# ***ADAP Neural Network Forecasts Diabetes Mellitus Onset***

Bài báo cung cấp một cái nhìn tổng quan về việc sử dụng thuật toán học ADAP (một mô hình mạng thần kinh sơ khai) để dự báo sự khởi phát của bệnh đái tháo đường (DM) trong vòng năm năm ở một nhóm người có nguy cơ cao là người da đỏ Pima. Các tác giả giải thích rằng các mô hình mạng thần kinh đang được quan tâm trở lại và rất có giá trị trong các bài toán nhận dạng mẫu mà các kỹ thuật thống kê truyền thống gặp khó khăn. Nghiên cứu này đã sử dụng dữ liệu từ 8 biến đầu vào—bao gồm cả một chức năng di truyền được phát triển mới gọi là Hàm Phả hệ Đái tháo đường (DPF)—trên một tập huấn luyện gồm 576 trường hợp và tập dự báo gồm 192 trường hợp. Kết quả cho thấy thuật toán đạt được độ nhạy và độ đặc hiệu là 76% tại điểm phân biệt 0.448, và các nỗ lực đang được tiến hành để so sánh hiệu suất của ADAP với mô hình hồi quy logistic và perceptron tuyến tính.

**Tóm tắt (Abstract)**

Phần này giới thiệu tổng quan về nghiên cứu, trong đó các tác giả đã thử nghiệm một mô hình mạng nơ-ron ban đầu có tên là **ADAP** để dự báo sự khởi phát của bệnh tiểu đường trong một quần thể người da đỏ Pima có nguy cơ cao. Hiệu suất của thuật toán được phân tích bằng các thước đo tiêu chuẩn cho các xét nghiệm lâm sàng, bao gồm độ nhạy và độ đặc hiệu. Kết quả cho thấy điểm giao nhau của độ nhạy và độ đặc hiệu là 0,76. Các tác giả cũng cho biết họ đang so sánh kết quả của ADAP với các mô hình khác như hồi quy logistic và perceptron tuyến tính.

**1. Giới thiệu (Introduction)**

Phần giới thiệu đặt ra bối cảnh và lý do thực hiện nghiên cứu.

• **Vấn đề**: Các phương pháp thống kê tiêu chuẩn có thể không hiệu quả trong việc dự báo và phân biệt trong y học khi kích thước mẫu nhỏ, không rõ dạng của mối quan hệ chức năng cơ bản, hoặc khi các mối quan hệ này liên quan đến các tương tác phức tạp và tương quan chéo.

• **Giải pháp đề xuất**: Trong những tình huống như vậy, các mô hình mạng nơ-ron mang lại nhiều hứa hẹn.

• **Nghiên cứu cụ thể**: Các tác giả đã chọn kiểm tra thuật toán ADAP, một thuật toán được phát triển từ năm 1961, có khả năng khám phá các mẫu trong dữ liệu thông qua một tập dữ liệu huấn luyện. Nghiên cứu áp dụng thuật toán này để dự báo sự khởi phát của bệnh tiểu đường không phụ thuộc insulin (DM) trong vòng 5 năm trên dữ liệu của người da đỏ Pima.

• **Quy trình và kết quả sơ bộ**: Thuật toán được huấn luyện trên 576 trường hợp và sau đó được dùng để dự báo cho 192 trường hợp thử nghiệm. Khi buộc phải đưa ra kết luận cho tất cả các trường hợp thử nghiệm, ADAP đạt độ nhạy và độ đặc hiệu là 76%.

**2. Nguyên liệu và Phương pháp (Materials and Methods)**

Phần này mô tả chi tiết cách thức nghiên cứu được tiến hành.

• **2.1 Quần thể nghiên cứu (Study Population)**: Nghiên cứu tập trung vào người da đỏ Pima, một quần thể đã được Viện Quốc gia về Bệnh tiểu đường, Tiêu hóa và Bệnh thận nghiên cứu liên tục từ năm 1965 do tỷ lệ mắc bệnh tiểu đường cao. Bệnh tiểu đường được chẩn đoán theo tiêu chuẩn của Tổ chức Y tế Thế giới (WHO).

• **2.2 Lựa chọn biến số (Selection of the Variables)**: **Tám biến số đầu vào** đã được chọn để dự báo sự khởi phát của bệnh tiểu đường ở phụ nữ Pima trong vòng 5 năm. Các biến này được chọn vì chúng là những yếu tố nguy cơ đáng kể. Các biến đó là:

1. Số lần mang thai.

2. Nồng độ Glucose huyết tương sau 2 giờ trong xét nghiệm dung nạp glucose đường uống (GTIT).

3. Huyết áp tâm trương (mm Hg).

4. Độ dày nếp gấp da cơ tam đầu (mm).

5. Insulin huyết thanh sau 2 giờ (Uh/ml).

6. Chỉ số khối cơ thể (BMI).

7. Hàm phả hệ bệnh tiểu đường (DPF).

8. Tuổi (năm).

◦ **Hàm phả hệ bệnh tiểu đường (DPF)** là một hàm do các tác giả phát triển để tổng hợp lịch sử bệnh tiểu đường của người thân và mối quan hệ di truyền của họ với đối tượng nghiên cứu. Nó đo lường ảnh hưởng di truyền dự kiến từ những người thân bị ảnh hưởng và không bị ảnh hưởng đến nguy cơ mắc bệnh tiểu đường của đối tượng.

• **2.3 Lựa chọn trường hợp (Case Selection)**: Các trường hợp được rút ra từ một nhóm các kỳ khám đáp ứng các tiêu chí cụ thể, bao gồm đối tượng là nữ, từ 21 tuổi trở lên, và tại thời điểm khám chưa bị tiểu đường. Tổng cộng 768 kỳ khám đã được chọn, sau đó được chia ngẫu nhiên thành 576 trường hợp cho tập huấn luyện và 192 trường hợp cho tập dự báo.

**3. Mô tả Thuật toán (Description of the Algorithm)**

Phần này giải thích chi tiết về thuật toán ADAP.

• **Tổng quan về ADAP**: ADAP là một quy trình học thích ứng, tạo ra các "chất tương tự" kỹ thuật số của các thiết bị giống perceptron. Nó học bằng cách thực hiện các điều chỉnh nội bộ khi dự báo của nó không chính xác. Một trong những khác biệt chính của ADAP so với các mạng nơ-ron truyền thống hơn là cách tổ chức các đơn vị cảm biến (sensor units) thành các "phân vùng" (partitions).

• **Khởi tạo ADAP (Initialization of ADAP)**: Để chuẩn bị dữ liệu đầu vào, các giá trị của mỗi biến phải được chia thành các loại (categories). Mạng nơ-ron được tạo ra bao gồm một ma trận cố định (mạng lưới các cảm biến và đơn vị kết hợp) và một mảng biến đổi (các kết nối có trọng số thay đổi).

• **Ma trận cố định (The Fixed Matrix)**: Chứa 8 phân vùng tương ứng với 8 biến đầu vào. Mỗi phân vùng được chia thành các hàng, tương ứng với một phạm vi giá trị của biến. Ma trận cố định trong nghiên cứu này chứa 100.000 cột (tương tự các đơn vị kết hợp).

• **Mảng biến đổi (The Variable Array)**: Gồm một hàng duy nhất chứa một ô cho mỗi cột trong ma trận cố định. Mỗi ô chứa một giá trị tương ứng với sức mạnh của kết nối giữa đơn vị kết hợp và bộ phản hồi (responder). Các giá trị này ban đầu được đặt bằng 0 và thay đổi trong quá trình học.

• **Quá trình dự báo/học (The Forecasting/Learning Process)**:

1. Một "trường hợp" được đọc vào, kích hoạt một tập hợp các hàng trong ma trận cố định.

2. ADAP kiểm tra từng cột. Nếu số lượng kết nối dương (+) trong các hàng được kích hoạt của một cột lớn hơn hoặc bằng một ngưỡng (trong nghiên cứu này là 4), thì cột đó được kích hoạt.

3. Tổng các giá trị trong mảng biến đổi tương ứng với các cột được kích hoạt sẽ tạo thành **DỰ BÁO (FORECAST)** cho trường hợp đó.

4. Nếu ở chế độ học, ADAP sẽ so sánh dự báo với giá trị thực tế (0 nếu không mắc bệnh, 1 nếu mắc bệnh) và tính toán một giá trị sai số.

5. Giá trị sai số này được cộng vào mỗi ô được gắn cờ trong mảng biến đổi, qua đó điều chỉnh các trọng số cho lần dự báo tiếp theo.

**4. Kết quả và Kết luận (Results and Conclusions)**

Phần cuối cùng trình bày kết quả của nghiên cứu.

• Sau khi huấn luyện với 576 trường hợp, thuật toán ADAP được chuyển sang chế độ không học để dự báo.

• Khi sử dụng giá trị ngưỡng là 0,448 để phân biệt, **độ nhạy và độ đặc hiệu đều đạt 76%**.

• Một đường cong ROC (Receiver Operating Characteristic) được vẽ để thể hiện sự ảnh hưởng của việc thay đổi điểm phân biệt.

• Các tác giả kết luận rằng họ đang tiếp tục kiểm tra cùng một bộ dữ liệu bằng cách sử dụng hồi quy logistic và mô hình perceptron tuyến tính để so sánh kết quả