**Tóm Tắt Bài Báo**

**Heart Disease Prediction System Using Model Of**

**Machine Learning and Sequential Backward**

**Selection Algorithm for Features Selection**

**Nhóm người của khoa liên quan khoa học máy tính đến từ Đại học Khoa học và Công nghệ Điện tử Trung Quốc**

Dưới đây là tóm tắt nội dung chính của bài báo:

1. Mục tiêu và ý nghĩa nghiên cứu:

Bài báo đề xuất một hệ thống dự đoán bệnh tim bằng cách sử dụng các mô hình học máy kết hợp với thuật toán chọn đặc trưng Sequential Backward Selection (SBS). Mục tiêu của nghiên cứu là phát triển một phương pháp dự đoán không xâm lấn, có khả năng phát hiện bệnh tim ở giai đoạn sớm, từ đó giúp cải thiện hiệu quả điều trị và phục hồi cho bệnh nhân.

2. Dữ liệu và xử lý sơ bộ:

- Dữ liệu: Sử dụng Bộ dữ liệu bệnh tim từ Cleveland năm 2016, ban đầu có 303 ca với 76 đặc trưng. Sau bước xử lý dữ liệu, 6 bản ghi có giá trị thiếu được loại bỏ, giúp thu được một tập dữ liệu gồm 297 mẫu với 13 đặc trưng được giữ lại.

- Phương pháp xử lý: Áp dụng các kỹ thuật tiền xử lý như xóa bỏ dữ liệu thiếu, chuẩn hóa dữ liệu bằng Standard Scaler (đưa về giá trị trung bình 0 và phương sai 1) và Min-Max Scaler (giá trị được thu gọn trong khoảng từ 0 đến 1) nhằm làm cho các đặc trưng có cùng tầm quan trọng và thuận tiện cho quá trình học của mô hình.

3. Phương pháp nghiên cứu:

- Chọn đặc trưng với SBS:

Thuật toán Sequential Backward Selection (SBS) được sử dụng để giảm số lượng đặc trưng không cần thiết nhằm giảm độ phức tạp của mô hình và tránh hiện tượng overfitting. SBS tiến hành loại bỏ lần lượt những đặc trưng ít quan trọng – tức là đặc trưng mà khi loại bỏ đi sẽ không làm giảm đáng kể hiệu năng phân loại của mô hình, và thậm chí có thể cải thiện hiệu quả dự đoán.

- Xây dựng mô hình phân loại với KNN:

Sau khi có các tập đặc trưng con được tạo ra từ SBS, bài báo sử dụng thuật toán K-Nearest Neighbor (K-NN) làm bộ phân loại chính. Nghiên cứu thử nghiệm với nhiều giá trị khác nhau của k (từ 1 đến 8) và tính trung bình hiệu năng của mô hình trên từng tập đặc trưng đã chọn.

- Đánh giá hiệu năng:

Hệ thống được đánh giá qua các chỉ số dựa trên ma trận nhầm lẫn (confusion matrix) với các thành phần như TP, TN, FP, FN. Phương pháp chia dữ liệu theo tỉ lệ 70% cho huấn luyện và 30% cho kiểm tra được áp dụng để đánh giá độ chính xác của mô hình.

4. Kết quả thực nghiệm:

Thông qua các thí nghiệm, nhóm nghiên cứu đã đánh giá hiệu năng của K-NN trên 13 tập con của đặc trưng được tạo ra từ SBS. Kết quả cho thấy, khi sử dụng tập con gồm 6 đặc trưng, độ chính xác của mô hình đạt tới 90% – cao hơn so với các tập đặc trưng khác. Biểu đồ minh họa trong bài báo cũng cho thấy xu hướng tăng độ chính xác khi giảm bớt số lượng đặc trưng không cần thiết.

5. Kết luận:

Bài báo kết luận rằng việc kết hợp thuật toán SBS trong việc chọn lọc đặc trưng, cùng với mô hình K-NN, đã tạo ra một hệ thống dự đoán bệnh tim hiệu quả với độ chính xác cao (90%). Phương pháp này không chỉ giảm thời gian tính toán mà còn giúp cải thiện khả năng phân loại giữa bệnh nhân bị tim mạch và những người khỏe mạnh. Do đó, hệ thống được đề xuất có tiềm năng ứng dụng trong thực tế như một công cụ hỗ trợ chẩn đoán sớm bệnh tim.

Ngoài ra, bài báo còn nêu rõ các đóng góp của nghiên cứu, nền tảng tài chính và học thuật hỗ trợ quá trình thực hiện dự án, cũng như so sánh với các phương pháp và kết quả đã có trong nghiên cứu trước đó. Đây là một hướng tiếp cận hữu ích nhằm cải thiện quá trình chẩn đoán bệnh tim thông qua ứng dụng các kỹ thuật học máy tiên tiến.