TRƯỜNG ĐẠI HỌC MỞ THÀNH PHỐ HỒ CHÍ MINH

**KHOA CÔNG NGHỆ THÔNG TIN**

---🙡 🕮 🙣---



**BÁO CÁO MÔN HỌC**

**PHÂN TÍCH DỮ LIỆU**

**Đề tài**

**MÔ HÌNH PHÂN TÍCH**

**VÀ DỰ ĐOÁN BỆNH TIM MẠCH**

**Sinh viên thực hiện: Đỗ Thị Thảo Nguyên**

**Mã số sinh viên: 2251010064**

**Sinh viên thực hiện: Nguyễn Thanh Nhi**

**Mã số sinh viên: 2251010067**

**Lớp: DH22CS01**

**Giảng viên hướng dẫn: GV Hồ Hướng Thiên**

**Tháng 04 năm 2025**

**MỤC LỤC**

**CHƯƠNG 1:** TỔNG QUAN

**1.** Giới thiệu: Mô hình phân tích và dự đoán bệnh tim mạch.

**2.** Khái quát bệnh tim mạch và thực trạng.

**3.** Dựa trên dữ liệu y tế (dataset heart disease), xây dựng mô hình máy học để dự đoán bệnh tim.

**4.** Phân tích các yếu tố gây bệnh hoặc có nguy cơ gây bệnh.

**CHƯƠNG 2:** PHƯƠNG PHÁP NGHIÊN CỨU

1. Thu thập dữ liệu y tế
2. Tiền xử lý dữ liệu
3. Mô hình học máy (Machine Learning)
4. Huấn luyện mô hình

**CHƯƠNG 3:** KẾT QUẢ NGHIÊN CỨU

1. Phân tích dữ liệu ban đầu
2. Đánh giá mô hình
3. Các dữ liệu quan trọng ảnh hưởng đến kết quả nghiên cứu

**CHƯƠNG 4:** KẾT LUẬN

1. Kết luận nghiên cứu
2. Đề xuất áp dụng thực tế

**CHƯƠNG 5:** TÀI LIỆU THAM KHẢO

**CHƯƠNG 1: TỔNG QUAN**

1. **Giới thiệu:**

Mô hình phân tích và dự đoán là sự kết hợp giữa khoa học dữ liệu và học máy giúp người dùng nghiên cứu và xây dựng các dự đoán có tỷ lệ chính xác cao, phục vụ cho các nhu cầu thiết yếu của con người trong mọi lĩnh vực. Chúng tôi đã sử dụng mô hình trên để xây dựng một dự án trong lĩnh vực y tế. Mô hình phân tích và dự đoán bệnh tim mạch được xây dựng dựa trên dữ liệu y tế được thu thập trên nền tảng **Kaggle,** kết hợp với các thuật toán máy học (**Machine Learning**).

1. **Khái quát về bệnh tim và thực trạng:**

**- Bệnh tim:** là một tập hợp các tình trạng ảnh hưởng đến sức khỏe của trái tim và hiệu suất hoạt động của các mạch máu chịu trách nhiệm cung cấp oxy, dưỡng chất đến nuôi cơ tim. Các loại bệnh tim mạch thường gặp bao gồm: rối loạn nhịp tim, cao huyết áp, bệnh mạch vành, các bệnh về van tim cùng nhiều loại khác. Đây là những nguyên nhân chính dẫn đến tình trạng tử vong.

**- Thực trạng:** Bệnh lý tim mạch và chuyển hóa là một trong những nhóm bệnh lý phức tạp có tỷ lệ mắc cao trong cộng đồng. Bệnh thường để lại di chứng nặng nề, làm giảm chất lượng cuộc sống của người bệnh, gây gánh nặng cho gia đình và xã hội. Các bác sỹ đã tiếp nhận cấp cứu nhiều trường hợp bị nhồi máu cơ tim ở độ tuổi trước 40. Về nguyên nhân góp phần làm gia tăng bệnh tim mạch, theo chuyên gia có nhiều yếu tố, trong đó toàn cầu hóa và đô thị hóa, sự thay đổi môi trường như là những tác nhân làm tăng lối sống không lành mạnh như hút thuốc lá, lạm dụng uống rượu, bia, ăn uống không hợp lý, ít vận động thể lực là yếu tố nguy cơ này làm phát triển bệnh lý tim mạch. Ngoài ra, yếu tố công việc căng thẳng, stress tinh thần cũng làm gia tăng căn bệnh này.

**3. Dựa trên dữ liệu y tế ( dataset heart disease ), xây dựng mô hình máy học để dự đoán bệnh tim.**

**- Mục tiêu chính**: Mục tiêu của việc xây dựng mô hình máy học là dự đoán khả năng mắc bệnh tim mạch dựa trên các yếu tố y tế như tuổi, giới tính, huyết áp, cholesterol, thói quen sống và các chỉ số sức khỏe khác. Mô hình này sẽ giúp xác định những cá nhân có nguy cơ cao mắc bệnh tim, từ đó hỗ trợ phát hiện sớm và đưa ra các biện pháp phòng ngừa hoặc điều trị phù hợp. Nhờ đó, bệnh nhân có thể thực hiện các can thiệp kịp thời, như điều chỉnh lối sống và nhận được sự chăm sóc y tế sớm, giảm thiểu nguy cơ phát triển các biến chứng nghiêm trọng.

**- Lợi ích**: Sử dụng mô hình máy học trong việc dự đoán nguy cơ mắc bệnh tim mang lại nhiều lợi ích rõ rệt. Trước tiên, nó giúp nhận diện nhanh chóng những người có nguy mắc bệnh, từ đó tạo điều kiện để thực hiện các biện pháp can thiệp sớm. Điều này không chỉ giảm tỷ lệ tử vong mà còn giúp tiết kiệm chi phí điều trị, vì các biện pháp phòng ngừa sớm có thể ngăn chặn bệnh phát triển nghiêm trọng. Bên cạnh đó, việc phát hiện nguy cơ bệnh tim từ sớm sẽ khuyến khích bệnh nhân thay đổi lối sống, cải thiện chế độ ăn uống và tăng cường tập luyện, giúp giảm thiểu nguy cơ mắc bệnh và duy trì sức khỏe lâu dài.

**4. Phân tích các yếu tố gây bệnh hoặc có nguy cơ gây bệnh.**

- Theo Tổ chức Y tế Thế giới, hàng năm có đến hơn 17 triệu người tử vong vì bệnh tim mạch. Con số này cho thấy mức độ phổ biến của các bệnh lý về tim, trong đó thường gặp nhất là cao huyết áp, bệnh mạch vành và đột quỵ.

- Bệnh tim mạch có rất nhiều hình thức, và vì vậy cũng có nhiều yếu tố nguy cơ dẫn đến các bệnh lý về tim được chia thành hai nhóm. Nhóm thứ nhất là những yếu tố nguy cơ không thể thay đổi như tuổi tác, giới tính, yếu tố di truyền, nhóm thứ hai là những yếu tố nguy cơ có thể thay đổi như thói quen sinh hoạt, vận động… Hiểu rõ các yếu tố nguy cơ này sẽ giúp bạn phòng tránh bệnh tim mạch tốt nhất.

**+ Các yếu tố nguy cơ không thể thay đổi:**

* **Giới tính**: Nhìn chung, nam giới có nguy cơ bị đau tim nhiều hơn phụ nữ. Tuy nhiên ở phụ nữ ở thời kỳ mãn kinh, nguy cơ sẽ tăng cao hơn và sau tuổi 65, nguy cơ mắc bệnh tim mạch ở nam giới và phụ nữ là như nhau.
* **Di truyền**: Nếu trong gia đình có tiền sử bệnh tim mạch, ví dụ như cha mẹ hoặc anh chị em của bạn đã từng gặp phải các vấn đề về tim mạch trước tuổi 55, bạn sẽ có nguy cơ mắc bệnh tim mạch nhiều hơn bình thường. Ngoài ra những yếu tố rủi ro dẫn đến nguy cơ tim mạch như cao huyết áp, tiểu đường và béo phì cũng có thể mang tính di truyền.
* **Tuổi tác**: Tuổi già là một trong những yếu tố nguy cơ của bệnh tim mạch. Càng lớn tuổi, hoạt động của tim càng kém hiệu quà. Thành tim dày lên, các động mạch cứng lại khiến cho quá trình bơm máu cũng trở nên khó khăn, đó là lý do vì sao nguy cơ mắc bệnh tim mạch gia tăng theo tuổi.

+ **Các yếu tố nguy cơ có thể thay đổi:**

* **Béo phì và thừa cân**: Việc thừa cân sẽ góp phần gia tăng tổng mức cholesterol trong máu, đồng thời mang đến nguy cơ huyết áp cao và bệnh mạch vành. Nếu ở cấp độ béo phì, lượng cholesterol sẽ tăng vọt, huyết áp đặc biệt cao và kéo theo bệnh tiểu đường. Trong nhiều trường hợp, chứng béo phì chịu trách nhiệm cho hầu hết các bệnh tật liên đới, trong đó đa phần là bệnh tim mạch. Có thể nói béo phì là một yếu tố nguy cơ mang đến nhiều yếu tố nguy cơ khác có thể gây ra bệnh tim mạch.
* **Nghiện thuốc lá:** Hầu hết mọi người đều biết rằng việc hút thuốc lá làm tăng nguy cơ ung thư phổi, nhưng ít ai nhận ra nó cũng làm tăng đáng kể nguy cơ mắc bệnh tim mạch. Nghiên cứu cho thấy hút thuốc làm tăng nhịp tim, thắt chặt các động mạch lớn và làm cho nhịp tim bất thường, có thể dẫn đến loạn nhịp tim...khiến hoạt động của tim kém hiệu quả và trở nên khó khăn hơn. Việc hút thuốc cũng làm tăng huyết áp và tăng nguy cơ đột quỵ ở những người đã có sẵn chứng huyết áp cao. Ngoài nicotine, các hóa chất khác có trong khói thuốc như carbon monoxide cũng có hại cho tim. Những chất này dẫn đến tích tụ mảng bám trong động mạch, ảnh hưởng đến cholesterol và mức fibrinogen một yếu tố làm đông máu, điều này khiến cho nguy cơ đông máu tăng và có thể dẫn đến đau tim.
* **Thiếu vận động thể chất:** Những người lười vận động hoặc ít có cơ hội vận động thể chất sẽ có nguy cơ mắc bệnh tim mạch cao hơn những người thường xuyên tập thể dục. Việc hoạt động và tập luyện sẽ đốt cháy calo, giúp kiểm soát mức cholesterol và bệnh tiểu đường đồng thời có thể hạ huyết áp. Tập thể dục cũng tăng cường sự dẻo dai cho cơ tim và làm cho các động mạch linh hoạt hơn. Những người tích cực đốt cháy 500-3500 calo mỗi tuần bằng cách tập thể dục hoặc các hình thức vận đông khác thường sống lâu hơn những người không tập thể dục. Ngay cả việc tập thể dục với cường độ trung bình nhưng đều đặn và thường xuyên cũng rất hữu ích.
* **Cao huyết áp:** Chứng cao huyết áp sẽ dẫn đến nguy cơ đau tim và đột quỵ. Nếu kết hợp thêm cả béo phì, nghiện thuốc lá và cholesterol cao thì nguy cơ bệnh tim hay đột quỵ sẽ tăng lên đáng kể.Huyết áp có thể thay đổi tùy theo điều kiện vận động và tuổi tác, nhưng về cơ bản, chỉ số huyết áp ở người lớn khi đang nghỉ ngơi nên ở mức 120/80.
* **Cholesterol trong máu cao:** Cholesterol là một chất tương tự chất béo có sẵn trong máu. Gan sản xuất cholesterol để hình thành màng tế bào và tạo một số hormon nhất định. Ngoài lượng cholesterol này ra, cơ thể còn nhận thêm một lượng cholesterol khác từ những thực phẩm được dung nạp, những thực phẩm này đa phần có nguồn gốc động vật như thịt, trứng, các sản phẩm từ sữa...Mặc dù ta vẫn thường đổ lỗi cho lượng cholesterol trong thức ăn khiến cho chỉ số cholesterol trong máu tăng cao, nhưng thực tế thủ phạm chính lại là các chất béo bão hòa có trong thực phẩm. Hãy cẩn thận ở điềm này, vì có những thực phẩm không chứa cholesterol nhưng vẫn chứa một lượng lớn các chất béo bão hòa khác. Những loại thực phẩm chứa nhiều chất béo bão hòa thường là thực phẩm có nguồn gốc bơ sữa và thịt đỏ.Các chất béo bão hòa này làm gia tăng lượng cholesterol “xấu” (LDL), tạo ra những mảng bám trên thành động mạch và bắt đầu quá trình xơ vữa động mạch. Khi các mảng bám tích tụ trong các động mạch vành cung cấp máu cho tim, nguy cơ đau tim sẽ tăng cao.
* **Tiểu đường:** Ước tính có đến 65% số người tiểu đường tử vong do các bệnh tim mạch. Tiểu đường làm tăng nguy cơ bệnh tim và đột quỵ. Một phần của nguyên nhân này là do bệnh tiểu đường làm ảnh hưởng đến cholesterol và triglyceride, ngoài ra người bị tiểu đường cũng có thể bị huyết áp cao và béo phì kèm theo, do vậy nguy cơ cũng cao hơn.Có thể thấy phần lớn nguyên nhân dẫn tới bệnh lý tim mạch thuộc về nhóm yếu tố nguy cơ có thể thay đổi được. Do đó để duy trì một trái tim mạnh khỏe của mình ngay từ hôm nay bạn hãy thực hiện chế độ ăn uống và lối sống lành mạnh đồng thời kiểm tra sức khỏe định kỳ thường xuyên để phát hiện sớm các bệnh lý nếu có.

**CHƯƠNG 2: PHƯƠNG PHÁP NGHIÊN CỨU**

1. **Thu thập dữ liệu y tế:** Dữ liệu y tế hỗ trợ nghiên cứu mô hình phân tích và dự đoán bệnh tim được tìm hiểu và thu thập dựa trên nền tảng Kaggle. Nguồn dữ liệu thu thập được 10000 dòng với các chỉ số sức khỏe liên quan đến bệnh tim mạch, gồm 21 cột: Age, Gender, Blood Pressure, Cholesterol Level, Exercise Habits, Smoking, Family Heart Disease, Diabetes, BMI, High Blood Pressure, Low HDL Cholesterol, High LDL Cholesterol, Alcohol Comsumption, Stress Level, Sleep Hours, Sugar Comsumption, Triglyceride Level, Fasting Blood Sugar, CRP Level, Homocysteine Level, Heart Disease Status. Dựa trên các thông tin nói trên ta có thể xây dựng mô hình dự đoán bệnh tim để hỗ trợ lĩnh vực y tế trong tương lai.

[Tại sao tỷ lệ mắc bệnh tim mạch ở Việt Nam ngày càng báo động?](https://tamanhhospital.vn/ty-le-mac-benh-tim-mach-o-viet-nam/#:~:text=Trong%20nh%E1%BB%AFng%20n%C4%83m%20g%E1%BA%A7n%20%C4%91%C3%A2y%20t%E1%BB%B7%20l%E1%BB%87%20m%E1%BA%AFc,m%E1%BA%A1nh%2C%20%C3%A1p%20l%E1%BB%B1c%20c%C3%B4ng%20vi%E1%BB%87c%20ng%C3%A0y%20c%C3%A0ng%20cao.)

[Các yếu tố nguy cơ dẫn đến bệnh Tim Mạch - Chương trình mục tiêu quốc gia - Cổng thông tin Bộ Y tế](https://moh.gov.vn/chuong-trinh-muc-tieu-quoc-gia/-/asset_publisher/7ng11fEWgASC/content/cac-yeu-to-nguy-co-dan-en-benh-tim-mach?inheritRedirect=false)