ỦY BAN NHÂN DÂN THÀNH PHỐ HỒ CHÍ MINH

**TRƯỜNG ĐẠI HỌC SÀI GÒN**

****

**BÁO CÁO**

Học phần: Phân tích dữ liệu

**ĐỀ TÀI:  
Phân tích khám phá về bệnh đái tháo đường**

Giảng viên hướng dẫn: ThS. Đỗ Như Tài

Thành viên nhóm:

Nguyễn Vũ Hào – MSSV: 3122410098

Trần Duy Khương – MSSV: 3122410192

Võ Thị Thương – MSSV: 3122410408

Tp.Hồ Chí Minh, 2025

|  |  |  |
| --- | --- | --- |
| **Họ và Tên** | **Mã số sinh viên** | **Phân công công việc** |
| Nguyễn Vũ Hào | 3122410098 |  |
| Trần Duy Khương | 3122410192 |  |
| Võ Thị Thương | 3122410408 |  |

**CHƯƠNG I: ĐỊNH NGHĨA, CHUẨN ĐOÁN VÀ PHÂN LOẠI BỆNH ĐÁI THÁO ĐƯỜNG VÀ CÁC BIẾN CHỨNG**

1. **Giới thiệu**

Thuật ngữ “đái tháo đường (diabetes mellitus)” mô tả một rối loạn chuyển hóa có nhiều nguyên nhân, đặc trưng bởi tình trạng tăng đường huyết mạn tính (chronic hyperglycaemia), kèm theo rối loạn chuyển hóa carbohydrate, lipid và protein. Nguyên nhân chủ yếu là do thiếu hụt tiết insulin, giảm tác dụng insulin, hoặc cả hai.

Các hậu quả lâu dài của bệnh bao gồm:  
- Tổn thương, rối loạn chức năng và suy yếu nhiều cơ quan.  
- Các biến chứng đặc hiệu: bệnh võng mạc (có thể gây mù), bệnh thận (dẫn tới suy thận), bệnh thần kinh (tăng nguy cơ loét bàn chân, cắt cụt chi, biến dạng khớp, rối loạn thần kinh tự động, rối loạn tình dục).  
- Nguy cơ tăng cao mắc bệnh tim mạch, bệnh mạch ngoại vi và bệnh mạch máu não.

Bệnh có thể biểu hiện triệu chứng điển hình như: khát nhiều, tiểu nhiều, sụt cân, nhìn mờ. Trong thể nặng có thể xuất hiện nhiễm toan ceton hoặc tình trạng tăng thẩm thấu không ceton, dẫn tới hôn mê, thậm chí tử vong nếu không điều trị. Tuy nhiên, nhiều trường hợp bệnh diễn tiến âm thầm, triệu chứng không rõ ràng, khiến tình trạng tăng đường huyết kéo dài trước khi được chẩn đoán.  
Về cơ chế bệnh sinh, có nhiều tiến trình khác nhau:  
- Quá trình phá hủy tế bào beta tụy → thiếu insulin tuyệt đối.  
- Quá trình gây đề kháng insulin → giảm tác động của insulin ở mô đích.

1. **Định nghĩa và tiêu chuẩn chuẩn đoán bệnh đái tháo đường và các dạng rối loạn dung nạp glucose**
   1. **Định nghĩa**

Đái tháo đường là một rối loạn chuyển hóa do nhiều nguyên nhân, đặc trưng bởi tình trạng tăng đường huyết mạn tính, kèm theo rối loạn chuyển hóa carbohydrate, mỡ và protein. Hậu quả là tổn thương lâu dài và suy chức năng của nhiều cơ quan.

* 1. **Chuẩn đoán và tiêu chuẩn chuẩn đoán**
     1. **Chuẩn đoán**

- Khi chẩn đoán đái tháo đường, bác sĩ cần chắc chắn và xác nhận vì chẩn đoán này ảnh hưởng suốt đời bệnh nhân.  
- Trường hợp bệnh nhân có triệu chứng rõ rệt kèm tăng đường huyết nặng, chẩn đoán đơn giản hơn so với người không triệu chứng.  
- Tăng đường huyết cấp tính do nhiễm trùng, chấn thương, stress tuần hoàn có thể tạm thời, không được coi là chẩn đoán đái tháo đường.  
- Với người không có triệu chứng, không bao giờ được chẩn đoán chỉ dựa trên một xét nghiệm đơn lẻ. Cần có ít nhất một xét nghiệm khác.  
- Nếu chưa rõ ràng, bệnh nhân cần được theo dõi và xét nghiệm định kỳ.  
- Các yếu tố cần cân nhắc: chủng tộc, tiền sử gia đình, tuổi, béo phì, bệnh lý kèm theo.  
- HbA1c có thể phản ánh đường huyết trung bình nhiều tuần, nhưng chưa đủ tiêu chuẩn hóa để thay thế OGTT trong chẩn đoán.

* + 1. **Đái tháo đường ở trẻ em**

- Thường khởi phát với triệu chứng nặng, đường huyết rất cao, nhiều glucose và ceton trong nước tiểu.  
- Thường được chẩn đoán ngay bằng xét nghiệm đường huyết, sau đó cần tiêm insulin khẩn cấp.  
- OGTT không cần thiết trong những trường hợp này.  
- Một số trẻ biểu hiện nhẹ hơn có thể cần OGTT hoặc xét nghiệm đường huyết lúc đói.

* 1. **Tiêu chuẩn chuẩn đoán**

- Chẩn đoán lâm sàng thường dựa vào triệu chứng: khát nhiều, tiểu nhiều, nhiễm trùng tái diễn, sụt cân không rõ nguyên nhân, hôn mê.  
- Đường huyết cao trong một lần xét nghiệm vượt ngưỡng chẩn đoán đã đủ để xác định bệnh.  
- Nếu kết quả nằm trong vùng nghi ngờ, có thể cần làm OGTT.  
- Ở người lớn và trẻ em, ngưỡng chẩn đoán là như nhau.

**2.3.1. Thay đổi đường huyết lúc đói**

- Trước đây: ≥7.8 mmol/L (140 mg/dl).  
 - Mới: ≥7.0 mmol/L (126 mg/dl) với huyết tương, hoặc ≥6.1 mmol/L

(110mg/dl) với máu toàn phần.  
 - Lý do: nhiều nghiên cứu cho thấy nguy cơ biến chứng vi mạch tăng rõ khi đường huyết lúc đói ≥7.0 mmol/L.

**2.3.2. Trong nghiên cứu dịch tễ học**

- Có thể phân loại dựa trên đường huyết lúc đói hoặc sau 2h làm OGTT.  
- Nhiều nghiên cứu chỉ dùng giá trị sau 2h OGTT.  
- Tuy nhiên, hai giá trị này không hoàn toàn trùng nhau.

**2.3.3. Trong chuẩn đoán cá thể**

- Không được chẩn đoán dựa trên một lần đo.  
- Cần xác nhận bằng triệu chứng hoặc một xét nghiệm khác.  
- Tốt nhất nên xét nghiệm cả lúc đói và sau 2h OGTT.

1. **Phân loại**
   1. **Các phân loại trước đó**

Phân loại đầu tiên được chấp nhận rộng rãi về bệnh đái tháo đường được WHO công bố vào năm 1980 (1) và, dưới dạng sửa đổi, vào năm 1985 (3). Phân loại bệnh đái tháo đường năm 1980 và 1985 và các nhóm liên quan đến rối loạn dung nạp glucose bao gồm các nhóm lâm sàng và hai nhóm nguy cơ thống kê. Ủy ban Chuyên gia năm 1980 đã đề xuất hai nhóm chính của bệnh đái tháo đường và đặt tên cho chúng là IDDM hay Loại 1, và NIDDM hay Loại 2. Trong Báo cáo Nhóm Nghiên cứu năm 1985, các thuật ngữ Loại 1 và Loại 2 đã bị loại bỏ, nhưng các nhóm IDDM và NIDDM được giữ lại, và một nhóm Bệnh đái tháo đường liên quan đến suy dinh dưỡng (MRDM) đã được giới thiệu. Trong cả hai báo cáo năm 1980 và 1985, các nhóm bệnh đái tháo đường khác bao gồm các loại khác và rối loạn dung nạp glucose (IGT) cũng như đái tháo đường thai kỳ (GDM). Những điều này đã được phản ánh trong Danh pháp Bệnh tật Quốc tế (IND) sau đó vào năm 1991, và lần sửa đổi thứ mười của Phân loại Bệnh tật Quốc tế (ICD-10) vào năm 1992. Phân loại năm 1985 đã được chấp nhận rộng rãi và được sử dụng trên toàn thế giới. Nó thể hiện sự dung hòa giữa phân loại lâm sàng và phân loại nguyên nhân, cho phép phân loại từng đối tượng và bệnh nhân một cách hữu ích về mặt lâm sàng ngay cả khi nguyên nhân hoặc nguyên nhân cụ thể chưa được biết. Phân loại được khuyến nghị bao gồm cả việc phân giai đoạn bệnh đái tháo đường dựa trên các tiêu chí mô tả lâm sàng và phân loại nguyên nhân bổ sung.

* 1. **Phân loại sửa đổi**

Phân loại này bao gồm cả các giai đoạn lâm sàng và các loại nguyên nhân của bệnh đái tháo đường và các loại tăng đường huyết khác, theo đề xuất của Kuzuya và Matsuda (15).

Phân loại lâm sàng phản ánh rằng bệnh đái tháo đường, bất kể nguyên nhân, đều tiến triển qua nhiều giai đoạn lâm sàng trong diễn biến tự nhiên của bệnh. Hơn nữa, mỗi cá thể có thể chuyển từ giai đoạn này sang giai đoạn khác theo bất kỳ hướng nào. Những người mắc hoặc đang mắc bệnh đái tháo đường có thể được phân loại theo giai đoạn dựa trên các đặc điểm lâm sàng, ngay cả khi không có thông tin liên quan đến nguyên nhân cơ bản.

Phân loại theo loại nguyên nhân là kết quả của việc hiểu biết tốt hơn về nguyên nhân gây bệnh đái tháo đường.

* + - 1. **Các ứng dụng của phân loại mới**

Phân loại mới bao gồm các giai đoạn phản ánh các mức độ tăng đường huyết khác nhau ở từng cá thể mắc bất kỳ quá trình bệnh lý nào có thể dẫn đến bệnh đái tháo đường.

Tất cả các đối tượng mắc bệnh đái tháo đường đều có thể được phân loại theo giai đoạn lâm sàng, và điều này có thể đạt được trong mọi trường hợp. Giai đoạn đường huyết có thể thay đổi theo thời gian tùy thuộc vào mức độ của các quá trình bệnh lý tiềm ẩn (Hình 2). Quá trình bệnh lý có thể hiện diện nhưng có thể chưa tiến triển đủ xa để gây tăng đường huyết.

Phân loại nguyên nhân phản ánh thực tế là khiếm khuyết hoặc Quá trình có thể dẫn đến bệnh tiểu đường có thể được xác định ở bất kỳ giai đoạn nào trong quá trình phát triển bệnh tiểu đường – ngay cả ở giai đoạn đường huyết bình thường. Do đó, sự hiện diện của kháng thể tế bào đảo tụy ở một cá nhân có đường huyết bình thường khiến người đó có khả năng mắc bệnh tự miễn dịch loại 1. Đáng tiếc là hiện nay có rất ít chỉ số nhạy cảm hoặc đặc hiệu cao của bệnh tiểu đường loại 2, mặc dù những chỉ số này có thể được phát hiện khi nguyên nhân được xác định rõ ràng hơn. Các quá trình bệnh lý tương tự có thể gây ra rối loạn đường huyết lúc đói và/hoặc rối loạn dung nạp glucose mà không đáp ứng các tiêu chuẩn chẩn đoán bệnh tiểu đường. Ở một số cá nhân mắc bệnh tiểu đường, việc kiểm soát đường huyết đầy đủ có thể đạt được bằng việc giảm cân, tập thể dục và/hoặc thuốc uống. Do đó, những cá nhân này không cần insulin và thậm chí có thể trở lại IGT hoặc đường huyết bình thường. Những cá nhân khác cần insulin để kiểm soát đường huyết đầy đủ nhưng có thể sống sót mà không cần insulin. Theo định nghĩa, những cá nhân này có một lượng insulin tiết ra còn lại. Những người bị phá hủy tế bào beta rộng rãi, và do đó không còn tiết insulin, cần insulin để sống sót. Mức độ nghiêm trọng của bất thường chuyển hóa có thể thoái triển (ví dụ khi giảm cân), tiến triển (ví dụ khi tăng cân), hoặc giữ nguyên.

1. **Các biến chứng**
   1. **Đái tháo đường**

Đái tháo đường, bất kể nguyên nhân tiềm ẩn, được chia thành: Cần insulin để sống sót (tương ứng với phân loại lâm sàng trước đây là "Đái tháo đường phụ thuộc insulin - IDDM"), ví dụ: thiếu hụt peptide C; Cần insulin để kiểm soát, tức là kiểm soát chuyển hóa, thay vì để sống sót, ví dụ: tiết một lượng insulin nội sinh nhưng không đủ để đạt được đường huyết bình thường nếu không bổ sung insulin ngoại sinh; và Không cần insulin, tức là những người có thể được kiểm soát thỏa đáng bằng các phương pháp không dùng thuốc hoặc thuốc khác ngoài insulin. Hai phân nhóm sau cùng tạo thành phân nhóm trước đây của NIDDM.

* 1. **Rối loạn điều hòa glucose – Rối loạn dung nạp glucose (IGT) và rối loạn đường huyết lúc đói (IFG)**

Rối loạn điều hòa glucose (IGT và IFG) đề cập đến một trạng thái chuyển hóa trung gian giữa cân bằng glucose bình thường và bệnh tiểu đường. Tuy nhiên, cần phải khẳng định rõ ràng rằng IFG và IGT không thể thay thế cho nhau và đại diện cho những bất thường khác nhau về điều hòa glucose, một ở trạng thái đói và một sau bữa ăn.

IGT, thay vì là một nhóm như trong phân loại trước, được phân loại là một giai đoạn trong lịch sử tự nhiên của rối loạn chuyển hóa carbohydrate. Một giai đoạn của IFG cũng được công nhận vì những đối tượng này, giống như những người mắc IGT, có nguy cơ tiến triển thành bệnh tiểu đường và bệnh mạch máu lớn cao hơn, mặc dù dữ liệu triển vọng còn ít và dữ liệu ban đầu cho thấy nguy cơ tiến triển thấp hơn IGT (18), mặc dù hồ sơ yếu tố nguy cơ tim mạch tương tự đã được cho thấy ở những đối tượng mắc IFG và IGT (19). IFG đề cập đến nồng độ glucose lúc đói thấp hơn nồng độ cần thiết để chẩn đoán bệnh tiểu đường nhưng cao hơn phạm vi tham chiếu "bình thường". Giá trị của IFG là nồng độ glucose huyết tương lúc đói là 6,1 mmol l–1 (110 mg dl–1) hoặc cao hơn (máu toàn phần 5,6 mmol l–1; 100 mg dl–1), nhưng thấp hơn 7,0 mmol l–1 (126 mg dl–1) (máu toàn phần 6,1 mmol l–1; 110 mg dl–1). Nếu thực hiện OGTT, một số người mắc IFG sẽ bị IGT hoặc tiểu đường, nhưng điều này không thể xác định được nếu không có OGTT. Nếu nguồn lực cho phép, khuyến cáo tất cả những người mắc IFG nên thực hiện OGTT để loại trừ chẩn đoán tiểu đường. Những người đáp ứng các tiêu chí của IGT hoặc IFG có thể bình thường đường huyết trong cuộc sống hàng ngày, thể hiện qua nồng độ hemoglobin glycated bình thường hoặc gần bình thường. IGT và IFG không phải là thực thể lâm sàng riêng biệt, mà là các nhóm nguy cơ cho bệnh tiểu đường và/hoặc bệnh tim mạch trong tương lai (20,21).

Chúng có thể xảy ra như một giai đoạn trung gian trong bất kỳ quá trình bệnh lý nào được liệt kê trong Bảng 2. IGT thường liên quan đến Hội chứng Chuyển hóa (Hội chứng Kháng Insulin) (22). Do đó, IGT có thể không liên quan trực tiếp đến cơ chế bệnh sinh của bệnh tim mạch, mà có thể đóng vai trò là một chỉ số hoặc dấu hiệu của nguy cơ gia tăng do mối tương quan của nó với các yếu tố khác của Hội chứng Chuyển hóa, vốn là các yếu tố nguy cơ tim mạch. Rõ ràng, những người mắc IGT chỉ biểu hiện tình trạng không dung nạp glucose khi được thử nghiệm với glucose đường uống.

**CHƯƠNG II: SỬ DỤNG THUẬT TOÁN HỌC ADAP ĐỂ DỰ ĐOÁN SỰ KHỞI PHÁT CỦA BỆNH TIỂU ĐƯỜNG**

1. **Giới thiệu về thuật toán**

Khả năng dự báo và phân biệt đóng vai trò trung tâm trong nhiều tình huống y tế. Các kỹ thuật thống kê tiêu chuẩn như phân tích phân biệt, phân tích hồi quy và phân tích nhân tố đã được sử dụng để cung cấp khả năng này. Tuy nhiên, ngay cả khi tồn tại các mối quan hệ hàm ẩn có thể cung cấp khả năng dự báo, các kỹ thuật thống kê tiêu chuẩn vẫn có thể không thành công. Các phương pháp thống kê tiêu chuẩn có thể mang lại kết quả không như mong đợi khi:

1. Kích thước mẫu nhỏ.
2. Dạng của mối quan hệ hàm cơ bản chưa được biết rõ.
3. Các mối quan hệ hàm cơ bản bao gồm các tương tác và tương quan phức tạp giữa một số biến.

Những tình trạng này không phải là hiếm gặp trong các vấn đề y tế. Trong những tình huống như vậy, một số phương pháp tiếp cận mạng nơ-ron nhân tạo mang lại nhiều hứa hẹn.

Mô hình mạng nơ-ron là một lớp thuật toán học tập được mô tả lần đầu vào cuối những năm 1950.[2] Những thuật toán ban đầu này đã được mở rộng gần đây, và sự quan tâm đến chúng đã tăng lên đáng kể.[3] Các thuật toán này có khả năng sử dụng một tập dữ liệu huấn luyện để khám phá các mẫu trong dữ liệu. Một thuật toán như vậy, được gọi là ADAP, đã được phát triển bởi hai tác giả [Smith, Dickson] vào 1961.[4,5] Chúng tôi đã chọn kiểm tra thuật toán ADAP và thử nghiệm việc sử dụng nó trong việc dự báo sự khởi phát của bệnh tiểu đường không phụ thuộc insulin trong khoảng thời gian năm năm. Dữ liệu được sử dụng trong nghiên cứu này được lấy từ quần thể người da đỏ Pima gần Phoenix, Arizona. Sau khi thuật toán được huấn luyện bằng 576 trường hợp, ADAP đã được sử dụng để dự báo liệu 192 trường hợp thử nghiệm khác có phát triển bệnh tiểu đường trong vòng năm năm hay không. Việc buộc ADAP đưa ra kết luận về tất cả các trường hợp thử nghiệm đã tạo ra độ nhạy và độ đặc hiệu là 76%. Đường cong đặc tính hoạt động của máy thu (ROC) đã được xác định. Những nỗ lực đang được tiến hành để kiểm tra cùng một tập dữ liệu bằng cách sử dụng mô hình hồi quy logistic và perceptron tuyến tính.

1. **Áp dụng vào mô hình để phân tích**
   1. **Đối tượng và phương pháp**

Dân số trong nghiên cứu này là dân số người da đỏ Pima gần Phoenix, Arizona. Dân số này đã được Viện Quốc gia về Bệnh tiểu đường, Tiêu hóa và Thận nghiên cứu liên tục từ năm 1965 do tỷ lệ mắc bệnh tiểu đường cao.[6,7,8] Mỗi cư dân cộng đồng trên 5 tuổi được yêu cầu trải qua một cuộc kiểm tra tiêu chuẩn hai năm một lần, bao gồm xét nghiệm dung nạp glucose đường uống. Bệnh tiểu đường được chẩn đoán theo Tiêu chuẩn của Tổ chức Y tế Thế giới[10]; nghĩa là, nếu nồng độ glucose huyết tương sau 2 giờ sau khi thử nghiệm ít nhất là 200 mg/dl (11,1 mmol/l) tại bất kỳ cuộc kiểm tra khảo sát nào hoặc nếu Bệnh viện Dịch vụ Y tế Người da đỏ phục vụ cộng đồng phát hiện nồng độ glucose ít nhất là 200 mg/dl trong quá trình chăm sóc y tế thường quy. Ngoài việc là một cơ sở dữ liệu quen thuộc với các nhà nghiên cứu, bộ dữ liệu này còn cung cấp một nguồn dữ liệu được xác thực tốt để khám phá dự đoán ngày khởi phát bệnh tiểu đường theo chiều dọc.

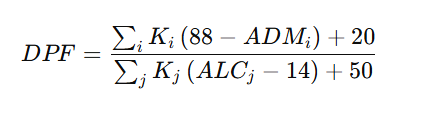
* 1. **Lựa chọn các thuộc tính**

Khi áp dụng ADAP vào vấn đề tiểu đường, tám biến số đã được chọn làm cơ sở dự báo sự khởi phát của bệnh tiểu đường trong vòng năm năm ở phụ nữ Ấn Độ Pima. Các biến số này được chọn vì chúng được phát hiện là những yếu tố nguy cơ đáng kể đối với bệnh tiểu đường ở người Pima hoặc các nhóm dân số khác.

Các biến đầu vào bao gồm:

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 (mmHg)
4. Độ dày nếp gấp da cơ tam đầu (mm)
5. Nồng độ insulin huyết thanh sau 2 giờ (Uh/ml)
6. Chỉ số khối cơ thể (Cân nặng (kg)/(Chiều cao (in))
7. Chức năng phả hệ bệnh tiểu đường
8. Tuổi (năm)

Hàm phả hệ bệnh tiểu đường:



được tính toán cho mỗi kỳ thi chỉ bằng cách sử dụng dữ liệu có sẵn vào ngày thi đó và trong đó:

i bao gồm tất cả những người thân đã mắc bệnh tiểu đường tính đến ngày khám của đối tượng;

j bao gồm tất cả những người thân KHÔNG mắc

bệnh tiểu đường tính đến ngày khám của đối tượng;

K là tỷ lệ phần trăm gen được chia sẻ bởi người thân (*relativex*) và

* bằng 0,500 khi *relativex* là cha mẹ hoặc anh chị em ruột,
* bằng 0,250 khi người thân là anh chị em cùng cha khác mẹ, ông bà, dì hoặc chú, và
* bằng 0,125 khi người thân là dì, chú, bác hoặc anh chị em họ;

ADMi  là độ tuổi tính theo năm của người thân khi bệnh tiểu đường được chẩn đoán;

ACLj  là độ tuổi tính theo năm của người thân tại lần khám không mắc bệnh tiểu đường gần nhất (trước ngày khám của đối tượng);

Hằng số

Các hằng số 88 và 14, ngoại trừ một số ít trường hợp, biểu thị độ tuổi tối đa và tối thiểu mà người thân của các đối tượng trong nghiên cứu này mắc bệnh tiểu đường.

Các hằng số 20 và 50 được chọn sao cho:

1. Một đối tượng không có người thân sẽ có giá trị DPF thấp hơn một chút so với mức trung bình
2. Giá trị DPF sẽ giảm tương đối chậm khi những người thân trẻ không mắc bệnh tiểu đường tham gia cơ sở dữ liệu
3. Giá trị DPF sẽ tăng tương đối nhanh khi những người thân đã biết mắc bệnh tiểu đường.

Lưu ý rằng giá trị của DPF tăng khi số lượng người thân mắc bệnh tiểu đường tăng, độ tuổi mắc bệnh tiểu đường của những người thân này giảm, và tỷ lệ phần trăm gen mà họ chia sẻ với đối tượng tăng. Cũng lưu ý rằng giá trị của DPF giảm khi số lượng người thân chưa bao giờ mắc bệnh tiểu đường tăng, độ tuổi của họ khi khám lần cuối tăng, và tỷ lệ phần trăm gen mà họ chia sẻ với đối tượng tăng.

* 1. **Lựa chọn các trường hợp**

Bệnh tiểu đường được định nghĩa là nồng độ glucose huyết tương lớn hơn 200 mg/dl hai giờ sau khi uống 75 gam dung dịch carbohydrate. Các trường hợp được chọn từ nhóm kiểm tra đáp ứng các tiêu chí sau:

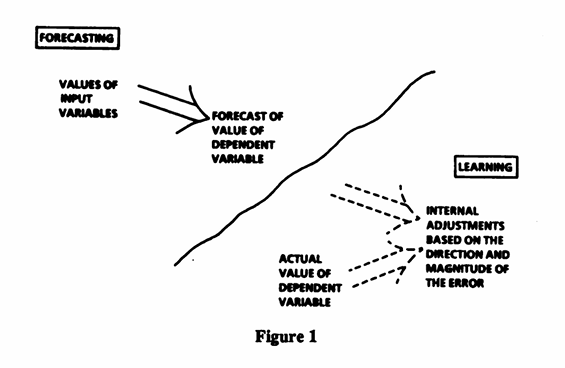
1. Đối tượng là nữ.
2. Đối tượng nghiên cứu được 221 tuổi tại thời điểm kiểm tra chỉ số. Kiểm tra chỉ số đề cập đến nghiên cứu được chọn để sử dụng trong mô hình này. Nó không nhất thiết phải tương ứng với lần kiểm tra đầu tiên theo thứ tự thời gian cho đối tượng này.
3. Mỗi đối tượng chỉ được chọn một xét nghiệm. Xét nghiệm đó phải cho thấy GTT không bị tiểu đường và đáp ứng một trong hai tiêu chí sau:
   1. Bệnh tiểu đường được chẩn đoán trong vòng năm năm sau khi kiểm tra, HOẶC
   2. Xét nghiệm GTT được thực hiện sau đó 5 năm hoặc hơn không phát hiện được bệnh tiểu đường.
4. Nếu bệnh tiểu đường xảy ra trong vòng một năm kể từ lần khám, lần khám đó sẽ bị loại khỏi nghiên cứu để loại bỏ khỏi mô hình dự báo những trường hợp có khả năng dễ dự báo hơn. Trong 75% số lần khám bị loại, bệnh tiểu đường được chẩn đoán trong vòng sáu tháng.

Sử dụng các tiêu chí này, 768 bài kiểm tra đã được chọn. Trong số đó, 576 bài được chọn ngẫu nhiên để sử dụng trong tập huấn luyện hoặc tập học tập và 192 trường hợp còn lại trở thành tập dự báo. Giả thuyết của chúng tôi là ADAP có thể học cách dự báo liệu một cá nhân nhất định có mắc bệnh tiểu đường trong vòng năm năm hay không, dựa trên giá trị của tám biến đầu vào.

1. **Mô tả thuật toán**

ADAP là một chương trình học tập thích ứng tạo ra và thực thi các tín hiệu tương tự kỹ thuật số của các thiết bị giống percepton. Nó có thể tạo ra nhiều loại tín hiệu tương tự như vậy. Tuy nhiên, nó được thiết kế đặc biệt để tạo ra các tín hiệu tương tự giống perceptron "phân vùng" chuyên dụng, phù hợp với từng vấn đề cụ thể.

Các mô hình tương tự ADAP học bằng cách thực hiện các điều chỉnh nội tại khi dự đoán của chúng trở nên không chính xác. Giả sử một hiện tượng nào đó được biểu thị dưới dạng giá trị hoặc trạng thái và bạn giả định rằng nó có thể được dự báo bằng một hàm số chưa biết của các biến khác. Chu kỳ điều chỉnh của ADAP được xác định bằng cách đọc giá trị của các biến trong một trường hợp (kiểm tra chỉ số trong nghiên cứu này). Sau đó, ADAP tạo ra một dự báo về giá trị hoặc trạng thái của hiện tượng, có thể là một hàm số của các biến đầu vào. Sau đó, nếu nó ở chế độ điều chỉnh, nó sẽ đọc trạng thái hoặc giá trị thực tế mà nó vừa cố gắng dự báo. Sau đó, nó so sánh dự báo với giá trị thực và thực hiện các điều chỉnh nội tại dựa trên hướng và độ lớn của lỗi. [HÌNH 1] Về mặt này, ADAP giống như các thiết bị học tập mạng nơ-ron khác.



Một điểm khác biệt lớn giữa ADAP và các mạng nơ-ron truyền thống hơn là cách các đơn vị cảm biến có thể được tổ chức và được "kết nối". Hình 2 là một ví dụ về một mạng tương tự ADAP. Các trường hợp đầu vào kích thích các cảm biến được tổ chức thành các phân vùng. Đến lượt mình, các cảm biến đó được kết nối với các đơn vị phân vùng được kết nối thông qua các trọng số có thể điều chỉnh với một đơn vị phản hồi. Một đơn vị cảm biến biểu diễn một giá trị rời rạc hoặc một dải giá trị có thể được giả định bởi một biến đầu vào nhất định. Tập hợp tất cả các cảm biến cho một đầu vào nhất định khả dụng bao gồm phân vùng cho biến đó. Mặc dù mỗi đơn vị liên kết được kết nối với mỗi cảm biến, nhưng ADAP thường hạn chế mô hình này bằng cách yêu cầu chỉ có một và duy nhất một kết nối khác không trong mỗi phân vùng. Tất cả các kết nối khác không thường được đặt thành trọng số 1, và các trọng số đó không được sử dụng trong quá trình học. Kết nối khác không này trong một phân vùng được chọn ngẫu nhiên. ADAP thường sử dụng một số lượng lớn các đơn vị liên kết như vậy. (Trong nghiên cứu này, chúng tôi đã sử dụng 100.000) Mỗi ​​đơn vị liên kết có một và chỉ một kết nối đầu ra với người phản hồi. Quy tắc kích hoạt cho một đơn vị liên kết với một tập hợp các giá trị biến đầu vào cụ thể là một hàm ngưỡng.

Tất cả các giá trị của một trường hợp khớp với một kết nối khác không được cho là kích thích các kết nối đó. Nếu tổng số kết nối được kích thích cho một đơn vị liên kết nhất định lớn hơn một giá trị định trước - ngưỡng - thì đơn vị liên kết đó được cho là đã được kích hoạt. Giá trị ngưỡng được cố định sao cho từ 2% đến 10% tổng số đơn vị liên kết sẽ được kích hoạt cho bất kỳ trường hợp đầu vào nào. Cuối cùng, các giá trị phản hồi được kết nối với mỗi đơn vị liên kết đã kích hoạt được cộng lại và tổng này tạo thành dự báo cho trường hợp đầu vào đã cho. Nếu ADAP đang ở chế độ học, giá trị đã biết được so sánh với dự báo vừa được thực hiện và chênh lệch được tính toán và phân bổ giữa các đơn vị đã kích hoạt cho trường hợp huấn luyện đó. Toàn bộ quá trình dự báo/học có thể được lặp lại cho các trường hợp bổ sung. Với ADAP, phản hồi đầu ra có thể được diễn giải là dự báo về một giá trị liên tục (ví dụ: 65,4) hoặc trạng thái (hiện diện hoặc không hiện diện).

Do thuật toán ADAP ra đời trước khi sự quan tâm đến các mô hình kết nối trỗi dậy, một số tính năng của nó khác với các mô hình đương đại. Tuy nhiên, sử dụng thuật ngữ của Rumelhart, Hinton và McClelland, nó giống nhất với một mô hình học tập liên tưởng tương tác sử dụng quy tắc học tập Hebbian.

**CHƯƠNG III: TỔNG QUAN, PHÂN LOẠI VÀ CHUẨN ĐOÁN, THẢO LUẬN CÁC LỚP GLUCOSE VÀ ĐIỀU TRỊ**

**CHƯƠNG IV: KHÁM PHÁ VÀ PHÂN TÍCH KHÁM PHÁ VỀ BỆNH ĐÁI THÁO ĐƯỜNG**

1. **Định nghĩa vấn đề**
2. **Chuẩn bị vấn đề**
3. **Chuẩn bị dữ liệu**
4. **Phân tích dữ liệu**