Bộ Giáo dục và Đào tạo

**Trường Đại học Sài Gòn**

**0B**

Khoa Công Nghệ Thông Tin

Bộ môn PPNCKH trong CNTT

# ĐỀ CƯƠNG CHI TIẾT

Ngày thực hiện đề cương: 27/03/2025

|  |  |  |
| --- | --- | --- |
| **TÊN ĐỀ TÀI** | Heart Disease Risk Prediction | |
| **LĨNH VỰC CHUYÊN NGÀNH** | Khoa học máy tính | |
| **LOẠI HÌNH NGHIÊN CỨU** | Sản phẩm | |
| **NGƯỜI HƯỚNG DẪN** | TS. Đỗ Như Tài | |
| **SINH VIÊN THỰC HIỆN 1** | Hồ Hoàng Nhật Tâm | 3122410368 |
| **SINH VIÊN THỰC HIỆN 2** | Phạm Hữu Nghĩa | 3122410261 |
| **SINH VIÊN THỰC HIỆN 3** | Bùi Lê Duy Thái | 3122410377 |
| **SINH VIÊN THỰC HIỆN 4** | Phạm Công Thành | 3122410386 |

## GIỚI THIỆU

|  |  |  |
| --- | --- | --- |
| **CÂU HỎI GỢI Ý** | **Linh vực và nghiên cứu liên quan đã và đang phát triển như thế nào? Các vấn đề, và bài toán đặt ra cần giải quyết là gi?̀** |  |
| **HƯỚNG DẪN** | Giới thiệu tổng quan về đề tài – những vấn đề và linh vực liên quan đến đề tài. |
| Bệnh tim mạch là một trong các nguyên nhân gây tử vong hàng đầu trên thế giới, cũng như tại Việt Nam. Theo số liệu của Tổ chức Y tế Thế giới năm 2019, tử vong do bệnh tim mạch chiếm tới 39,5%, trong đó: bệnh mạch máu não (55,4%), bệnh tim thiếu máu cục bộ (32%), bệnh tim do tăng huyết áp (6,9%) và bệnh tim mạch khác (5,7%).  Tại Việt Nam, xu hướng tử vong do bệnh tim mạch đang ngày càng tăng, trong đó tử vong do bệnh mạch máu não chiếm tỷ lệ lớn nhất, với tỷ suất tử vong tăng từ 127,3/100.000 dân (năm 2000) lên 164,9/100.000 dân hiện nay  Gánh nặng bệnh tật vẫn tiếp tục gia tăng, đặc biệt là gánh nặng về bệnh lý tim mạch và xu hướng trẻ hóa ở những người trong độ tuổi lao động.  Bệnh tim không chỉ ảnh hưởng đến sức khỏe cá nhân mà còn tạo ra áp lực lớn lên hệ thống y tế do chi phí điều trị cao và tỷ lệ tử vong cao. Nếu có thể phát hiện sớm nguy cơ mắc bệnh tim, bệnh nhân có thể điều chỉnh lối sống hoặc can thiệp y tế kịp thời để giảm nguy cơ biến chứng nghiêm trọng. Các phương pháp chẩn đoán bệnh tim truyền thống thường dựa trên xét nghiệm lâm sàng, đo huyết áp, xét nghiệm máu và chẩn đoán hình ảnh (ECG, siêu âm tim). Tuy nhiên, những phương pháp này có thể tốn kém, mất thời gian và đôi khi không phát hiện được bệnh ở giai đoạn đầu.  Với sự phát triển mạnh mẽ của khoa học dữ liệu và trí tuệ nhân tạo (AI), việc sử dụng các mô hình học máy (Machine Learning) để dự đoán nguy cơ mắc bệnh tim ngày càng được quan tâm. Những mô hình này có thể phân tích dữ liệu bệnh nhân, nhận diện mối quan hệ giữa các yếu tố nguy cơ và bệnh tim, từ đó hỗ trợ bác sĩ đưa ra quyết định chính xác hơn trong chẩn đoán và điều trị. Ứng dụng các thuật toán Machine Learning như Logistic Regression, Decision Trees, Random Forest, Support Vector Machine (SVM), và Neural Networks để dự đoán nguy cơ bệnh tim. Xử lý dữ liệu y tế và tối ưu hóa mô hình nhằm nâng cao độ chính xác trong chẩn đoán. Kết hợp AI với dữ liệu sinh trắc học, xét nghiệm máu và hình ảnh y khoa để tạo ra hệ thống hỗ trợ chẩn đoán thông minh. | |
| **GHI CHÚ** | |
|  | |

## TÌNH HÌNH NGHIÊN CỨU TRONG VÀ NGOÀI NƯỚC

|  |  |  |
| --- | --- | --- |
| **CÂU HỎI GỢI Ý** | **Lĩnh vực và nghiên cứu liên quan đã và đang phát triển như thế nào? Các vấn đề, và bài toán đặ t ra cần giải quyết là gi?̀** |  |
| **HƯỚ NG DẪN** | Tim hiểu các nghiên cứu đã công bố gần nhất (5 năm trở lại) về linh vực liên quan trong và ngoài nước. Phân tich các kết quả đóng góp, nhận xét các hạn chế còn tồn tại. |
| Nghiên cứu được công bố trên tạp chí *British Journal of Ophthalmology* cho thấy rằng việc phân tích mạch máu trong võng mạc có thể giúp xác định những người có nguy cơ cao mắc bệnh tim mạch. Các nhà khoa học phát hiện ra rằng kích thước và hình dạng của mạch máu võng mạc có mối liên hệ với nguy cơ nhồi máu cơ tim, mở ra tiềm năng cho việc sử dụng khám mắt như một công cụ dự đoán sớm.  Tại Việt Nam, bệnh tim mạch là nguyên nhân hàng đầu gây tử vong, chiếm 33% tổng số ca tử vong mỗi năm, tương đương khoảng 200.000 người. Các yếu tố nguy cơ chính bao gồm tăng huyết áp, tiểu đường, béo phì và lối sống ít vận động.  Tại Bệnh viện Quân y 103, một nghiên cứu đã được thực hiện nhằm xác định các yếu tố nguy cơ tim mạch và dự báo nguy cơ mắc bệnh mạch vành trong 10 năm tới bằng thang điểm Framingham. Nghiên cứu này giúp nhận diện các yếu tố nguy cơ phổ biến trong cộng đồng Việt Nam và đề xuất các biện pháp phòng ngừa phù hợp  Năm 2020, Bộ Y tế Việt Nam ban hành tài liệu chuyên môn về dự phòng tiên phát bệnh tim mạch, nhấn mạnh tầm quan trọng của việc phòng ngừa từ sớm thông qua kiểm soát các yếu tố nguy cơ như tăng huyết áp, tiểu đường và lối sống không lành mạnh. Tài liệu này cung cấp hướng dẫn chi tiết cho các cơ sở y tế trong việc đánh giá và quản lý nguy cơ tim mạch. ​  Các nghiên cứu đã giới thiệu nhiều phương pháp mới, từ xét nghiệm máu, khám mắt đến sử dụng thang điểm đánh giá nguy cơ, giúp đa dạng hóa công cụ dự đoán và phòng ngừa bệnh tim mạch. | |
| Những phát hiện mới giúp nâng cao nhận thức cộng đồng về các yếu tố nguy cơ tiềm ẩn và khuyến khích việc kiểm tra sức khỏe định kỳ để phát hiện sớm.  Mặc dù các phương pháp mới hứa hẹn, nhưng cần thêm nghiên cứu để xác định độ chính xác, tính khả thi và hiệu quả kinh tế khi áp dụng rộng rãi. Việc áp dụng các phương pháp dự đoán mới đòi hỏi cơ sở hạ tầng y tế phù hợp, đào tạo nhân lực và sự chấp nhận từ cộng đồng, đặc biệt ở các khu vực có nguồn lực hạn chế. | |
|  | |
| **GHI CHÚ** | |
|  | |

## MỤC TIÊU VÀ PHẠM VI NGHIÊN CỨU



Với sự phát triển của trí tuệ nhân tạo (AI) và học máy (Machine Learning), có thể xây dựng các mô hình dự đoán nguy cơ mắc bệnh tim với độ chính xác cao hơn, từ đó hỗ trợ bác sĩ đưa ra quyết định tốt hơn. Tuy nhiên, vẫn còn nhiều thách thức cần giải quyết, bao gồm:

* Cách thu thập và xử lý dữ liệu y tế để phù hợp với các mô hình học máy.;
* Độ chính xác và khả năng diễn giải của mô hình trong bối cảnh y tế;
* Ứng dụng thực tiễn của các mô hình vào hệ thống chăm sóc sức khỏe.

**́**

## CÁC GIẢ THUYẾT VÀ CÁCH TIẾP CẬ N

|  |  |  |
| --- | --- | --- |
| **CÂU HỎ I GỢ I Ý** | **Các giả thuyết đặ t ra để giả i quyết bài toán chính ? Các cách tiếp cậ n để giả i quyết bài toán đặ t ra?** |  |
| **HƯỚ NG DẪN** | Đặt ra những giả thuyết, hay vấn đề bài toán con cần phải giải quyết để đạt được mục tiêu nghiên cứu đề tài.  Các cách tiếp cận (dự kiến) để giải quyết các giả thuyết, bài toán con đã đặ̃t. |
| **Giả thuyết:**  Để xây dựng mô hình dự đoán nguy cơ mắc bệnh tim mạch, đề tài đặt ra các giả thuyết sau:   * Các yếu tố nguy cơ truyền thống như tuổi, huyết áp, mức cholesterol, chỉ số BMI, tiểu đường, hút thuốc lá có ảnh hưởng đáng kể đến nguy cơ mắc bệnh tim; * Mô hình học máy có thể phân tích và dự đoán nguy cơ mắc bệnh tim với độ chính xác cao hơn so với các phương pháp thống kê truyền thống; * Các thuật toán học máy khác nhau có thể có hiệu suất khác nhau, và việc tối ưu hóa mô hình có thể cải thiện độ chính xác dự đoán; * Việc kết hợp nhiều nguồn dữ liệu khác nhau (như xét nghiệm máu, hình ảnh y khoa, dữ liệu di truyền) có thể làm tăng độ chính xác của mô hình; * Mô hình có thể triển khai dưới dạng một công cụ hỗ trợ ra quyết định cho bác sĩ, giúp họ đánh giá nhanh nguy cơ mắc bệnh tim của bệnh nhân.   **Từ mục tiêu nghiên cứu, chúng tôi đặt ra hai bài toán chính của đề tài:**  Bài toán thứ nhất: Thu thập và tiền xử lý dữ liệu;  Tìm kiếm và thu thập tập dữ liệu y tế phù hợp (ví dụ: Framingham Heart Study, Cleveland Heart Disease Dataset, dữ liệu từ bệnh viện trong nước...).  Làm sạch dữ liệu, xử lý giá trị bị thiếu, loại bỏ dữ liệu nhiễu và chuẩn hóa dữ liệu.  Lựa chọn các biến đầu vào quan trọng nhất để tăng hiệu suất mô hình.  Bài toán thứ hai: Lựa chọn thuật toán học máy phù hợp.  So sánh các thuật toán học máy phổ biến như Logistic Regression, Random Forest, Gradient Boosting (XGBoost, LightGBM), Deep Learning (Neural Networks) Đánh giá độ chính xác của từng thuật toán bằng các chỉ số như Accuracy, Precision, Recall, F1-score, AUC-ROC..  Bài toán thứ ba: Cải thiện hiệu suất mô hình  Sử dụng kỹ thuật tối ưu hóa tham số (Hyperparameter tuning) như Grid Search, Random Search, Bayesian Optimization.  **Cách tiếp cậ n:**   1. Thu thập dữ liệu từ nguồn y tế đáng tin cậy. Nâng cấp hoặc chỉ nh sử a mộ t số hoạt độ ng cho phù hợ p với hệ họ c thich nghí 2. Lựa chọn thuật toán học máy: Chạy thử các thuật toán Logistic Regression, Random Forest, XGBoost, Deep Learning để so sánh hiệu suất. 3. Cải thiện hiệu suất mô hình: Chạy thử các thuật toán Logistic Regression, Random Forest, XGBoost, Deep Learning để so sánh hiệu suất. 4. Xây dựng hệ thống hỗ trợ ra quyết định: Phát triển ứng dụng web với Streamlit hoặc Flask để hiển thị kết quả dự đoán. Đưa ra biểu đồ và thông tin dễ hiểu cho người dùng (ví dụ: hiển thị % nguy cơ mắc bệnh tim). | |
|  | |
| **GHI CHÚ** | |
|  | |

## TÀI LIỆU THAM KHẢO

|  |  |  |
| --- | --- | --- |
| **CÂU HỎI GỢI Ý** | **Các tài liệu tham khảo thuộc lĩnh vực nghiên cứu liên quan là gi?̀** |  |
| **HƯỚ NG DẪN** | 1. Tổ chức Y tế Thế giới (WHO) 2. Hiệp hội Tim mạch Hoa Kỳ (AHA - American Heart Association) 3. Các nghiên cứu khoa học và bài báo trên PubMed, Google Scholar.   [4] Mayo Clinic, WebMD  [5] [https://www.who.int/vietnam/vi/health-topics/cardiovascular-disease?](https://www.who.int/vietnam/vi/health-topics/cardiovascular-disease?utm_source=chatgpt.com)  [6] https://thanhnien.vn/xet-nghiem-mau-co-the-du-doan-nguy-co-mac-benh-tim-mach-185847114.htm? |
|  **Tiếng Anh**   1. Han J, Kamber M, Pei J. [Khai thác dữ liệu: Các khái niệm và kỹ thuật, Waltham](https://homes.di.unimi.it/ceselli/IM/2012-13/slides/02-KnowYourData.pdf) . Nhà xuất bản Morgan Kaufmann. 2012. 2. Palaniappan S, Awang R. [Hệ thống dự đoán bệnh tim thông minh sử dụng các kỹ thuật khai thác dữ liệu](https://ieeexplore.ieee.org/abstract/document/4493524/) . Trong hội nghị quốc tế IEEE/ACS năm 2008 về hệ thống máy tính và ứng dụng. IEEE. 2008; 108-115. | |

|  |
| --- |
|  **Tiếng Việt**  [1] Machine Learning cơ bản – Vũ Hữu Tiệp  [2] Deep Learning cơ bản – Nguyễn Thanh Tuấn |
| LCMS nguôǹ m̉ơ Moodle.   **Trang web** |
|  |
| **GHI CHÚ** |
|  |

|  |  |  |
| --- | --- | --- |
| **NGƯỜI THỰC HIỆ N ĐỀ CƯƠNG** |  |  |
| **SINH VIÊN THỰ C HIỆ N 1** | Hồ Hoàng Nhật Tâm | 3122410368 |
| **SINH VIÊN THỰ C HIỆ N 2** | Phạm Hữu Nghĩa | 3122410261 |
| **SINH VIÊN THỰ C HIỆ N 3** | Bùi Lê Duy Thái | 3122410377 |
| **SINH VIÊN THỰ C HIỆ N 4** | Phạm Công Thành | 3122410386 |

|  |  |  |
| --- | --- | --- |
| **XÁC NHẬ N CỦ A GIÁO VIÊN HƯỚ NG DẪN** | |  |
| **NGƯỜI HƯỚ NG DẪN** | TS. Đỗ Như Tài |  |
| **Ý KIẾN** |  |  |
|  |  đồng ý hướng dâñ |  không đồng ý hướng dâñ |