

# BỆNH ĐỘNG MẠCH VÀNH

Nguyễn Phương Bích Đỗ Thảo Giang Nguyễn Thị Phương Thảo

Mô hình Toán sinh thái  
Khoa Toán - Cơ - Tin học  
Trường Đại học Khoa học Tự nhiên

Hà Nội, 2024

# Mục lục

- 1 Tổng quan về bệnh
- 2 Tập dữ liệu
- 3 Phân lớp Naive Bayes
- 4 Mô hình hồi quy Ordinal Logistics
- 5 Kết luận
- 6 Tài liệu tham khảo

# Mục lục

1 Tổng quan về bệnh

2 Tập dữ liệu

3 Phân lớp Naive Bayes

4 Mô hình hồi quy Ordinal Logistics

5 Kết luận

6 Tài liệu tham khảo

# Tìm hiểu cấu tạo hệ tim mạch

**Hệ tim mạch** bao gồm tim và hệ mạch máu. Đảm nhận 3 chức năng chính:

- Chức năng vận chuyển (oxi, chất dinh dưỡng, chất thải).
- Chức năng bảo vệ.
- Điều chỉnh nhiệt độ, pH cơ thể.

# Tìm hiểu cấu tạo hệ tim mạch

- **Tim:** cơ quan trung tâm của hệ thống tim mạch, có nhiệm vụ chính là bơm máu đi khắp cơ thể.
- **Hệ thống mạch vành:** hệ thống mạch đặc biệt bao quanh trái tim có chức năng duy trì dinh dưỡng cho cơ tim.



# Tổng quan về bệnh lý

**Bệnh động mạch vành (Coronary Artery Disease - CAD)** hay còn gọi là bệnh tim thiếu máu cục bộ. Do hẹp lòng ĐMV, nguyên nhân do xơ vữa động mạch hoặc bị tắc nghẽn xơ vữa động mạch.

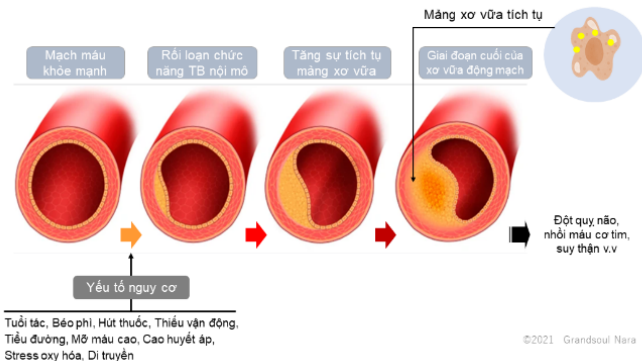
- Theo WHO, ước tính có 17,9 triệu người tử vong do bệnh tim mạch vào năm 2019 và phần lớn trong đó là do xơ vữa.
- Ở Việt Nam, tỷ lệ tử vong do bệnh này chiếm 11 - 36%.
- Tỷ lệ mắc bệnh ở nam giới nhiều hơn nữ giới, 50-60 tuổi.

# Sự hình thành các mảng xơ vữa

Cholesterol được vận chuyển trong máu với 3 loại Lipoprotein được ký hiệu là: LDL-c, VLDL-c, HDL-c.

**LDL-c (Low Density Lipoprotein)** có liên quan trực tiếp đến sự hình thành mảng xơ vữa.

## SỰ HÌNH THÀNH XƠ VỮA ĐỘNG MẠCH



## Sự hình thành các mảng xơ vữa

<b>Chỉ số</b>	<b>Bình thường</b>	<b>Cao</b>
1/Cholesterol toàn phần	<200 mg/dL (< 5,2 mmol/L)	> 240 mg/dL (> 6,2 mmol/L)
2/ LDL-c	< 130 mg/dL (< 3,3 mmol/L)	> 160 mg/dL (> 4,1 mmol/L)
3/HDL-c	> 50 mg/dL (> 1,3 mmol/L)	< 40 mg/dL (< 1 mmol/L)

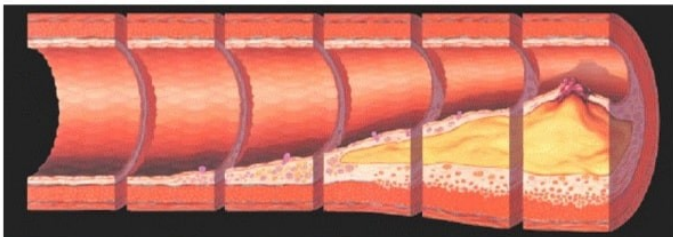


# Diễn biến tự nhiên của mảng xơ vữa

## Diễn biến tự nhiên của mảng xơ vữa

**YTNC: Tăng huyết áp, ĐTD, RL lipid máu, béo phì, tuổi, hút thuốc lá**

Tế bào Bọt    Vết mỡ    Tổn thương Trung gian    Xơ vữa Động mạch    Mảng xơ vữa    Biến chứng/Vỡ mảng XVDM



Từ thập kỉ  
thứ nhất

Từ thập kỉ  
thứ ba

Từ thập kỉ  
thứ tư

# Dấu hiệu và nguyên nhân gây bệnh

- Dấu hiệu:



Chest Pain



Vomiting



Sweating



Dizziness

- Nguyên nhân

- Các yếu tố không thể thay thế được: Tuổi tác, giới tính, tiền sử gia đình, chủng tộc.
- Các yếu tố thay đổi được: Cao huyết áp, hút thuốc và tiêu thụ rượu bia, tăng hàm lượng cholesterol máu (mỡ trong máu cao), bệnh tiểu đường và kháng insulin, thừa cân và béo phì, lối sống thiếu vận động.

# Biện pháp phòng bệnh

- Hình thành lối sống: lành mạnh, tập thể dục, vận động, đủ giấc, hạn chế chất kích thích..
- Chế độ ăn uống hợp lý: đủ dinh dưỡng, đúng giờ đúng bữa, hạn chế thức ăn nhanh,...



# Mục lục

- 1 Tổng quan về bệnh
- 2 Tập dữ liệu
- 3 Phân lớp Naive Bayes
- 4 Mô hình hồi quy Ordinal Logistics
- 5 Kết luận
- 6 Tài liệu tham khảo

# Tập dữ liệu

**Bộ dữ liệu** UCI Heart Disease Data , các dữ liệu được lấy từ các bệnh nhân thuộc phòng khám

- Cleveland
- Viện tim mạch Hungari
- Trung tâm quản lý cựu chiến binh Long Beach
- Các bệnh viện Zurich và Basel (Thụy Sĩ)  
Có 920 mẫu 16 trường dữ liệu

# Tập dữ liệu

## Xử lý dữ liệu

Công thức chuẩn hóa dữ liệu:

$$x' = \frac{x - \mu(x)}{\sigma(x)}$$

Với:

$$\sigma(x) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x - \mu(x))^2}$$

Dữ liệu được khảo sát và xóa đi một số cột:

- Không sử dụng đến trong mô hình như : "ID", "Dataset"
- Loại bỏ những dữ liệu có khả năng sai lệch do xét nghiệm hoặc nhập liệu của biến "chol" (có giá trị bằng 0).

# Mục lục

- 1 Tổng quan về bệnh
- 2 Tập dữ liệu
- 3 Phân lớp Naive Bayes**
- 4 Mô hình hồi quy Ordinal Logistics
- 5 Kết luận
- 6 Tài liệu tham khảo

# Phân lớp Naive Bayes

## Định lý Bayes

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

- $P(A|B)$ : Xác suất của A xảy ra khi biết B đã xảy ra
- $P(B|A)$ : Xác suất của B xảy ra khi biết A đã xảy ra
- $P(A)$ : Xác suất của A xảy ra
- $P(B)$ : Xác suất của B xảy ra



# Phân lớp Naive Bayes

**Giả định của Naive Bayes:** Naive Bayes giả định rằng tất cả các đặc trưng đều độc lập và có đóng góp như nhau vào xác suất kết quả.

Xét bài toán phân lớp với  $m$  class ( $C_1, C_2, \dots, C_m$ ). Giả sử có một điểm dữ liệu  $X \in \mathbb{R}^n$  ( $X = (x_1, x_2, \dots, x_n)$ ).

$$\begin{aligned} y &= \arg \max_{C_j \in C} P(C_j | x_1, x_2, \dots, x_n) \\ &= \arg \max_{C_j \in C} \frac{P(x_1, x_2, \dots, x_n | C_j) P(C_j)}{P(x_1, x_2, \dots, x_n)} \\ &= \arg \max_{C_j \in C} P(x_1, x_2, \dots, x_n | C_j) P(C_j) \end{aligned}$$

# Mô hình Navie Bayes

Dựa vào giả định độc lập:

$$P(x_1, x_2, \dots, x_n | C_j) = P(x_1 | C_j)P(x_2 | C_j) \dots P(x_n | C_j)$$

Thay vào giá trị biểu thức trên:

$$y = \arg \max_{C_j \in C} P(C_j) \prod_{i=1}^n P(x_i | C_j)$$

# Gaussian Naive Bayes

Với mỗi chiều dữ liệu  $i$  và một class  $c$ :

$$x_i \sim \mathcal{N}(\mu_{ci}, \sigma_{ci}^2)$$

$$P(x_i|C) = \frac{1}{\sqrt{2\pi\sigma_{ci}^2}} \exp\left(-\frac{(x_i - \mu_{ci})^2}{2\sigma_{ci}^2}\right)$$

- $x_i$ : Giá trị của đặc trưng thứ  $i$ .
- $C$ : Nhãn của lớp.
- $\mu_{ci}$ : Giá trị trung bình của đặc trưng  $x_i$  trong lớp  $C$ .
- $\sigma_{ci}^2$ : Phương sai của đặc trưng  $x_i$  trong lớp  $C$ .

# Mục lục

- 1 Tổng quan về bệnh
- 2 Tập dữ liệu
- 3 Phân lớp Naive Bayes
- 4 Mô hình hồi quy Ordinal Logistics**
- 5 Kết luận
- 6 Tài liệu tham khảo

# Mô hình hồi quy Ordinal Logistic

## Khái niệm odds

Odds được định nghĩa là tỷ lệ giữa xác suất xảy ra một sự kiện so với xác suất không xảy ra sự kiện đó.

$$\text{odds} = \frac{P}{1-P} \quad (1)$$

Tương ứng trong hồi quy Ordinal Logistic: Giả sử  $Y$  là kết quả dự báo thứ tự với  $J$  danh mục thứ tự khi đó:

$$\text{odds} = \frac{P(Y \leq j|X)}{P(Y > j|X)} = \frac{P(Y \leq j|X)}{1 - P(Y \leq j|X)} \quad \text{với } j = 1, \dots, J - 1 \quad (2)$$

# Mô hình hồi quy Ordinal Logistic

**Khái niệm logit:** Logit được xác định là log cơ số tự nhiên của odds

$$\log \left( \frac{P(Y \leq j|X)}{P(Y > j|X)} \right) = \log \left( \frac{P(Y \leq j|X)}{1 - P(Y \leq j|X)} \right) = \text{logit}(P(Y \leq j|X)) \quad (3)$$

# Mô hình hồi quy Ordinal Logistic

## Hồi quy Ordinal Logistic

Hồi quy Ordinal Logistic được sử dụng để xác định mối quan hệ giữa một tập hợp các biến dự báo và một biến phụ thuộc có yếu tố thứ tự

$$\text{logit}(P(Y \leq j|X)) = \log \left( \frac{P(Y \leq j|X)}{1 - P(Y \leq j|X)} \right) = \alpha_j - \beta_1 x_1 - \dots - \beta_p x_p$$

Trong đó:

$\alpha_j$  là hệ số chặn ứng với mỗi lớp.

$\beta_1, \dots, \beta_p$  là các hệ số ứng với mỗi biến dự báo.

# Mô hình hồi quy Ordinal Logistic

## Các hệ số của mô hình

Biến	Hệ số	Độ lệch chuẩn	Giá trị W
age	0.28150	0.09364	3.0063
sex	1.03861	0.23271	4.4631
cp	0.62407	0.10278	6.0716
trestbps	0.05056	0.08744	0.5783
chol	-0.55104	0.08275	-6.6594
fbs	0.48154	0.21271	2.2638
restecg	0.20164	0.10444	1.9306
thalch	-0.27037	0.09617	-2.8113
exang	0.66870	0.18294	3.6554
oldpeak	0.60682	0.08746	6.9385
slope	0.18849	0.12095	1.5584
ca	0.75244	0.12489	6.0250
thal	0.10002	0.11742	0.8518

Hệ số chặn	Giá trị
$\alpha_0$	2.3353
$\alpha_1$	4.3003
$\alpha_2$	5.2951
$\alpha_3$	7.3956

Bảng: Hệ số chặn

Bảng: Hệ số của các biến



# Mô hình hồi quy Ordinal Logistic

## Kiểm định Wald

Được sử dụng để đánh giá xem từng hệ số của biến trong mô hình có ý nghĩa đối với mô hình hay không.

$$\text{Giả thuyết: } \begin{cases} H_0 : \hat{\beta}_i = 0 \\ H_1 : \hat{\beta}_i \neq 0 \end{cases}$$

Tiêu chuẩn kiểm định:

$$W = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \quad (4)$$

Trong đó:

- $\hat{\beta}$  là hệ số ước lượng của biến.
- $SE(\hat{\beta})$  là sai số chuẩn của hệ số ước lượng.

# Mô hình hồi quy Ordinal Logistic

Giá trị p-value được xác định là:

$$p\text{-value} = 2 \cdot P(Z > |W|) \quad (5)$$

Với mức ý nghĩa  $\alpha = 0.05$ :

- Nếu  $W > 1.96$  hoặc  $W < -1.96$  ứng với  $p\text{-value} < \alpha$ : Bác bỏ  $H_0$  (Biến có ý nghĩa đối với mô hình).
- Nếu  $-1.96 < W < 1.96$  ứng với  $p\text{-value} > \alpha$ : Chấp nhận  $H_0$  (Biến không có ý nghĩa đối với mô hình).

# Mục lục

- 1 Tổng quan về bệnh
- 2 Tập dữ liệu
- 3 Phân lớp Naive Bayes
- 4 Mô hình hồi quy Ordinal Logistics
- 5 Kết luận**
- 6 Tài liệu tham khảo

# Kết luận

Dựa vào Accuracy, Precision, Recall và F1-score để đánh giá hiệu suất phân loại của mô hình.

$$\textbf{Accuracy} = \frac{TN + TP}{TN + FN + FP + TP}$$

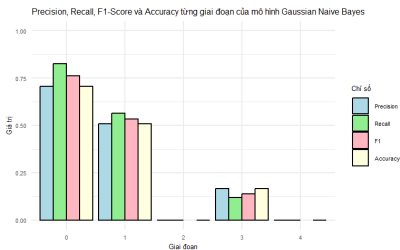
$$\textbf{Precision} = \frac{TP}{TP + FP}$$

$$\textbf{Recall} = \frac{TP}{TP + FN}$$

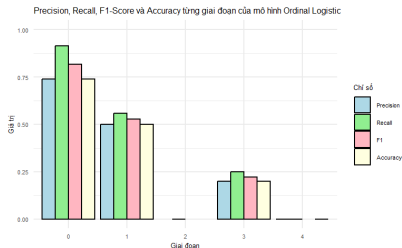
$$\textbf{F1-score} = 2 * \frac{\textit{Precision} * \textit{Recall}}{\textit{Precision} + \textit{Recall}}$$

# Kết luận

So sánh các giá trị Accuracy, Precision, Recall và F1-score của 2 mô hình.



Hình: Mô hình Gaussian Naive Bayes



Hình: Mô hình Ordinal Logistic

# Kết luận

Thông qua các hệ số của mô hình, chúng ta có thể xác định những yếu tố nào có ảnh hưởng mạnh nhất đến nguy cơ mắc bệnh, từ đó có những biện pháp can thiệp kịp thời và hiệu quả.

Việc ứng dụng mô hình không chỉ giúp nâng cao độ chính xác trong chẩn đoán mà còn góp phần vào việc cải thiện chất lượng chăm sóc sức khỏe, giảm thiểu tỷ lệ tử vong và biến chứng liên quan đến bệnh động mạch vành.

# Mục lục

- 1 Tổng quan về bệnh
- 2 Tập dữ liệu
- 3 Phân lớp Naive Bayes
- 4 Mô hình hồi quy Ordinal Logistics
- 5 Kết luận
- 6 Tài liệu tham khảo

## Tài liệu tham khảo

- [1 ] (N.d.). Nih.gov.Retrieved October 27, 2024, from <https://pmc.ncbi.nlm.nih.gov/articles/PMC11458606/>
- [2 ] Hasanah, S. H. (2022). Application of machine learning for heart disease classification using Naive Bayes. *Jurnal Matematika MANTIK*. <https://doi.org/10.15642/mantik.2022.8.1.68-77>
- [3 ] Chapter 12 Ordinal Logistic Regression | Companion to BER 642: Advanced Regression Methods. (2024). Retrieved 27 October 2024, from [https://bookdown.org/chua/ber642\\_advanced\\_regression/ordinal-logistic-regression.html](https://bookdown.org/chua/ber642_advanced_regression/ordinal-logistic-regression.html)



**Thanks for listening !!!**