**MỤC LỤC**

**1. Phần MỞ ĐẦU**

**1.1. Tính cấp thiết của đề tài**

* Giới thiệu vấn đề bệnh tim và tác động toàn cầu.
* Tầm quan trọng của việc dự đoán sớm bệnh tim để can thiệp kịp thời.

**1.2. Mục đích của đề tài**

* Xây dựng mô hình dự đoán bệnh tim với độ chính xác cao.
* Đóng góp vào việc hỗ trợ chẩn đoán y khoa và giảm thiểu rủi ro.

**1.3. Cách tiếp cận và phương pháp nghiên cứu**

* Sử dụng học máy (machine learning) để phân tích dữ liệu y khoa.
* Áp dụng các thuật toán như Random Forest, SVM, Neural Network.
* Quy trình: thu thập dữ liệu, tiền xử lý, huấn luyện mô hình, đánh giá.
* **Đối tượng nghiên cứu**: Dữ liệu bệnh nhân (tuổi, huyết áp, cholesterol, v.v.).
* **Phạm vi nghiên cứu**: Tập trung vào dữ liệu từ bệnh viện hoặc bộ dữ liệu công khai (ví dụ: UCI Heart Disease Dataset).

**2. Phần NỘI DUNG**

**Chương 1: Cơ sở lý thuyết và tổng quan**

**1.1. Tổng quan về bệnh tim**

* Định nghĩa và các loại bệnh tim phổ biến.
* Các yếu tố nguy cơ (di truyền, lối sống, bệnh lý nền).

**1.2. Tổng quan về học máy trong y khoa**

* Ứng dụng của học máy trong chẩn đoán bệnh.
* Các thuật toán học máy phổ biến trong dự đoán bệnh tim.

**Chương 2: Phương pháp nghiên cứu**

**2.1. Thu thập và tiền xử lý dữ liệu**

* Nguồn dữ liệu: Bộ dữ liệu UCI Heart Disease hoặc bệnh viện.
* Tiền xử lý: xử lý giá trị thiếu, chuẩn hóa dữ liệu, mã hóa biến phân loại.

**2.2. Xây dựng và huấn luyện mô hình**

* Lựa chọn thuật toán: Random Forest, SVM, Neural Network.
* Quy trình huấn luyện: chia tập train/test, tối ưu siêu tham số.
* Đánh giá mô hình: sử dụng độ chính xác, AUC, F1-score.

**Chương 3: Kết quả và thảo luận**

**3.1. Kết quả thực nghiệm**

* Hiệu suất của các mô hình (bảng so sánh độ chính xác, ROC curve).
* Phân tích đặc trưng quan trọng (feature importance).

**3.2. Thảo luận**

* So sánh hiệu quả các thuật toán.
* Hạn chế của mô hình và hướng cải thiện.
* Ứng dụng thực tiễn của mô hình trong y khoa.

**3. Phần KẾT LUẬN**

* Tóm tắt kết quả đạt được.
* Đề xuất hướng nghiên cứu tiếp theo (ví dụ: tích hợp dữ liệu thời gian thực, cải thiện thuật toán).

**4. Tài liệu tham khảo**

[1] UCI Machine Learning Repository. Heart Disease Dataset.  
[2] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.  
[3] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer.

**KẾ HOẠCH THỰC HIỆN**

| **STT** | **Thời gian** | **Công việc** | **Ghi chú** |
| --- | --- | --- | --- |
| 1 | 05/05/2025 - 12/05/2025 | Thu thập tài liệu và xây dựng cơ sở lý thuyết | Phân công nhóm nghiên cứu tài liệu |
| 2 | 13/05/2025 - 20/05/2025 | Thu thập và tiền xử lý dữ liệu | Kiểm tra chất lượng dữ liệu |
| 3 | 21/05/2025 - 05/06/2025 | Xây dựng, huấn luyện và đánh giá mô hình | Lưu ý tối ưu hóa siêu tham số |
| 4 | 06/06/2025 - 15/06/2025 | Phân tích kết quả và viết báo cáo | Hoàn thiện thảo luận và kết luận |
| 5 | 16/06/2025 - 20/06/2025 | Kiểm tra và nộp báo cáo | Đảm bảo định dạng theo yêu cầu |