1. **Breast Cancer by SEER**

Link: <https://www.kaggle.com/datasets/reihanenamdari/breast-cancer> (1\_Breast\_Cancer.csv)

Tập dữ liệu về bệnh nhân ung thư vú này được lấy từ bản cập nhật tháng 11 năm 2017 của Chương trình SEER của NCI, nơi cung cấp thông tin về thống kê ung thư dựa trên dân số. Bộ dữ liệu liên quan đến các bệnh nhân nữ bị ung thư vú ung thư biểu mô ống xâm nhập và ung thư biểu mô tiểu thùy (SEER trích dẫn chính mã lại mã mô học NOS 8522/3) được chẩn đoán vào năm 2006-2010. Những bệnh nhân có kích thước khối u không rõ, LN khu vực được kiểm tra, LN khu vực dương tính và những bệnh nhân có tháng sống sót dưới 1 tháng đã bị loại trừ. Và, 4024 bệnh nhân cuối cùng đã được đưa vào.

- **Age**: Độ tuổi.

**- Race**: Chủng tộc.

**- Marital State:** Tình trạng kết hôn.

**- T Stage:**

Giai đoạn T trong ung thư vú đề cập đến kích thước và mức độ lan rộng của khối u nguyên phát và là thành phần quan trọng của hệ thống phân giai đoạn TNM được sử dụng để xác định giai đoạn tổng thể của bệnh.

T1 (Tumor size 1): Khối u có kích thước tối đa là 2 centimet (cm) trong chiều lớn nhất.

T2 (Tumor size 2): Khối u có kích thước từ 2 cm đến 5 cm trong chiều lớn nhất.

T3 (Tumor size 3): Khối u có kích thước lớn hơn 5 cm trong chiều lớn nhất.

T4 (Tumor size 4): Có thể được chia thành các phụ giai đoạn T4a, T4b, T4c, và T4d:

T4a: Khối u lan ra ngoài vú, nhưng không tác động lên cơ bắp hoặc da.

T4b: Khối u lan ra ngoài vú và tác động lên cơ bắp.

T4c: Khối u lan ra ngoài vú và tác động lên da, có thể bao gồm sưng, đau, và vẻ ngoại hình bất thường của da.

T4d: Mọi giai đoạn của T4 có thể bao gồm sự lan truyền đến nút lymph cục.

**- N Stage:**

Giai đoạn N trong ung thư vú đề cập đến sự liên quan của các hạch bạch huyết ở nách (nách) và các khu vực khác gần vú,

đồng thời đóng vai trò quan trọng trong việc xác định giai đoạn tổng thể và tiên lượng của bệnh. Trong hệ thống phân loại giai đoạn ung thư vú theo AJCC 6th, giá trị N (Node) chỉ ra mức độ lan truyền của ung thư đến các nút lymph cục xung quanh. Dưới đây là mô tả tổng quan về các giá trị N:

N1 (Node 1): Sự lan truyền đến một hoặc nhiều nút lymph cục cục bộ.

N2 (Node 2): Sự lan truyền có thể bao gồm một trong các điều kiện sau:

Sự lan truyền đến 4-9 nút lymph cục cục bộ.

Sự lan truyền đến nút lymph cục xa cách vú, nhưng không phải là nút lymph cục bên đối diện.

N3 (Node 3): Sự lan truyền có thể bao gồm một trong các điều kiện sau:

Sự lan truyền đến 10 nút lymph cục cục bộ trở lên.

Sự lan truyền đến nút lymph cục bên đối diện của vú.

**- 6th Stage:**

Giai đoạn của ung thư vú theo hệ thống AJCC 6th được biểu diễn bằng các pha từ I đến IV, với mỗi giai đoạn được chia thành các phụ giai đoạn A, B, C, v.v., tùy thuộc vào một số yếu tố như kích thước của khối u, sự lan truyền đến nút lymph, và sự lan truyền đến các bộ phận khác của cơ thể.

Các giá trị:

IIA (Stage IIA): Đây là một giai đoạn trong hệ thống AJCC 6th và thường ám chỉ một mức độ lan truyền ở mức độ trung bình. Các yếu tố như kích thước của khối u và sự lan truyền đến nút lymph có thể đặc trưng cho giai đoạn này.

IIB (Stage IIB): Giai đoạn này thường ám chỉ một mức độ lan truyền ở mức độ cao hơn so với IIA. Điều này có thể bao gồm các yếu tố như kích thước lớn hơn của khối u hoặc sự lan truyền đến nút lymph xa hơn.

IIIA (Stage IIIA): Đây là một giai đoạn trong đó ung thư vú đã có mức độ lan truyền mạnh mẽ. Thông thường, các yếu tố như kích thước của khối u, sự lan truyền đến nút lymph cục, và vị trí của nút lymph cục bị ảnh hưởng đến việc xác định giai đoạn này. Ví dụ, giai đoạn IIIA có thể ám chỉ khối u lớn, có sự lan truyền đến nút lymph cục, nhưng không có sự lan truyền đến các cơ bản khác.

IIIB (Stage IIIB): Giai đoạn IIIB thường ám chỉ một mức độ lan truyền mạnh mẽ hơn so với giai đoạn IIIA, nhưng không đạt đến mức độ lan truyền của giai đoạn IIIC. Các yếu tố như kích thước của khối u, sự lan truyền đến nút lymph cục, và vị trí của nút lymph cục có thể được xem xét để xác định giai đoạn này.

IIIC (Stage IIIC): Đây là một giai đoạn nâng cao, trong đó sự lan truyền của ung thư đã mở rộng đến nút lymph cục xa và có thể liên quan đến các cơ bản lân cận khác. Giai đoạn IIIC thường là một giai đoạn nâng cao, chỉ ra một mức độ lan truyền rất mạnh mẽ của ung thư.

**- Differentiate:**

Điều này đề cập đến mức độ trưởng thành của các tế bào khối u. Các tế bào biệt hóa tốt giống với tế bào vú bình thường

và có tiên lượng tốt hơn, trong khi các tế bào kém biệt hóa thì bất thường và hung hãn hơn.

(Có thể)

**- Grade:**

Trong ung thư vú, thuật ngữ “độ” dùng để chỉ mức độ trưởng thành hoặc sự biệt hóa của các tế bào khối u. Về cơ bản,

nó là thước đo mức độ giống nhau của các tế bào ung thư với các tế bào vú bình thường.

Đánh giá tốc độ tế bào ung thư đang phân chia.

Điểm số cụ thể từ mỗi đặc điểm này sau đó được cộng lại để xác định bậc độ tổng thể, theo các phân loại thông thường như Bậc độ 1 (tốt), Bậc độ 2 (trung bình), và Bậc độ 3 (kém).

(Và nó có phân chia các trường hợp, mình sẽ tìm hiểu thêm.)

**- A Stage:**

Có 2 giá trị:

Regional - Một khối u đã lan rộng

Distant - Một khối u đã lan đến các bộ phận của cơ thể ở xa

**- Tumor Size:**

Kích thước khối u là yếu tố quan trọng quyết định giai đoạn và tiên lượng của bệnh ung thư vú. Nó được đo ở điểm rộng

nhất và thường được báo cáo bằng milimét (mm) hoặc centimét (cm).

(Nó cũng được chia ra các thành phần nhỏ hơn mình sẽ tìm hiểu sau.)

**- Estrogren Status:**

Estrogen đóng một vai trò quan trọng trong sự phát triển và tiến triển của ung thư vú. Hiểu được tình trạng thụ thể estrogen

(ER) của khối u là rất quan trọng để xác định các chiến lược điều trị hiệu quả nhất. Các thụ thể estrogen là các protein được tìm thấy trong tế bào vú. Khi estrogen liên kết với các thụ thể này, nó sẽ báo hiệu cho các tế bào phát triển và phân chia. Các khối u có số lượng thụ thể estrogen cao được gọi là dương tính với thụ thể estrogen (ER+). Những khối u này dựa vào estrogen để phát triển và có khả năng đáp ứng với các loại thuốc trị liệu bằng hormone ngăn chặn tác dụng của estrogen.

**- Progesterone Status:**

Các thụ thể progesterone là các protein được tìm thấy trong một số tế bào vú. Khi progesterone liên kết với các thụ thể này, nó có thể thúc đẩy sự phát triển và phân chia tế bào. Các khối u có số lượng thụ thể progesterone cao được gọi là dương tính với thụ thể progesterone (PR+). Những khối u này có thể bị ảnh hưởng bởi progesterone và có khả năng đáp ứng với liệu pháp hormone nhắm vào cả estrogen và progesterone.

**- Regional Node Examined:**

Trong ung thư vú, việc kiểm tra hạch vùng tùy thuộc vào giai đoạn và vị trí của khối u

**- Reginol Node Positive**

**- Survival Months Status**

1. **Breast Cancer Wisconsin (Diagnostic) Data Set**

Link: <https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data> (2\_Breast\_Cancer.csv)

Các đặc điểm được tính toán từ hình ảnh số hóa của chọc hút bằng kim nhỏ (FNA) của khối vú. Chúng mô tả đặc điểm của nhân tế bào có trong hình ảnh. n không gian 3 chiều được mô tả trong: [K. P. Bennett và O. L. Mangasarian: "Phân biệt lập trình tuyến tính mạnh mẽ của hai tập hợp tuyến tính không thể tách rời", Phương pháp tối ưu hóa và phần mềm 1, 1992, 23-34].

**Attribute Information:**

1) ID number

2) Diagnosis (M = malignant, B = benign)

3-32)

Mười tính năng có giá trị thực được tính toán cho mỗi nhân tế bào:

a) **radius** (mean of distances from center to points on the perimeter) - bán kính (trung bình khoảng cách từ tâm đến các điểm trên chu vi)

b) **texture** (standard deviation of gray-scale values) - kết cấu (độ lệch chuẩn của các giá trị thang độ xám)

Texture trong ung thư vú là một đặc trưng được sử dụng để phân tích biến đổi của các mẫu hình ảnh y tế. Để hiểu rõ hơn về texture, chúng ta có thể nói về độ lệch chuẩn (standard deviation) của các giá trị xám (gray-scale values). Trong hình ảnh y tế, các giá trị xám thường biểu thị mức độ sáng tối của các pixel. Độ lệch chuẩn là một chỉ số thống kê dùng để đo sự phân tán của dữ liệu. Trong trường hợp này, độ lệch chuẩn của các giá trị xám trong hình ảnh ung thư vú được sử dụng để đánh giá sự biến đổi trong mức độ sáng tối của các khu vực khác nhau trên hình ảnh. Thông thường, các vùng ung thư có xu hướng có texture không đều, không đồng nhất và thay đổi nhiều so với các vùng bình thường trong hình ảnh. Điều này có thể được phản ánh bằng độ lệch chuẩn lớn hơn của các giá trị xám trong các khu vực ung thư so với các khu vực bình thường.

c) **perimeter** – Chu vi

d) **area** – diện tích

e) **smoothness** (local variation in radius lengths) - độ mịn (thay đổi cục bộ về độ dài bán kính)

Smoothness (độ mịn) trong ung thư vú là một đặc trưng hình học được sử dụng để đánh giá sự biến thiên cục bộ trong độ dài bán kính của các vùng trong hình ảnh y tế của ung thư vú.

Để hiểu rõ hơn về smoothness, chúng ta có thể nói về sự biến thiên cục bộ trong độ dài bán kính của các vùng trong hình ảnh ung thư vú. Thông thường, trong ung thư vú, các vùng bất thường hoặc ung thư thường có xu hướng có độ biến thiên cục bộ cao hơn so với các vùng bình thường.

**Smoothness** được đo bằng cách xác định sự khác biệt giữa các bán kính của các vùng lân cận nhau. Nếu sự khác biệt này nhỏ, tức là các vùng có độ mịn cao, có xu hướng ít biến đổi và liền mạch hơn. Ngược lại, nếu sự khác biệt lớn, tức là các vùng có độ mịn thấp, có xu hướng biến đổi đột ngột và không liền mạch.

f) **compactness** (perimeter^2 / area - 1.0) - nhỏ gọn (chu vi ^ 2 / diện tích - 1.0)

Nếu một vùng có compactness gần với 0, điều đó cho thấy chu vi bình phương và diện tích của vùng tương đối tương đồng, tức là vùng đó có hình dạng gần như là một hình tròn hoặc hình vuông. Tuy nhiên, trong trường hợp ung thư vú, các vùng bất thường thường có compactness cao hơn. Điều này có nghĩa là chu vi bình phương của vùng lớn hơn đáng kể so với diện tích, cho thấy hình dạng của vùng không gọn gàng và có nhiều biến đổi.

g) **concavity** (severity of concave portions of the contour) - lõm (mức độ nghiêm trọng của các phần lõm của đường viền)

Concavity (độ lõm) trong ung thư vú là một đặc trưng hình học được sử dụng để đánh giá mức độ lõm của các phần lõm trên đường viền của một vùng trong hình ảnh y tế.

Để hiểu rõ hơn về concavity, chúng ta có thể xem xét cách đo lường mức độ lõm của các phần trên đường viền. Trong trường hợp ung thư vú, các phần lõm thường là các vùng không đều và không liền mạch trên đường viền của khối u. Concavity được sử dụng để đo lường mức độ lõm này, và thường được đánh giá dựa trên các đặc trưng hình học của đường viền của vùng.

h) **concave points** (number of concave portions of the contour) - điểm lõm (số phần lõm của đường viền)

Concave points (số lõm) trong ung thư vú là một đặc trưng hình học được sử dụng để đếm số lượng điểm lõm trên đường viền của một vùng trong hình ảnh y tế.

Để hiểu rõ hơn về concave points, chúng ta có thể xem xét cách đếm số lượng điểm lõm trên đường viền của một vùng. Trong trường hợp ung thư vú, các điểm lõm thường là các điểm trên đường viền của khối u mà có hình dạng lõm vào bên trong.

Số lượng concave points được sử dụng để đo lường mức độ lõm của đường viền và có thể cho thấy sự không đều và không liền mạch của vùng ung thư.

i) **symmetry** – sự đối xứng

j) **fractal dimension** ("coastline approximation" - 1) - kích thước fractal ("xấp xỉ đường bờ " - 1)

Fractal dimension (kích thước fraktal) trong ung thư vú là một đặc trưng hình học được tính dựa trên khái niệm "xấp xỉ đường bờ" của một vùng trong hình ảnh y tế.

Để hiểu rõ hơn về fractal dimension, chúng ta có thể xem xét cách xấp xỉ đường bờ của vùng. Trong trường hợp ung thư vú, đường bờ của vùng thường có hình dạng không đều, rối ren, và có nhiều chi tiết phức tạp.

Fractal dimension được tính bằng cách áp dụng khái niệm "xấp xỉ đường bờ" vào vùng. Đầu tiên, đường bờ của vùng được xem như đường bờ của một hình fraktal, trong đó các chi tiết nhỏ hơn được lặp lại trên các tỷ lệ khác nhau. Sau đó, fractal dimension được tính bằng cách áp dụng phương pháp tính toán kích thước fraktal vào đường bờ của vùng.