# **1.Dataset**

* Dataset là một tập dữ liệu đa trung tâm (UE‐MICAP) bao gồm 19,464 ảnh siêu âm được thu thập từ tháng 1/2014 đến tháng 12/2022. Dữ liệu được tổng hợp từ 8 bệnh viện lớn ở Trung Quốc (The First Affiliated Hospital of Sun Yat‐sen University, The Third/Sixth/Seventh Affiliated Hospital of Sun Yat‐sen University, The First Affiliated Hospital of Guangzhou Medical University, The First Affiliated Hospital of Guangxi Medical University, Foshan Sanshui District People’s Hospital, và West China Xiamen Hospital of Sichuan University).
* Tiêu chí:
  + Bao gồm: Bệnh nhân có nguy cơ HCC (bao gồm có chẩn đoán lâm sàng, bằng chứng hình ảnh xơ gan, hoặc xác nhận bằng bệnh học), nam trên 40 tuổi và nữ trên 50 tuổi với tiền sử viêm gan virus, và những bệnh nhân đã được sàng lọc siêu âm với tổn thương được xác nhận.
  + Loại trừ: Người dưới 18 tuổi, ảnh có chất lượng không đạt yêu cầu, hoặc ảnh chứa cả dấu hiệu lành tính và ác tính.
* Vì lý do liên quan đến bảo mật thông tin bệnh nhân, dataset gốc có thể không được công khai hoàn toàn.

# **2. Input, Output, Metrics, Phương pháp và Cải tiến của mô hình**

* Input:
  + Dữ liệu đầu vào là ảnh siêu âm được chuyển đổi từ DICOM sang PNG bằng OpenCV.
  + Tiền xử lý bao gồm việc loại bỏ thông tin văn bản (như thông tin bệnh nhân, ngày tháng, mã thiết bị,…) sử dụng mô hình YOLO nhằm bảo vệ tính ẩn danh và tiêu chuẩn hóa ảnh cho quá trình đào tạo.
* Output:
  + Mô hình HSQformer trả về dự đoán chẩn đoán ung thư gan (phân loại giữa các ca lành tính và ác tính) với đầu ra ở dạng xác suất hoặc nhãn phân loại.
* Metrics (chỉ số đánh giá):
  + Accuracy:
    - Là tỷ lệ phần trăm của các dự đoán đúng trên tổng số mẫu.
    - Chỉ số này cho biết tổng thể mô hình có thể phân loại chính xác bao nhiêu phần trăm các ca.
  + Precision:
    - Là tỷ lệ phần trăm các dự đoán dương tính trên tổng số các dự đoán được gán nhãn dương tính.
    - Chỉ số này rất quan trọng trong y tế vì nó phản ánh mức độ tin cậy của các ca mà mô hình dự đoán là có bệnh (ví dụ ung thư).
  + Recall:
    - Là tỷ lệ phần trăm các ca dương tính thật (true positives) được mô hình nhận diện trên tổng số ca dương tính thực tế.
    - Recall quan trọng trong các ứng dụng y tế vì nó cho biết mô hình có khả năng phát hiện bệnh như thế nào (đặc biệt cần tránh bỏ sót các ca bệnh ác tính).
  + F1 Score:
    - Là trung bình điều hòa của precision và recall, cung cấp một số đo cân bằng giữa độ chính xác và khả năng phát hiện của mô hình.
    - F1 score càng cao, nghĩa là mô hình vừa có độ chính xác cao vừa có khả năng nhận diện tốt các ca dương tính.
    - Chỉ số này hữu ích khi ta cần một đánh giá tổng thể, nhất là trong các trường hợp mà có sự mất cân bằng giữa số lượng mẫu của các lớp.
  + AUC (Diện tích dưới đường cong ROC):
    - ROC (Receiver Operating Characteristic) là đường cong biểu diễn mối quan hệ giữa recall (tỉ lệ dương tính thật) và false positive rate (tỉ lệ âm tính bị nhầm thành dương tính) ở các ngưỡng khác nhau.
    - AUC là diện tích dưới đường cong này; giá trị AUC càng cao (trong khoảng từ 0 đến 1), mô hình càng có khả năng phân biệt giữa các lớp (ví dụ, phân biệt giữa ca ung thư và không ung thư).
    - AUC cung cấp cái nhìn tổng quát về hiệu suất phân loại của mô hình, bất kể ngưỡng phân loại được chọn.
* **Phương pháp:**
  + **Kiến trúc:** HSQformer là một kiến trúc mới dựa trên việc kết hợp các đặc trưng của CNN (sử dụng ConvNext) và Vision Transformer (sử dụng SwinTransformer) thông qua một cấu trúc truy vấn phân cấp (Hierarchical Sparse Query Transformer).
  + **ConvNext (CNN):** Đây là một kiến trúc dựa trên các Convolutional Neural Networks, chuyên trích xuất các đặc trưng cục bộ như kết cấu, vi cấu trúc của ảnh. CNN rất hiệu quả trong việc nhận dạng các mẫu nhỏ và chi tiết trong ảnh.
  + **SwinTransformer (ViT):** Đây là một biến thể của Vision Transformer, chuyên trích xuất các đặc trưng tổng quát và mô hình hóa mối quan hệ giữa các vùng khác nhau của ảnh. SwinTransformer cho phép mạng hiểu được bối cảnh tổng thể của hình ảnh.
  + **Cách tiếp cận:**
    - Sử dụng biểu diễn không gian tiềm ẩn (latent space) để giảm sự dư thừa của đặc trưng.
    - Tích hợp các khối Attention (Cross-Attention và Self-Attention) cùng với mô hình Mixture-of-Experts (MOE) nhằm phân bổ tính toán một cách có chọn lọc, chỉ kích hoạt các experts cần thiết, từ đó cải thiện hiệu quả tính toán và khả năng mở rộng.
* **Cải tiến và Thành công:**
  + **So với các phương pháp trước đây:** Các phương pháp kết hợp CNN và ViT thông thường (late fusion) thường gặp phải vấn đề dư thừa đặc trưng. HSQformer khắc phục hạn chế này bằng cách áp dụng chiến lược truy vấn phân cấp, từ đó vừa khai thác được các đặc trưng cục bộ (low-level) vừa trích xuất các đặc trưng tổng quát (high-level) một cách hiệu quả mà không cần thay đổi cấu trúc mạng phức tạp.
  + **Thành công:** HSQformer-B đã đạt được hiệu suất vượt trội trong việc phân loại HCC so với các mô hình hiện trạng (như ConvNext, SwinTransformer) và thậm chí so sánh được với hiệu năng chẩn đoán của các bác sĩ chuyên khoa (radiologists), đặc biệt là vượt trội so với bác sĩ trẻ (junior radiologists) và tương đương hoặc vượt trội so với bác sĩ nhiều kinh nghiệm (senior radiologists).

A black and white text with red text

AI-generated content may be incorrect.

Bảng 1. Bảng so sánh giữa các mô hình cũ và HSQformer

A table with numbers and symbols

AI-generated content may be incorrect.

Bảng 2. Bảng só sánh giữa chuẩn đoán của bác sĩ so với mô hình HSQformer