, ECG) hoặc vào PPG cho các bài toán *khác* ước tính HA.
* Các tiêu chuẩn đánh giá thiết bị đo HA (AAMI/ESH/ISO).

**6. Phương pháp nghiên cứu (Methodology)**

* **Nguồn dữ liệu:** Sử dụng các bộ dữ liệu công khai lớn, chất lượng cao có tín hiệu PPG và HA động mạch xâm lấn (ABP) đồng bộ, độ phân giải cao (ví dụ: MIMIC-III/MIMIC-IV Waveform Database, UCI Cuffless Blood Pressure Estimation dataset). Nêu rõ bộ dữ liệu, các bước tiền xử lý (lọc, phân đoạn), tiêu chí chọn bệnh nhân/bản ghi.
* **Biểu diễn đầu vào:** Thử nghiệm các cách đưa dữ liệu PPG vào Transformer:
  + Chuỗi PPG thô đã được phân đoạn.
  + Chuỗi các vector đặc trưng được trích xuất từ mỗi nhịp hoặc cửa sổ thời gian.
  + Chuỗi các "patch": Chia tín hiệu thành các đoạn ngắn (có thể chồng lấn), sau đó chiếu tuyến tính thành các vector embedding (tương tự ViT).
  + Kết hợp với các kỹ thuật mã hóa vị trí phù hợp cho chuỗi thời gian.
* **Kiến trúc mô hình:**
  + Bắt đầu với kiến trúc Transformer Encoder chuẩn.
  + Thử nghiệm các siêu tham số: số lớp encoder, số lượng attention head, kích thước embedding, kích thước mạng feed-forward.
  + Xem xét các cải tiến/biến thể dành riêng cho chuỗi thời gian (ví dụ: dùng lớp tích chập (convolutional layer) để tạo embedding ban đầu, các loại positional encoding khác nhau).
* **Huấn luyện mô hình:**
  + Xác định hàm mất mát (loss function) phù hợp cho bài toán hồi quy (ví dụ: Mean Squared Error - MSE, Mean Absolute Error - MAE, Smooth L1 Loss).
  + Lựa chọn trình tối ưu hóa (optimizer), ví dụ: Adam, AdamW.
  + Chiến lược phân chia dữ liệu (tập huấn luyện/kiểm định/kiểm thử, chú ý phân chia độc lập theo bệnh nhân - subject-independent split).
  + Nghiên cứu chiến lược hiệu chuẩn (ví dụ: sử dụng một phép đo HA ban đầu bằng bao hơi, transfer learning, fine-tuning cho từng người dùng).
* **Đánh giá:**
  + Độ đo chính: MAE và SDE cho cả SBP và DBP.
  + Độ đo phụ: Hệ số tương quan Pearson (r), Root Mean Squared Error (RMSE).
  + Phân tích Bland-Altman để đánh giá sự phù hợp giữa giá trị ước tính và giá trị tham chiếu.
  + Đối chiếu kết quả với các tiêu chuẩn (ví dụ: AAMI: MAE ≤ 5 mmHg, SDE ≤ 8 mmHg).
  + So sánh một cách công bằng và chặt chẽ với các mô hình baseline (LSTM, GRU, CNN,...) được huấn luyện trên cùng điều kiện dữ liệu.
* **Tính diễn giải:** Trực quan hóa bản đồ chú ý (attention maps) để hiểu mô hình tập trung vào vùng nào của chuỗi tín hiệu đầu vào khi đưa ra dự đoán HA.

**7. Kết quả dự kiến/Đóng góp của đề tài (Expected Results/Contributions)**

* Một mô hình học sâu dựa trên Transformer để ước tính HA liên tục từ tín hiệu PPG.
* Báo cáo phân tích so sánh hiệu năng (độ chính xác, độ bền vững) của phương pháp Transformer so với các phương pháp hiện có.
* Các phát hiện về cách biểu diễn đầu vào và lựa chọn kiến trúc tối ưu khi áp dụng Transformer cho dữ liệu PPG.
* Tiềm năng cải thiện độ chính xác của việc ước tính HA không dùng bao hơi, đóng góp vào việc theo dõi sức khỏe tim mạch tốt hơn.
* Phân tích cơ chế chú ý giúp tăng tính diễn giải cho các dự đoán của mô hình học sâu.
* Đóng góp vào việc ứng dụng các kiến trúc học sâu tiên tiến (Transformer) vào lĩnh vực xử lý tín hiệu y sinh.

Chắc chắn rồi! Sử dụng Diffusion Models, một loại mô hình sinh (generative model) đang rất nổi bật, cho tín hiệu y sinh như PPG là một hướng đi rất mới mẻ và có nhiều tiềm năng. Dưới đây là dàn ý mẫu cho đề tài theo hướng này, tập trung vào việc tạo dữ liệu PPG tổng hợp.

**MẪU ĐỀ TÀI NGHIÊN CỨU**

**1. Tên đề tài:** Tổng hợp Tín hiệu PPG Thực tế sử dụng Mô hình Khuếch tán (Diffusion Models) cho Tăng cường Dữ liệu và Bảo vệ Quyền riêng tư

**(Tên tiếng Anh:** Realistic PPG Signal Synthesis using Diffusion Models for Data Augmentation and Privacy Preservation)

**2. Đặt vấn đề/Tổng quan (Introduction/Background)**

* **Vai trò của tín hiệu PPG:** Tín hiệu Quang phổ đồ (PPG) là nguồn thông tin y sinh quan trọng, được ứng dụng rộng rãi trong theo dõi nhịp tim, HRV, SpO2, giấc ngủ, mức độ căng thẳng, và tiềm năng ước tính huyết áp.
* **Thách thức về dữ liệu:** Việc thu thập các bộ dữ liệu PPG lớn, đa dạng, được gán nhãn chính xác và đại diện cho nhiều tình trạng sinh lý/bệnh lý khác nhau gặp nhiều khó khăn về chi phí, thời gian, và đặc biệt là các quy định nghiêm ngặt về quyền riêng tư dữ liệu y tế (như HIPAA, GDPR).
* **Hạn chế do thiếu dữ liệu:** Sự khan hiếm dữ liệu chất lượng cao cản trở việc xây dựng và kiểm chứng các mô hình học máy/học sâu mạnh mẽ và tổng quát hóa tốt cho các bài toán phân tích PPG.
* **Mô hình sinh và giải pháp:** Các mô hình sinh (Generative Models) có thể học phân phối xác suất của dữ liệu thực và tạo ra dữ liệu tổng hợp (synthetic data) mới. Dữ liệu tổng hợp này có thể dùng để:
  + **Tăng cường dữ liệu (Data Augmentation):** Bổ sung vào tập dữ liệu huấn luyện thực để cải thiện hiệu năng mô hình trên các tác vụ xuôi dòng (downstream tasks).
  + **Bảo vệ quyền riêng tư:** Chia sẻ dữ liệu tổng hợp thay vì dữ liệu bệnh nhân nhạy cảm.
  + **Tạo dữ liệu có kiểm soát:** Phục vụ mô phỏng, kiểm thử thuật toán.
* **Giới thiệu Diffusion Models:** Mô hình Khuếch tán (ví dụ: DDPM - Denoising Diffusion Probabilistic Models) là một lớp mô hình sinh mới nổi, đạt được kết quả ấn tượng trong việc tạo ảnh, âm thanh chất lượng cao. Chúng hoạt động dựa trên hai quá trình: quá trình thuận (forward process) - thêm nhiễu dần vào dữ liệu, và quá trình nghịch (reverse process) - học cách loại bỏ nhiễu để tái tạo dữ liệu từ nhiễu thuần túy. Ưu điểm tiềm năng so với GANs là quá trình huấn luyện ổn định hơn và chất lượng mẫu sinh ra cao hơn.
* **Giả thuyết:** Mô hình khuếch tán có khả năng học được phân phối phức tạp tiềm ẩn của tín hiệu PPG thực và tạo ra dữ liệu PPG tổng hợp có độ trung thực cao, nắm bắt được các biến thiên sinh lý và đặc điểm hình thái sóng thực tế.

**3. Lý do chọn đề tài/Tính cấp thiết (Problem Statement/Motivation)**

* **Nhu cầu dữ liệu PPG tổng hợp chất lượng cao:** Các bộ dữ liệu PPG công khai còn hạn chế về quy mô, sự đa dạng và chất lượng nhãn. Dữ liệu tổng hợp là giải pháp tiềm năng.
* **Vấn đề riêng tư:** Chia sẻ dữ liệu PPG thực tế gặp rào cản pháp lý và đạo đức.
* **Hạn chế của mô hình sinh hiện tại:** Các mô hình như GANs đôi khi gặp vấn đề về ổn định huấn luyện (mode collapse) hoặc tạo ra các chuỗi thời gian thiếu tính mạch lạc dài hạn hoặc các chi tiết tinh vi của tín hiệu PPG thực.
* **Thách thức khi áp dụng Diffusion Models cho PPG:**
  + **Kiến trúc phù hợp:** Cần điều chỉnh kiến trúc mạng nơ-ron (thường là U-Net trong diffusion models cho ảnh) để xử lý hiệu quả dữ liệu chuỗi thời gian 1D như PPG, nắm bắt được các phụ thuộc thời gian.
  + **Mô hình hóa ràng buộc sinh lý:** Làm sao để đảm bảo tín hiệu PPG sinh ra tuân thủ các quy luật sinh lý cơ bản (ví dụ: tính gần tuần hoàn, hình thái sóng hợp lý)?
  + **Đánh giá chất lượng:** Việc đánh giá mức độ "thực tế" và "hữu ích" của dữ liệu PPG tổng hợp là không tầm thường. Cần các độ đo định lượng và định tính phù hợp.
  + **Chi phí tính toán:** Huấn luyện diffusion models có thể đòi hỏi tài nguyên tính toán lớn.

**4. Câu hỏi nghiên cứu/Mục tiêu đề tài (Research Questions/Objectives)**

* **Câu hỏi nghiên cứu:**
  + Mô hình khuếch tán có thể học hiệu quả phân phối của tín hiệu PPG thực và tạo ra các mẫu tổng hợp giống với tín hiệu thực về mặt hình ảnh và thống kê không?
  + Kiến trúc mạng nơ-ron nào (ví dụ: dựa trên U-Net với tích chập 1D, hoặc kết hợp Transformer) là phù hợp nhất để mô hình hóa quá trình khuếch tán thuận và nghịch cho tín hiệu PPG?
  + Chất lượng (độ trung thực, độ đa dạng) của tín hiệu PPG do mô hình khuếch tán tạo ra so với các phương pháp khác (GANs, VAEs) như thế nào?
  + Dữ liệu PPG tổng hợp từ mô hình khuếch tán có thực sự hữu ích khi dùng để tăng cường dữ liệu, cải thiện hiệu năng của các mô hình trên tác vụ xuôi dòng (ví dụ: phân loại giai đoạn ngủ, phát hiện rối loạn nhịp tim) không?
  + Có thể điều khiển quá trình sinh của mô hình khuếch tán để tạo ra tín hiệu PPG với các đặc điểm mong muốn cụ thể (ví dụ: theo nhịp tim, mức độ hoạt động, giai đoạn ngủ) không? (Conditional Generation)
* **Mục tiêu đề tài:**
  + Triển khai và điều chỉnh các kiến trúc mô hình khuếch tán (ví dụ: DDPM, DDIM) để phù hợp với việc sinh chuỗi thời gian PPG 1D.
  + Huấn luyện các mô hình khuếch tán trên các bộ dữ liệu PPG thực tế quy mô lớn.
  + Xây dựng bộ độ đo định lượng và định tính để đánh giá độ thực tế, đa dạng và hợp lý về mặt sinh lý của tín hiệu PPG tổng hợp.
  + Thực hiện các thí nghiệm tăng cường dữ liệu: huấn luyện mô hình cho tác vụ xuôi dòng (ví dụ: phân loại) với và không có dữ liệu tổng hợp để chứng minh tính hữu dụng của dữ liệu sinh ra.
  + So sánh hiệu quả sinh dữ liệu của mô hình khuếch tán với các mô hình sinh cơ sở khác (GANs, VAEs).
  + (Tùy chọn/Nâng cao) Khám phá các phương pháp sinh có điều kiện (conditional generation) cho tín hiệu PPG sử dụng mô hình khuếch tán.

**5. Tổng quan tình hình nghiên cứu (Literature Review)**

* Đặc điểm tín hiệu PPG, phương pháp thu thập, các ứng dụng phổ biến.
* Thách thức về dữ liệu (khan hiếm, riêng tư) trong xử lý tín hiệu y sinh.
* Tổng quan về các mô hình sinh: GANs (WGAN-GP, StyleGAN,...), VAEs, Normalizing Flows, và ứng dụng của chúng cho chuỗi thời gian/tín hiệu y sinh.
* Mô hình Khuếch tán (Diffusion Models): DDPM, DDIM, Score-based generative models. Nền tảng toán học (phương trình vi phân ngẫu nhiên thuận/nghịch - SDEs). Kiến trúc mạng (thường là U-Net). Các thành công trong sinh ảnh/âm thanh.
* Các nghiên cứu (nếu có) áp dụng mô hình khuếch tán cho dữ liệu chuỗi thời gian nói chung hoặc tín hiệu y sinh nói riêng (đây là lĩnh vực đang phát triển rất nhanh). Các điều chỉnh cần thiết cho tín hiệu 1D.
* Các phương pháp đánh giá mô hình sinh, đặc biệt cho dữ liệu chuỗi thời gian (ví dụ: các độ đo trên miền đặc trưng, đánh giá bởi chuyên gia, đánh giá qua hiệu năng tác vụ xuôi dòng).
* Kỹ thuật tăng cường dữ liệu trong xử lý tín hiệu y sinh.

**6. Phương pháp nghiên cứu (Methodology)**

* **Nguồn dữ liệu:** Sử dụng các bộ dữ liệu PPG công khai lớn (ví dụ: MIMIC-III/IV, các CSDL trên PhysioNet,...) hoặc dữ liệu thu thập riêng (nếu có). Nêu rõ các bước tiền xử lý (lọc, chuẩn hóa, phân đoạn thành các cửa sổ có độ dài cố định).
* **Triển khai Mô hình Khuếch tán:**
  + Chọn một framework cơ sở (ví dụ: DDPM, DDIM).
  + Thiết kế kiến trúc mạng nơ-ron cho quá trình nghịch (mạng khử nhiễu - denoising network). Thường là kiến trúc U-Net được điều chỉnh cho chuỗi thời gian 1D (sử dụng lớp tích chập 1D, có thể kết hợp các lớp self-attention).
  + Xác định lịch trình thêm nhiễu (variance schedule βt) cho quá trình thuận.
  + Triển khai vòng lặp huấn luyện (lấy mẫu nhiễu, áp dụng quá trình thuận, huấn luyện mạng khử nhiễu dự đoán nhiễu hoặc tín hiệu gốc).
  + Triển khai quá trình lấy mẫu (sampling): Bắt đầu từ nhiễu ngẫu nhiên và khử nhiễu lặp đi lặp lại bằng mạng đã huấn luyện.
* **Độ đo đánh giá:**
  + **Định tính:** Chuyên gia y sinh/kỹ thuật đánh giá trực quan dạng sóng PPG sinh ra. Vẽ đồ thị phân phối của các đặc trưng hình thái quan trọng (biên độ, độ rộng đỉnh, sự hiện diện của khuyết dicrotic) và so sánh với dữ liệu thực.
  + **Định lượng:**
    - **Độ tương đồng thống kê:** So sánh phân phối của các đặc trưng miền thời gian (HR, HRV nếu có thể trích xuất) và miền tần số giữa tập dữ liệu thực và tổng hợp (ví dụ: dùng khoảng cách Wasserstein, KL divergence).
    - **(Thử nghiệm)** Điều chỉnh các độ đo như Frechet Inception Distance (FID) hoặc Inception Score (IS) bằng cách sử dụng đặc trưng trích xuất từ một mô hình đã huấn luyện trước (pre-trained model) trên PPG.
    - **Khả năng dự đoán (Tác vụ xuôi dòng):** Đây là thước đo quan trọng nhất về tính hữu dụng. Huấn luyện một mô hình chuẩn (ví dụ: bộ phân loại) trên các kịch bản: (1) Chỉ dữ liệu thực, (2) Dữ liệu thực + Dữ liệu tổng hợp, (3) Chỉ dữ liệu tổng hợp. So sánh hiệu năng trên một tập kiểm thử dữ liệu thực độc lập.
* **So sánh:** Huấn luyện các mô hình sinh cơ sở (ví dụ: WGAN-GP, VAE) trên cùng bộ dữ liệu và so sánh chất lượng PPG sinh ra bằng các độ đo đã xác định.
* **Sinh có điều kiện (Tùy chọn):** Triển khai cơ chế điều khiển (ví dụ: đưa nhãn lớp hoặc biến liên tục làm đầu vào bổ sung cho mạng khử nhiễu) và đánh giá khả năng tạo tín hiệu với đặc tính mong muốn.

**7. Kết quả dự kiến/Đóng góp của đề tài (Expected Results/Contributions)**

* Một mô hình khuếch tán có khả năng sinh dữ liệu PPG tổng hợp với độ trung thực cao, đa dạng và hợp lý về mặt sinh lý.
* Một quy trình phương pháp luận vững chắc cho việc huấn luyện và đánh giá mô hình khuếch tán cho việc sinh tín hiệu PPG.
* Các so sánh định lượng và định tính cho thấy ưu điểm tiềm năng của mô hình khuếch tán so với các kỹ thuật sinh khác cho PPG.
* Bằng chứng thực nghiệm cho thấy hiệu quả của dữ liệu PPG tổng hợp từ mô hình khuếch tán trong việc tăng cường dữ liệu thực và cải thiện hiệu năng mô hình trên tác vụ xuôi dòng.
* Đóng góp vào việc giải quyết các thách thức về khan hiếm dữ liệu và bảo vệ quyền riêng tư trong nghiên cứu sức khỏe dựa trên PPG.
* Tiềm năng tạo ra dữ liệu PPG có thể kiểm soát được cho các mục đích nghiên cứu hoặc mô phỏng cụ thể (thông qua sinh có điều kiện).