**Sử dụng thuật toán học ADAP để dự đoán sự khởi phát của bệnh tiểu đường**

Tổng quan nghiên cứu:

Mạng nơ-ron hay mô hình kết nối cho xử lý song song không phải là mới. Tuy nhiên, mối quan tâm đã trỗi dậy trở lại trong nửa thập kỷ qua.

Nghiên cứu này sử dụng thuật toán mạng nơ-ron ADAP để dự báo khả năng mắc bệnh tiểu đường trong vòng 5 năm ở nhóm dân số người da đỏ Pima - một cộng đồng có tỷ lệ tiểu đường rất cao tại Arizona, Mỹ.

**Kết quả chính:** Thuật toán đạt độ chính xác 76% trong cả độ nhạy (khả năng phát hiện người sẽ mắc bệnh) và độ đặc hiệu (khả năng xác định người không mắc bệnh).

1. **Tại sao chọn mạng nơ-ron?**

Các phương pháp thống kê truyền thống thường gặp khó khăn khi:

* Dữ liệu có kích thước nhỏ
* Mối quan hệ giữa các yếu tố phức tạp và chưa rõ ràng
* Có nhiều tương tác phức tạp giữa các biến số

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.

Mạng nơ-ron ADAP được chọn vì khả năng tự học và khám phá các mẫu ẩn trong dữ liệu mà các phương pháp thống kê thông thường có thể bỏ sót.

1. **Đối tượng và dữ liệu nghiên cứu**

**Đối tượng:** Phụ nữ người da đỏ Pima từ 21 tuổi trở lên

* **Dữ liệu huấn luyện:** 576 trường hợp
* **Dữ liệu kiểm tra:** 192 trường hợp
* **Thời gian theo dõi:** 5 năm

**Tiêu chí chẩn đoán:** Nồng độ glucose máu > 200 mg/dl sau 2 giờ uống dung dịch đường

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:

**8 yếu tố dự báo được sử dụng**

1. **Số lần mang thai** - Liên quan đến nguy cơ tiểu đường thai kỳ
2. **Nồng độ glucose sau 2 giờ** - Chỉ số quan trọng nhất
3. **Huyết áp tâm trương** - Yếu tố tim mạch
4. **Độ dày nếp gấp da** - Phản ánh lượng mỡ dưới da
5. **Nồng độ insulin** - Chỉ số kháng insulin
6. **Chỉ số BMI** - Tình trạng thừa cân, béo phì
7. **Chức năng phả hệ tiểu đường (DPF)** - Yếu tố di truyền (công thức riêng)
8. **Tuổi** - Nguy cơ tăng theo tuổi

**Điểm đặc biệt: Chỉ số Phả hệ Tiểu đường (DPF)**

**A black text on a white background

AI-generated content may be incorrect.**

Đây là một công thức độc đáo do nhóm nghiên cứu phát triển, tính toán dựa trên:

* Số lượng người thân mắc tiểu đường
  + 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
* Tuổi mắc bệnh của 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);
* K là tỷ lệ phần trăm gen được chia sẻ bởi người thân (*relativex*)
* Mức độ họ hàng (50% gen chung với bố mẹ, anh chị em; 25% với ông bà, cô chú; 12.5% với anh chị em họ)
* Số người thân không mắc bệnh và tuổi của họ

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.
4. **Cách hoạt động của thuật toán ADAP**

Khái niệm cơ bản: ADAP tạo ra một mạng lưới gồm 100.000 "đơn vị liên kết" - giống như các nơ-ron nhân tạo.

**Quá trình học:**

1. Chia nhóm dữ liệu: Mỗi yếu tố được chia thành các khoảng giá trị cụ thể
2. Kết nối ngẫu nhiên: Mỗi đơn vị liên kết kết nối với một yếu tố từ mỗi nhóm
3. Học từ sai lầm: Khi dự đoán sai, thuật toán tự điều chỉnh các kết nối
4. Ngưỡng kích hoạt: Chỉ 8% đơn vị liên kết được kích hoạt cho mỗi trường hợp

**Quá trình dự đoán:**

* Nhập thông tin 8 yếu tố của một người
* Kích hoạt khoảng 8.000 đơn vị liên kết phù hợp
* Tính tổng các tín hiệu từ các đơn vị đã kích hoạt
* Ra kết quả: xác suất mắc tiểu đường trong 5 năm

**Kết quả và ý nghĩa**

**Hiệu suất:**

* Độ nhạy: 76% (phát hiện đúng 76% người sẽ mắc bệnh)
* Độ đặc hiệu: 76% (xác định đúng 76% người không mắc bệnh)
* Điểm cân bằng tối ưu: 0.448

**So sánh với phương pháp khác:** Nghiên cứu đang tiến hành so sánh ADAP với hồi quy logistic và perceptron tuyến tính trên cùng bộ dữ liệu.

**Ứng dụng thực tế**

**Ưu điểm:**

* Không cần biết trước mối quan hệ toán học giữa các yếu tố
* Tự động khám phá các mẫu phức tạp
* Phù hợp với dữ liệu y tế có nhiều yếu tố tương tác

**Hạn chế:**

* Cần dữ liệu huấn luyện lớn và chất lượng
* Khó giải thích cách thuật toán đưa ra quyết định
* Chỉ áp dụng được cho nhóm dân số tương tự

**Kết luận**

Nghiên cứu chứng minh mạng nơ-ron ADAP có thể dự báo bệnh tiểu đường với độ chính xác khá cao (76%). Đây là bước tiến quan trọng trong việc ứng dụng trí tuệ nhân tạo vào y học dự phòng, giúp can thiệp sớm và giảm gánh nặng bệnh tật cho cộng đồng.

*Lưu ý: Đây là nghiên cứu từ những năm 1980s, mở đường cho các ứng dụng AI hiện đại trong y tế ngày nay.*