# [1] Đọc paper, khảo sát bài toán liên quan

## 1.1 Tóm tắt paper

- Paper WHO (1999) đưa ra định nghĩa, tiêu chuẩn chẩn đoán và phân loại bệnh tiểu đường.  
- Định nghĩa: Tiểu đường là rối loạn chuyển hóa đặc trưng bởi tăng đường huyết mạn tính do thiếu insulin, giảm tác dụng insulin hoặc cả hai.  
- Tiêu chuẩn chẩn đoán:   
 + FPG ≥ 7.0 mmol/L (126 mg/dl).  
 + 2h OGTT ≥ 11.1 mmol/L (200 mg/dl).  
 + IFG (6.1 – 6.9 mmol/L), IGT (7.8 – 11.0 mmol/L).  
- Phân loại: Type 1, Type 2, Other specific types, Gestational diabetes.  
- Hội chứng chuyển hóa: cụm yếu tố nguy cơ (béo bụng, tăng huyết áp, rối loạn lipid, kháng insulin).

## 1.2 Khảo sát bài toán liên quan

Từ nội dung paper, có thể đặt ra các bài toán liên quan đến phân tích dữ liệu y tế:  
- Phân loại bệnh nhân vào nhóm: Bình thường, IFG/IGT (nguy cơ), hoặc Tiểu đường.  
- Phân tích mối liên hệ giữa chỉ số glucose, BMI, huyết áp và khả năng mắc bệnh.  
- Phát hiện bệnh nhân có nguy cơ cao để đưa ra khuyến nghị can thiệp sớm.  
- Tìm hiểu sự phân bố bệnh theo nhân khẩu học (giới tính, độ tuổi).  
- Phát hiện giá trị ngoại lệ (glucose bất thường) và dữ liệu thiếu, để đảm bảo độ chính xác của phân tích.  
  
Như vậy, bài toán liên quan ở đây chính là: \*Khám phá và phân tích dữ liệu lâm sàng để phát hiện, phân loại, và dự đoán nguy cơ tiểu đường dựa trên chuẩn chẩn đoán WHO.\*

# [2] Phân tích khám phá dữ liệu (EDA)

## 2.1 Xác định tính chất dữ liệu

- Biến định lượng: FPG, OGTT, HbA1c, BMI, Huyết áp, Triglyceride, HDL...  
- Biến phân loại: Loại tiểu đường, giới tính, nhóm tuổi.  
- Đặc điểm: dữ liệu y tế thường chứa giá trị thiếu và ngoại lệ.

## 2.2 Phân tích đơn biến

- Kiểm tra phân phối các biến FPG, OGTT bằng histogram, boxplot.  
- So sánh với ngưỡng WHO:  
 + FPG ≥ 7.0 mmol/L → Diabetes.  
 + 6.1 ≤ FPG < 7.0 mmol/L → IFG.  
 + OGTT ≥ 11.1 mmol/L → Diabetes.  
 + 7.8 ≤ OGTT < 11.1 mmol/L → IGT.

## 2.3 Phân tích đa biến

- Phân tích tương quan giữa glucose, HbA1c, BMI, huyết áp.  
- Dùng scatter plot, heatmap để phát hiện mối quan hệ.  
- Ví dụ: BMI cao thường đi kèm nguy cơ Type 2 Diabetes.  
- Có thể dùng phân cụm để nhóm bệnh nhân theo yếu tố nguy cơ.

## 2.4 Xác định giá trị thiếu và mẫu bất thường

- Một số xét nghiệm không được thực hiện (thiếu OGTT, HbA1c).  
- Thiếu dữ liệu nhân khẩu học (tuổi, giới).  
- Mẫu bất thường: bệnh nhân có glucose cực cao/thấp, có thể do nhập sai hoặc ca cấp cứu đặc biệt.

## 2.5 Xác định ngoại lệ

- Dùng IQR (Interquartile Range) hoặc Z-score để phát hiện ngoại lệ.  
- Ngoại lệ cần được phân biệt:   
 + Sai số đo lường.  
 + Trường hợp bệnh hiếm (ví dụ: MODY, Wolfram syndrome).  
- So sánh với bảng tiêu chuẩn WHO (Bảng 1 trong paper) để xác định bất thường.

# Kết luận

- Paper WHO 1999 cung cấp chuẩn chẩn đoán và phân loại rõ ràng cho tiểu đường.  
- EDA giúp khám phá dữ liệu, phân loại bệnh nhân và phát hiện nguy cơ.  
- Type 2 chiếm đa số, nhóm IFG/IGT cần được theo dõi để phòng ngừa.  
- Cần xử lý giá trị thiếu và ngoại lệ để đảm bảo phân tích chính xác.  
- Kết quả EDA là cơ sở để xây dựng mô hình dự đoán và ra quyết định y tế.