ỦY BAN NHÂN DÂN THÀNH PHỐ HỒ CHÍ MINH

**TRƯỜNG ĐẠI HỌC SÀI GÒN**

KHOA CÔNG NGHỆ THÔNG TIN

**BÁO CÁO TỔNG KẾT**

**Môn Học: Phương Pháp Nghiên Cứu Khoa Học**

**Đề Tài: Dự Đoán Bệnh Tim Bằng Máy Học**

**Giảng viên: Đỗ Như Tài**

Nhóm

Phạm Tấn Khương – 3122410191

Hoàng Vũ - 3122560089

Huỳnh Thanh Bình - 3122410033

Nguyễn Minh Tú - 3120411167

Thành phố HCM, tháng 5 năm 2025

# **Tóm Tắt (Abstract)**

# **Chương 1: Tổng Quan Vấn Đề**

## **1 Lý do chọn đề tài**

Nghiên cứu dự đoán bệnh tim bằng máy học có nhiều lý do quan trọng, bao gồm:

1. **Phát hiện sớm và phòng ngừa**: Các mô hình máy học có thể giúp phát hiện bệnh tim ngay từ giai đoạn sớm, khi các triệu chứng chưa rõ ràng. Điều này có thể giúp bác sĩ can thiệp kịp thời, giảm nguy cơ tử vong hoặc các biến chứng nghiêm trọng.
2. **Cải thiện độ chính xác**: Máy học có khả năng xử lý và phân tích một lượng lớn dữ liệu y tế, bao gồm các yếu tố nguy cơ (như huyết áp, mức cholesterol, thói quen sinh hoạt, lịch sử gia đình) để đưa ra dự đoán chính xác hơn so với việc chẩn đoán chỉ dựa vào kinh nghiệm của bác sĩ.
3. **Giảm tải cho bác sĩ**: Với sự hỗ trợ của các mô hình máy học, các bác sĩ có thể nhanh chóng phân loại và đánh giá nguy cơ bệnh tim của bệnh nhân, từ đó giảm bớt khối lượng công việc và tập trung vào các ca bệnh phức tạp hơn.
4. **Cải thiện quản lý sức khỏe cộng đồng**: Dự đoán bệnh tim bằng máy học giúp các tổ chức y tế phân tích xu hướng bệnh tật trong cộng đồng, từ đó thiết kế các chiến lược phòng ngừa và can thiệp hợp lý hơn
5. **Ứng dụng trong hệ thống chăm sóc sức khỏe tự động**: Máy học có thể được tích hợp vào các ứng dụng y tế và hệ thống chăm sóc sức khỏe, cung cấp các dự đoán và cảnh báo cho người dùng và các chuyên gia y tế ngay lập tức, tạo ra một hệ thống chăm sóc sức khỏe tiên tiến và chủ động.
6. **Khả năng học từ dữ liệu lớn**: Các mô hình máy học có thể học và cải thiện từ những dữ liệu lớn, đồng thời phân tích mối quan hệ phức tạp giữa các yếu tố nguy cơ và bệnh lý mà không cần hiểu rõ tất cả các yếu tố đó một cách chi tiết.

Tóm lại, nghiên cứu dự đoán bệnh tim bằng máy học không chỉ giúp tăng cường khả năng chẩn đoán và điều trị mà còn góp phần vào việc tối ưu hóa và cải thiện hệ thống chăm sóc sức khỏe.

## **2 Vấn đề nghiên cứu**

Vấn đề nghiên cứu của chúng em là bệnh tim mạch là một trong những nguyên nhân gây tử vong hàng đầu trên toàn thế giới. Việc chẩn đoán bệnh tim thường phụ thuộc vào kinh nghiệm của bác sĩ và các xét nghiệm lâm sàng, tuy nhiên quá trình này có thể tốn thời gian, chi phí cao và đôi khi không phát hiện được bệnh ở giai đoạn sớm. Trong khi đó, nhiều yếu tố nguy cơ như tuổi tác, huyết áp, cholesterol, tiểu sử bệnh lý, thói quen sinh hoạt... có thể được thu thập dễ dàng và cung cấp dữ liệu phong phú để hỗ trợ quá trình chẩn đoán.

Máy học (Machine Learning) – một nhánh của trí tuệ nhân tạo – có khả năng xử lý và phân tích dữ liệu phức tạp, từ đó phát hiện các mẫu và mối quan hệ tiềm ẩn giữa các yếu tố đầu vào và nguy cơ mắc bệnh tim. Tuy nhiên, việc áp dụng máy học vào dự đoán bệnh tim cũng đặt ra nhiều thách thức, như lựa chọn thuật toán phù hợp, xử lý dữ liệu thiếu, đảm bảo độ chính xác, độ tin cậy của mô hình và khả năng triển khai thực tế trong môi trường y tế.

Do đó, đề tài “Dự đoán bệnh tim bằng máy học” được đặt ra nhằm mục tiêu nghiên cứu và xây dựng một mô hình có khả năng dự đoán chính xác nguy cơ mắc bệnh tim từ dữ liệu bệnh nhân, hỗ trợ các bác sĩ trong công tác chẩn đoán và phòng ngừa bệnh một cách hiệu quả hơn.

## **3 Mục tiêu nghiên cứu**

Mục tiêu chính của nghiên cứu là đào tạo ra mô hình máy học vói độ chính xác cao nhất và thời gian xử lý nhanh nhất để dự đoán bệnh tim.

## **4 Câu hỏi nghiên cứu**

Chúng em đặt ra 4 câu hỏi cần nghiên cứu:

1. Thuật toán máy học nào phù hợp nhất để áp dụng trong việc dự đoán bệnh tim ?
2. Các yếu tố nào trong dữ liệu (ví dụ: huyết áp, cholesterol, tuổi, giới tính, nhịp tim...) có mức độ ảnh hưởng cao đến kết quả dự đoán nguy cơ mắc bệnh tim ?
3. Làm thế nào để đánh giá hiệu suất của mô hình dự đoán một cách toàn diện ?
4. Làm thế nào để tối ưu hóa mô hình máy học để cân bằng giữa độ chính xác và thời gian xử lý, nhằm phục vụ tốt trong môi trường thực tế?

## **5 Phạm vi nghiên cứu**

### **5.1 Đối tượng nghiên cứu**

Đối tượng nghiên cứu của đề tài bao gồm các mô hình máy học và dữ liệu y tế của bệnh nhân dùng để dự đoán nguy cơ mắc bệnh tim. Cụ thể:

* Về mô hình: Nghiên cứu tập trung vào một số thuật toán máy học phổ biến như Logistic Regression, Random Forest, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), v.v., nhằm đánh giá và so sánh hiệu quả của chúng trong việc phân loại nguy cơ mắc bệnh tim.
* Về dữ liệu: Dữ liệu được sử dụng trong nghiên cứu là từ các bộ dữ liệu y tế công khai, đó là UCI Heart Disease Dataset thuộc UCI data Repository. Bộ dữ liệu này bao gồm các thông tin lâm sàng của bệnh nhân như tuổi, giới tính, huyết áp, mức cholesterol, nhịp tim, chỉ số đường huyết, và kết quả chẩn đoán bệnh tim.

### **5.2 Không gian và thời gian nghiên cứu**

Nghiên cứu không đi sâu vào các khía cạnh y học chuyên sâu như chẩn đoán hình ảnh (ECG, MRI), dữ liệu thời gian thực hoặc can thiệp lâm sàng. Phạm vi nghiên cứu chủ yếu mang tính mô phỏng và thử nghiệm trong môi trường học thuật, chưa triển khai trong hệ thống y tế thực tế. Nghiên cứu này chỉ kéo dài trong khoảng 15 tuấn