|  |
| --- |
| **BỘ GIÁO DỤC VÀ ĐÀO TẠO**  **TRƯỜNG ĐẠI HỌC NGUYỄN TẤT THÀNH**  **KHOA CÔNG NGHỆ THÔNG TIN**    **CHUYÊN ĐỀ CHUYÊN SÂU TRÍ TUỆ NHÂN TẠO 2**  **XÂY DỰNG HỆ THỐNG PHÂN LOẠI CÁC BỆNH VỀ MẮT**  **SỬ DỤNG THUẬT TOÁN HỌC SÂU**  **Giảng viên hướng dẫn: TRẦN CHÂU THANH THIỆN**  **Sinh viên thực hiện: HỒ CÔNG CHÍ**  **MSSV: 2100008526**  **Chuyên ngành: Trí tuệ nhân tạo**  **Khóa: 2021**  **Tp.HCM, tháng 12 năm 2024** |
| **BỘ GIÁO DỤC VÀ ĐÀO TẠO**  **TRƯỜNG ĐẠI HỌC NGUYỄN TẤT THÀNH**  **KHOA CÔNG NGHỆ THÔNG TIN**    **CHUYÊN ĐỀ CHUYÊN SÂU TRÍ TUỆ NHÂN TẠO 2**  **XÂY DỰNG HỆ THỐNG PHÂN LOẠI CÁC BỆNH VỀ MẮT**  **SỬ DỤNG THUẬT TOÁN HỌC SÂU**  **Giảng viên hướng dẫn: TRẦN CHÂU THANH THIỆN**  **Sinh viên thực hiện: HỒ CÔNG CHÍ**  **MSSV: 2100008526**  **Chuyên ngành: Trí tuệ nhân tạo**  **Khóa: 2021**  **Tp.HCM, tháng 12 năm 2024** |

|  |  |
| --- | --- |
| TRƯỜNG ĐẠI HỌC NGUYỄN TẤT THÀNH  **TRUNG TÂM KHẢO THÍ** | **KỲ THI KẾT THÚC HỌC PHẦN**  **HỌC KỲ 1 - NĂM HỌC 2024 - 2025** |

**PHIẾU CHẤM THI TIỂU LUẬN**

BM-ChT-11

Môn thi: **Chuyên đề chuyên sâu Trí Tuệ Nhân Tạo 2** Lớp học phần: **21DTH3D**

Nhóm sinh viên thực hiện:

1. **Hồ Công Chí (2100008526)** Tham gia đóng góp:

2. Tham gia đóng góp:

Ngày thi: **04/22/2025** Phòng thi:

Đề tài tiểu luận/báo cáo của sinh viên: **XÂY DỰNG HỆ THỐNG PHÂN LOẠI CÁC BỆNH VỀ MẮT SỬ DỤNG THUẬT TOÁN HỌC SÂU**

Phần đánh giá của giảng viên (căn cứ trên thang rubrics của môn học):

|  |  |  |  |
| --- | --- | --- | --- |
| **Tiêu chí (theo CĐR HP)** | **Đánh giá của GV** | **Điểm tối đa** | **Điểm đạt được** |
| Cấu trúc của báo cáo |  |  |  |
| Nội dung |  |  |  |
| - Các nội dung thành phần |  |  |  |
| - Lập luận |  |  |  |
| - Kết luận |  |  |  |
| Trình bày |  |  |  |
| **TỔNG ĐIỂM** |  |  |  |

|  |  |
| --- | --- |
|  | **Giảng viên chấm thi**  *(ký, ghi rõ họ tên)* |

# LỜI CẢM ƠN

Trước hết, em xin bày tỏ lòng biết ơn chân thành và sâu sắc đến thầy Thiện, người đã tận tình hướng dẫn, đồng hành và hỗ trợ em trong suốt quá trình thực hiện tiểu luận với đề tài "Xây dựng hệ thống phân loại các bệnh về mắt".

Trong suốt thời gian nghiên cứu và hoàn thiện bài tiểu luận, thầy không chỉ là người định hướng nội dung mà còn là nguồn động viên lớn lao, luôn sẵn sàng dành thời gian góp ý, chỉnh sửa và đưa ra những nhận xét quý báu để em có thể nhìn nhận vấn đề một cách sâu sắc và khoa học hơn. Những kiến thức, kinh nghiệm và tinh thần nghiêm túc trong học thuật mà thầy truyền đạt đã giúp em hoàn thiện bản thân không chỉ trong học tập mà còn trong tư duy nghiên cứu.

Đề tài về hệ thống phân loại các bệnh về mắt là một lĩnh vực vừa chuyên sâu vừa mang tính ứng dụng cao. Nếu không có sự chỉ bảo tận tâm và sự định hướng đúng đắn của thầy, em sẽ gặp rất nhiều khó khăn trong việc tiếp cận và xử lý thông tin. Chính sự tận tụy và nhiệt huyết trong công tác giảng dạy, nghiên cứu của thầy đã truyền cảm hứng để em cố gắng không ngừng, vượt qua những thử thách trong quá trình làm bài.

Em xin chân thành cảm ơn thầy vì tất cả những gì thầy đã truyền đạt và hỗ trợ.

# LỜI MỞ ĐẦU

Trong những năm gần đây, các bệnh lý về mắt đang ngày càng trở nên phổ biến và có xu hướng gia tăng do nhiều yếu tố như tuổi tác, môi trường sống, áp lực công việc và đặc biệt là thói quen sử dụng thiết bị điện tử trong thời gian dài. Việc phát hiện và chẩn đoán sớm các bệnh về mắt có vai trò vô cùng quan trọng trong việc bảo vệ thị lực và nâng cao chất lượng cuộc sống cho người bệnh.

Tuy nhiên, thực tế cho thấy việc chẩn đoán các bệnh lý về mắt vẫn đang phụ thuộc chủ yếu vào chuyên môn của bác sĩ và các thiết bị y tế hiện đại. Trong khi đó, tại nhiều khu vực, đặc biệt là vùng sâu vùng xa, việc tiếp cận với dịch vụ y tế chất lượng còn nhiều hạn chế. Điều này đặt ra yêu cầu cấp thiết về việc ứng dụng công nghệ vào lĩnh vực y tế, đặc biệt là trí tuệ nhân tạo (AI) và học máy (Machine Learning) trong việc hỗ trợ chẩn đoán và phân loại bệnh.

Xuất phát từ thực tế đó, tiểu luận với đề tài **“Xây dựng hệ thống phân loại các bệnh về mắt”** được thực hiện nhằm mục tiêu nghiên cứu, ứng dụng các kỹ thuật học sâu (Deep Learning) để xây dựng một hệ thống có khả năng tự động phân loại hình ảnh mắt theo các loại bệnh phổ biến. Qua đó, đề tài không chỉ mang ý nghĩa về mặt học thuật mà còn mở ra tiềm năng ứng dụng thực tiễn trong lĩnh vực chăm sóc sức khỏe cộng đồng.

Tiểu luận sẽ trình bày các cơ sở lý thuyết, phương pháp thực hiện, kết quả thu được và những đánh giá, nhận định về hệ thống được xây dựng. Mặc dù còn nhiều giới hạn, em hy vọng rằng đề tài sẽ là một bước khởi đầu để tiếp tục phát triển các nghiên cứu chuyên sâu hơn trong tương lai.

# MỤC LỤC

[LỜI CẢM ƠN 4](#_Toc196076809)

[LỜI MỞ ĐẦU 5](#_Toc196076810)

[MỤC LỤC 6](#_Toc196076811)

[DANH MỤC BẢNG 8](#_Toc196076812)

[DANH MỤC HÌNH ẢNH 9](#_Toc196076813)

[CHƯƠNG 1: 10](#_Toc196076814)

[GIỚI THIỆU ĐỀ TÀI 10](#_Toc196076815)

[1.1 Giới thiệu đề tài 10](#_Toc196076816)

[1.2 Lý do chọn đề tài 10](#_Toc196076817)

[1.3 Phương pháp và mục tiêu 11](#_Toc196076818)

[1.3.1 Phương pháp 11](#_Toc196076819)

[1.3.2 Mục tiêu 12](#_Toc196076820)

[CHƯƠNG 2: 14](#_Toc196076821)

[CƠ SỞ LÝ THUYẾT 14](#_Toc196076822)

[2.1 Deep learning 14](#_Toc196076823)

[2.1.1 Cách hoạt động Deep learning 14](#_Toc196076824)

[2.1.2 Ưu điểm và nhược điểm 14](#_Toc196076825)

[2.2 CNN (Convolutional Neural Network) 15](#_Toc196076826)

[2.2.1 Convolution Layer (Lớp tích chập) 16](#_Toc196076827)

[2.2.2 Bước nhảy - Stride 16](#_Toc196076828)

[2.2.3 Padding (Đường viền) 16](#_Toc196076829)

[2.2.4 Hàm phi tuyến - ReLU 17](#_Toc196076830)

[2.4.5 Lớp gộp - Pooling Layer 17](#_Toc196076831)

[2.3 MobileNets 18](#_Toc196076832)

[2.3.1 Standard Convolutions 19](#_Toc196076833)

[2.3.2 Depthwise Separable Convolutions 20](#_Toc196076834)

[2.4 MobileNetV2 21](#_Toc196076835)

[CHƯƠNG 3: 22](#_Toc196076836)

[MÔ HÌNH THỰC NGHIỆM 22](#_Toc196076837)

[3.1. Tổng quan dữ liệu 22](#_Toc196076838)

[3.2. Tiền xử lý dữ liệu 23](#_Toc196076839)

[3.2.1. Tiền xử lý hình ảnh 24](#_Toc196076840)

[3.2.2. Tăng cường dữ liệu (Data Augmentation) 24](#_Toc196076841)

[3.2.3. Phân chia tập dữ liệu 24](#_Toc196076842)

[3.3. Huấn luyện mô hình 25](#_Toc196076843)

[3.3.1. Xây dựng mô hình 25](#_Toc196076844)

[3.3.2. Huấn luyện mô hình 26](#_Toc196076845)

[3.3.3. Môi trường thực nghiệm 26](#_Toc196076846)

[3.4. Đánh giá mô hình 26](#_Toc196076847)

[3.4.1 Quá trình huấn luyện 27](#_Toc196076848)

[3.4.2 Kết quả trên tập test 27](#_Toc196076849)

[3.4.3 Ma trận nhầm lẫn (Confusion Matrix) 28](#_Toc196076850)

[KẾT LUẬN VÀ HƯỚNG CẢI TIẾN 29](#_Toc196076851)

[1. Kết Luận 29](#_Toc196076852)

[2. Hướng Cải Tiến 29](#_Toc196076853)

[TÀI LIỆU THAM KHẢO 30](#_Toc196076854)

# DANH MỤC BẢNG

**No table of figures entries found.**

# DANH MỤC HÌNH ẢNH

[Hình 1. Kiến trúc CNN 15](#_Toc196076396)

[Hình 2. Tích chập 16](#_Toc196076397)

[Hình 3. Trượt bộ lọc 17](#_Toc196076398)

[Hình 4. Max pooling với kernel 2x2 18](#_Toc196076399)

[Hình 5. Sự khác biệt giữa 2 loại tích chập 18](#_Toc196076400)

[Hình 6. sự khác biệt giữa standard CNN và kiến trúc MobileNet 20](#_Toc196076401)

[Hình 7. Kiến trúc Inverted Residual 21](#_Toc196076402)

[Hình 8. Cataract 22](#_Toc196076403)

[Hình 9. Diabetic Retinopathy 22](#_Toc196076404)

[Hình 10. Macular Scar 23](#_Toc196076405)

[Hình 11. Myopia 23](#_Toc196076406)

[Hình 12. Biểu đồ Pie Chart thể hiện phân bố dữ liệu 23](#_Toc196076407)

[Hình 13. Biểu đồ Bar Chart thể hiện phân chia dữ liệu huấn luyện 24](#_Toc196076408)

[Hình 14. Xây dựng mô hình 25](#_Toc196076409)

[Hình 15. Biểu đồ Accuracy/Loss 27](#_Toc196076410)

[Hình 16. Confusion Matrix 28](#_Toc196076411)

# CHƯƠNG 1:

# GIỚI THIỆU ĐỀ TÀI

## Giới thiệu đề tài

Mắt là một trong những cơ quan quan trọng nhất trong cơ thể con người, giúp chúng ta tiếp nhận thông tin và tương tác với thế giới xung quanh. Tuy nhiên, mắt cũng dễ bị ảnh hưởng bởi nhiều bệnh lý khác nhau, có thể gây suy giảm thị lực hoặc thậm chí dẫn đến mù lòa nếu không được chẩn đoán và điều trị kịp thời. Các bệnh về mắt ngày càng trở nên phổ biến, đặc biệt là trong bối cảnh lão hóa dân số và sự gia tăng các yếu tố môi trường như ô nhiễm, thói quen sinh hoạt không lành mạnh, và sự phát triển mạnh mẽ của công nghệ điện tử.

Phân loại các bệnh về mắt là một nhiệm vụ quan trọng trong lĩnh vực y học, giúp các bác sĩ và chuyên gia y tế nhận diện và điều trị sớm các bệnh lý, từ đó giảm thiểu các biến chứng nghiêm trọng. Các bệnh lý về mắt có thể chia thành nhiều nhóm khác nhau, bao gồm các bệnh về giác mạc, võng mạc, bệnh lý về dây thần kinh thị giác, và các bệnh khác như đục thủy tinh thể hay glocom. Trong thời đại hiện nay, nhờ vào sự phát triển của công nghệ hình ảnh và các phương pháp chẩn đoán tự động, việc phân loại và phát hiện các bệnh lý về mắt đã trở nên chính xác và nhanh chóng hơn.

Đề tài "Phân loại các loại bệnh về mắt" không chỉ mang lại giá trị trong việc cải thiện chất lượng chăm sóc sức khỏe mà còn mở ra cơ hội nghiên cứu về việc áp dụng các công nghệ hiện đại như học máy (machine learning) và trí tuệ nhân tạo (AI) trong việc phát triển các hệ thống chẩn đoán tự động. Đây là một hướng đi mới, tiềm năng giúp giảm tải cho các bác sĩ và tăng khả năng phát hiện bệnh lý ở giai đoạn sớm, từ đó nâng cao hiệu quả điều trị và bảo vệ thị lực cho người bệnh.

## 1.2 Lý do chọn đề tài

Đề tài "Phân loại các loại bệnh về mắt" được chọn vì tầm quan trọng của mắt trong cuộc sống hàng ngày và tác động sâu rộng của các bệnh lý mắt đến sức khỏe cộng đồng. Mắt không chỉ là cơ quan giúp con người tiếp nhận thông tin và hình ảnh từ thế giới bên ngoài, mà các bệnh về mắt như đục thủy tinh thể, glocom, và bệnh lý võng mạc có thể dẫn đến suy giảm thị lực nghiêm trọng hoặc mù lòa nếu không được phát hiện và điều trị kịp thời. Với tình trạng gia tăng các bệnh lý về mắt trong xã hội hiện đại, việc phân loại và chẩn đoán sớm sẽ giúp giảm thiểu nguy cơ mù lòa, đồng thời nâng cao chất lượng cuộc sống cho người bệnh.

Hơn nữa, sự phát triển nhanh chóng của công nghệ hình ảnh y tế và trí tuệ nhân tạo (AI) mở ra cơ hội ứng dụng các phương pháp phân loại bệnh mắt tự động, giúp tăng độ chính xác, tiết kiệm thời gian và giảm thiểu sai sót trong chẩn đoán. Các mô hình học máy, đặc biệt là học sâu (deep learning), có thể tự động phân tích hình ảnh mắt và phát hiện các dấu hiệu bệnh lý một cách hiệu quả, giúp bác sĩ đưa ra quyết định chẩn đoán chính xác và kịp thời hơn.

Ngoài việc cải thiện quy trình chẩn đoán, việc phân loại bệnh mắt còn có ý nghĩa lớn trong việc phát triển các công nghệ chăm sóc sức khỏe từ xa, mang lại sự tiếp cận dễ dàng hơn cho người dân ở các khu vực xa xôi hoặc những nơi thiếu chuyên gia. Đề tài này không chỉ mang lại giá trị nghiên cứu mà còn đóng góp vào sự cải thiện chất lượng chăm sóc sức khỏe cộng đồng, bảo vệ thị lực cho hàng triệu người trên thế giới.

## 1.3 Phương pháp và mục tiêu

### 1.3.1 Phương pháp

Để thực hiện đề tài phân loại các loại bệnh về mắt, chúng tôi sẽ sử dụng mô hình MobileNetV2, một trong những mô hình học sâu (deep learning) hiệu quả và nhẹ nhàng, đặc biệt phù hợp cho các ứng dụng trên thiết bị di động và các môi trường tài nguyên hạn chế. MobileNetV2 là một biến thể của mạng nơ-ron tích chập (Convolutional Neural Network - CNN), nổi bật với khả năng giảm thiểu số lượng tham số mà vẫn đảm bảo hiệu suất phân loại cao.

Các bước cụ thể trong phương pháp nghiên cứu bao gồm:

**Bước 1. Thu thập và xử lý dữ liệu:**

Chúng tôi sẽ sử dụng các bộ dữ liệu hình ảnh mắt có sẵn từ các nguồn uy tín như Kaggle hoặc các cơ sở dữ liệu y tế chuyên ngành, bao gồm các hình ảnh về các bệnh lý mắt cần phân loại (ví dụ: đục thủy tinh thể, glocom, bệnh lý võng mạc, tật khúc xạ, viêm mắt, và bệnh giác mạc).

Các hình ảnh này sẽ được tiền xử lý, bao gồm chuẩn hóa kích thước, chuyển sang màu xám hoặc các không gian màu khác nếu cần thiết, và tăng cường dữ liệu (data augmentation) để tăng cường tính đa dạng và độ chính xác của mô hình.

**Bước 2. Xây dựng và huấn luyện mô hình:**

MobileNetV2 sẽ được sử dụng làm mô hình chính cho việc phân loại các bệnh mắt. Chúng tôi sẽ sử dụng kỹ thuật transfer learning, tức là tận dụng các trọng số đã được huấn luyện trước của MobileNetV2 trên các bộ dữ liệu lớn, sau đó tinh chỉnh lại cho phù hợp với bộ dữ liệu bệnh lý mắt.

Các lớp đầu tiên của mô hình sẽ được giữ nguyên, trong khi các lớp cuối cùng sẽ được thay đổi để phù hợp với việc phân loại 4 loại bệnh lý mắt. Quá trình huấn luyện sẽ được thực hiện với các bộ dữ liệu đã được chia thành tập huấn luyện và tập kiểm tra, sử dụng các chỉ số như độ chính xác (accuracy), độ nhạy (sensitivity) và độ đặc hiệu (specificity) để đánh giá hiệu quả của mô hình.

**Bước 3. Tuning và tối ưu hóa mô hình:**

Chúng tôi sẽ sử dụng các kỹ thuật tối ưu hóa như Adam optimizer và điều chỉnh các siêu tham số của mô hình như tỷ lệ học (learning rate), batch size, và số lượng epochs để đạt được hiệu suất tốt nhất.

Cross-validation sẽ được sử dụng để đánh giá mô hình và tránh hiện tượng overfitting.

**Bước 4. Đánh giá mô hình:**

Sau khi huấn luyện, mô hình sẽ được đánh giá trên tập kiểm tra (test set) chưa được sử dụng trong quá trình huấn luyện để kiểm tra khả năng tổng quát của mô hình.

Chúng tôi sẽ sử dụng các chỉ số đánh giá như confusion matrix, ROC curve, và AUC (Area Under Curve) để đo lường độ chính xác và hiệu quả phân loại của mô hình.

### 1.3.2 Mục tiêu

Mục tiêu của đề tài này là xây dựng và phát triển một mô hình phân loại hình ảnh bệnh lý mắt bằng cách sử dụng MobileNetV2. Các mục tiêu cụ thể bao gồm:

**Phân loại chính xác loại bệnh lý mắt:**

Phân loại chính xác các bệnh về mắt như đục thủy tinh thể, glocom, bệnh lý võng mạc, tật khúc xạ, viêm mắt và bệnh giác mạc từ hình ảnh mắt.

**Tối ưu hóa mô hình MobileNetV2:**

Sử dụng kỹ thuật transfer learning để tối ưu hóa mô hình MobileNetV2 cho việc phân loại các bệnh về mắt, với mục tiêu đạt độ chính xác cao trong việc nhận diện và phân loại hình ảnh bệnh lý.

**Đánh giá mô hình với các chỉ số chẩn đoán:**

Đánh giá mô hình dựa trên các chỉ số như độ chính xác, độ nhạy, độ đặc hiệu, và AUC để đảm bảo rằng mô hình có thể phân loại chính xác các bệnh lý mắt và có khả năng ứng dụng trong môi trường thực tế.

**Ứng dụng trong chẩn đoán tự động:**

Xây dựng một hệ thống có thể hỗ trợ bác sĩ và chuyên gia y tế trong việc phát hiện và chẩn đoán sớm các bệnh lý mắt thông qua hình ảnh, từ đó nâng cao hiệu quả điều trị và bảo vệ sức khỏe người bệnh.

**Tính khả thi và hiệu quả của mô hình:**

Đảm bảo rằng mô hình có thể hoạt động hiệu quả trên các thiết bị có tài nguyên hạn chế, đặc biệt là các ứng dụng di động hoặc các hệ thống chẩn đoán y tế từ xa, góp phần vào việc mở rộng khả năng tiếp cận chăm sóc sức khỏe cho người dân.

# CHƯƠNG 2:

# CƠ SỞ LÝ THUYẾT

## 2.1 Deep learning

Deep learning (học sâu) là một nhánh của machine learning (học máy), tập trung vào việc xây dựng và huấn luyện mạng nơ-ron nhân tạo với nhiều lớp (deep neural networks) để mô phỏng cách con người học từ dữ liệu.

### 2.1.1 Cách hoạt động Deep learning

Mạng thần kinh nhân tạo trong Deep Learning được xây dựng để mô phỏng khả năng tư duy của bộ não con người. Cách thức hoạt động như sau:

* Một mạng nơ-ron bao gồm nhiều lớp (layer) khác nhau, càng nhiều lớp thì mạng sẽ càng “sâu”. Trong mỗi lớp là các node (nút mạng) và được liên kết với các lớp khác bên cạnh.
* Mỗi kết nối giữa các nút sẽ có một trọng số tương ứng, trọng số càng cao thì mức độ ảnh hưởng của kết nối này đến mạng nơ-ron càng lớn.
* Mỗi nơ-ron sẽ có một chức năng kích hoạt, về cơ bản chịu trách nhiệm "chuẩn hóa" đầu ra từ nơ-ron này.
* Dữ liệu được người dùng nhập vào mạng thần kinh đi qua tất cả các lớp và trả về kết quả ở layer cuối cùng, được gọi là lớp đầu ra (output layer).
* Trong quá trình đào tạo mô hình mạng nơ-ron, các trọng số sẽ được thay đổi và nhiệm vụ của mô hình là tìm tập giá trị của các trọng số sao cho phán đoán đúng nhất.

Các hệ thống Deep Learning yêu cầu phần cứng rất mạnh để có thể xử lý lượng lớn dữ liệu và thực hiện các phép tính phức tạp. Nhiều mô hình Deep Learning có thể mất hàng tuần hoặc thậm chí hàng tháng để triển khai trên phần cứng tiên tiến nhất hiện nay.

### 2.1.2 Ưu điểm và nhược điểm

Một số ưu điểm nổi bật của Deep Learning bao gồm:

* Cấu trúc neural networks linh hoạt, dễ dàng thay đổi để phù hợp với nhiều thuật toán khác nhau.
* Có khả năng giải nhiều vấn đề phức tạp với độ chính xác rất cao.
* Khả năng tự động hóa cao, tự điều chỉnh và tự tối ưu hóa.
* Có khả năng thực hiện tính toán song song, hiệu năng tốt, xử lý lượng dữ liệu lớn.

Bên cạnh ưu điểm, Deep Learning vẫn còn tồn tại nhiều hạn chế:

* Cần khối lượng dữ liệu khổng lồ để khai thác tối đa khả năng của Deep Learning.
* Chi phí tính toán cao vì phải xử lý nhiều mô hình phức tạp.
* Không có nền tảng lý thuyết vững chắc để chọn các công cụ tối ưu cho Deep Learning.

## 2.2 CNN (Convolutional Neural Network)

Convolutional Neural Network (CNNs – Mạng nơ-ron tích chập) là một trong những mô hình Deep Learning tiên tiến. Nó giúp cho chúng ta xây dựng được những hệ thống thông minh với độ chính xác cao như hiện nay.

A diagram of a diagram of a diagram

AI-generated content may be incorrect.

Hình 1. Kiến trúc CNN

CNN được sử dụng nhiều trong các bài toán nhận dạng các object trong ảnh. Để tìm hiểu tại sao thuật toán này được sử dụng rộng rãi cho việc nhận dạng (detection), chúng ta hãy cùng tìm hiểu về thuật toán này.

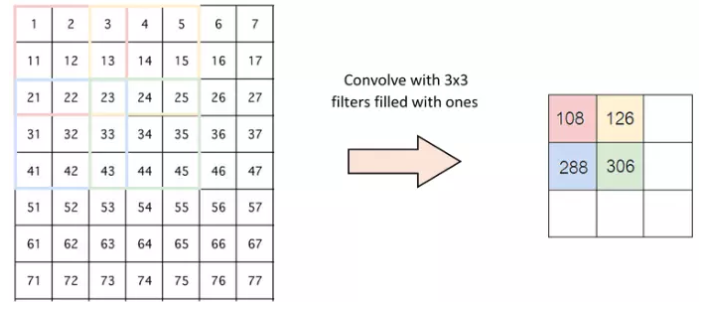
Các lớp cơ bản của thuật toán CNN:

### 2.2.1 Convolution Layer (Lớp tích chập)

Tích chập là lớp đầu tiên để trích xuất các tính năng từ hình ảnh đầu vào. Tích chập duy trì mối quan hệ giữa các pixel bằng cách tìm hiểu các tính năng hình ảnh bằng cách sử dụng các ô vương nhỏ của dữ liệu đầu vào. Nó là 1 phép toán có 2 đầu vào như ma trận hình ảnh và 1 bộ lọc hoặc hạt nhân.

### 2.2.2 Bước nhảy - Stride

Stride là số pixel thay đổi trên ma trận đầu vào. Khi stride là 1 thì ta di chuyển các kernel 1 pixel. Khi stride là 2 thì ta di chuyển các kernel đi 2 pixel và tiếp tục như vậy. Hình dưới là lớp tích chập hoạt động với stride là 2.



Hình 2. Tích chập

Hình ảnh này minh họa cách một bộ lọc 3x3 chứa toàn số 1 được sử dụng để thực hiện phép tích chập trên một ma trận đầu vào. Kết quả của phép tích chập là một ma trận đầu ra, trong đó mỗi phần tử là tổng của các phần tử tương ứng trong vùng 3x3 của ma trận đầu vào. Phép tích chập là một phép toán cơ bản trong xử lý ảnh và mạng nơ-ron tích chập

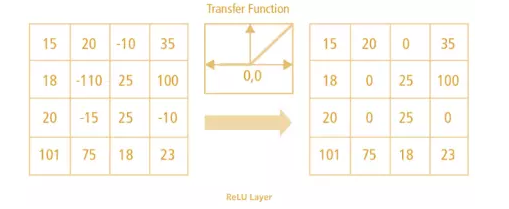
### 2.2.3 Padding (Đường viền)

Trong trường hợp kernel không tương xứng với hình ảnh đầu vào thì có thể lựa chọn 1 trong 2 cách sau:

* Chèn thêm số 0 vào 4 đường biên hình ảnh
* Cắt bớt những điểm không phù hợp với kernel

### 2.2.4 Hàm phi tuyến - ReLU

* ReLU viết tắt của Rectified Linear Unit, là 1 hàm phi tuyến. Với đầu ra là: ƒ (x) = max (0, x).
* Tại sao ReLU lại quan trọng: ReLU giới thiệu tính phi tuyến trong ConvNet. Vì dữ liệu trong thế giới mà chúng ta tìm hiểu là các giá trị tuyến tính không âm.



Hình 3. Trượt bộ lọc

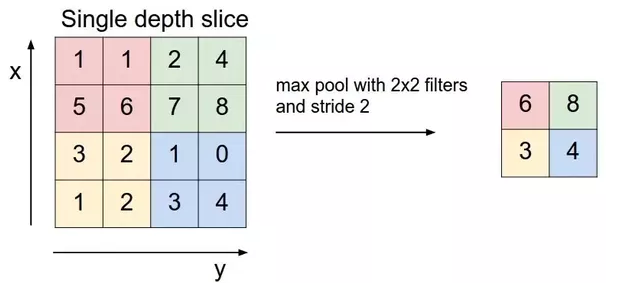
Hình ảnh này minh họa một bước cơ bản trong xử lý ảnh và mạng nơ-ron tích chập (CNNs), trong đó một bộ lọc được trượt qua ma trận đầu vào để tạo ra một ma trận đầu ra, với mỗi phần tử trong ma trận đầu ra là kết quả của phép toán tích chập giữa bộ lọc và một vùng tương ứng trên ma trận đầu vào. Trong trường hợp này, bộ lọc là một ma trận 3x3 chứa toàn bộ giá trị là 1, dẫn đến việc mỗi phần tử đầu ra là tổng của 9 phần tử lân cận trong ma trận đầu vào.

### 2.4.5 Lớp gộp - Pooling Layer

Lớp pooling sẽ giảm bớt số lượng tham số khi hình ảnh quá lớn. Không gian pooling còn được gọi là lấy mẫu con hoặc lấy mẫu xuống làm giảm kích thước của mỗi map nhưng vẫn giữ lại thông tin quan trọng. Các pooling có thể có nhiều loại khác nhau:

* Max Pooling
* Average Pooling
* Sum Pooling

Max pooling lấy phần tử lớn nhất từ ma trận đối tượng, hoặc lấy tổng trung bình. Tổng tất cả các phần tử trong map gọi là sum pooling

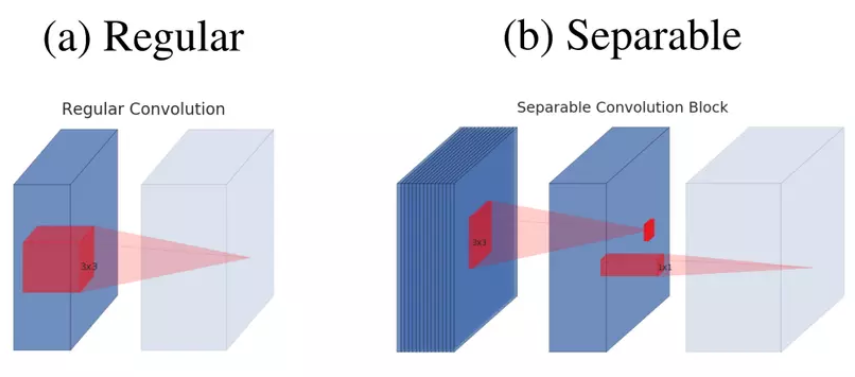


Hình 4. Max pooling với kernel 2x2

Hình ảnh này minh họa cách phép max pooling với bộ lọc 2x2 và bước nhảy 2 làm giảm kích thước của một lát cắt độ sâu và giữ lại các giá trị lớn nhất trong các vùng không chồng lấn. Đây là một kỹ thuật quan trọng trong xây dựng các mạng CNN hiệu quả.

## 2.3 MobileNets

MobileNet là một mô hình kiến trúc của mạng nơ-ron tích chập (CNN) tập trung rõ ràng vào Phân loại hình ảnh cho các ứng dụng di động. Thay vì sử dụng các lớp tích chập tiêu chuẩn, nó sử dụng các lớp tích chập tách biệt chiều sâu (Depthwise Separable Convolutions).



Hình 5. Sự khác biệt giữa 2 loại tích chập

Hình ảnh này minh họa sự khác biệt giữa hai loại phép tích chập (convolution) thường dùng trong các mạng nơ-ron tích chập (CNN):

1. **Regular (Tích chập thông thường/Tiêu chuẩn):**

* Hình ảnh cho thấy một bộ lọc (filter) duy nhất (khối màu đỏ, ví dụ kích thước 3x3) có chiều sâu bằng với chiều sâu của dữ liệu đầu vào (khối màu xanh lam đậm).
* Bộ lọc này trượt trên toàn bộ dữ liệu đầu vào và thực hiện phép tính tích chập trên cả không gian (chiều rộng, chiều cao) và chiều sâu (số kênh) cùng một lúc để tạo ra một điểm (hoặc một kênh) trong bản đồ đặc trưng đầu ra (khối màu xanh lam nhạt).
* Để tạo ra nhiều kênh ở đầu ra, bạn cần nhiều bộ lọc 3D như vậy.
* Đây là phép tích chập tiêu chuẩn, mạnh mẽ nhưng tốn kém về mặt tính toán và số lượng tham số.

1. **Separable (Tích chập tách biệt/Depthwise Separable Convolution):**

* Đây là loại tích chập được sử dụng trong các kiến trúc hiệu quả như MobileNet. Nó chia phép tích chập thông thường thành hai bước riêng biệt:
* Bước 1: Depthwise Convolution (Tích chập theo chiều sâu): Áp dụng một bộ lọc 2D (ví dụ 3x3, các hình vuông đỏ nhỏ) cho từng kênh đầu vào một cách độc lập. Nếu đầu vào có 10 kênh, bạn sẽ có 10 bộ lọc 2D riêng biệt, mỗi bộ lọc chỉ xử lý 1 kênh. Bước này chỉ lọc về mặt không gian (spatial filtering). Khối màu xanh lam ở giữa đại diện cho kết quả sau bước này, có cùng số kênh với đầu vào.
* Bước 2: Pointwise Convolution (Tích chập điểm): Áp dụng các bộ lọc có kích thước 1x1 (hình chữ nhật đỏ nhỏ, sâu) lên kết quả của bước Depthwise. Các bộ lọc 1x1 này hoạt động trên tất cả các kênh để kết hợp thông tin từ các kênh đã được lọc độc lập ở bước trước, tạo ra các đặc trưng mới ở đầu ra (khối màu xanh lam nhạt). Bước này thực hiện việc kết hợp kênh (channel combination).

### 2.3.1 Standard Convolutions

Standard Convolutions là tích chập tiêu chuẩn thực hiện lọc không gian và lọc kênh đồng thời. Dù MobileNet V1 chủ yếu dùng Depthwise Separable Convolutions để giảm chi phí, Standard Convolutions vẫn được dùng ở lớp đầu tiên và lớp 1x1 cuối để trích xuất đặc trưng ban đầu và kết hợp thông tin giữa các kênh.

### 2.3.2 Depthwise Separable Convolutions

Depthwise Separable Convolution là một kỹ thuật có thể cải thiện hiệu quả và tốc độ của các mô hình học sâu. Bằng cách chia quá trình tính toán thành hai bước, nó cho phép các mô hình học các tính năng tổng quát hơn, giảm tình trạng quá khớp và sử dụng tài nguyên tốt hơn.

1. **Deepwise Convolution**

Depthwise Convolution là một phép tích chập mà trong đó mỗi kênh đầu vào được lọc bởi một bộ lọc riêng biệt,bộ lọc sâu như đầu vào và cho phép chúng ta tự do trộn các kênh để tạo ra từng phần tử trong đầu ra. Giúp giảm đáng kể chi phí tính toán so với tích chập tiêu chuẩn bằng cách tách biệt quá trình lọc không gian với quá trình kết hợp kênh.

1. **Pointwise Convolution**

Pointwise Convolution là một loại hoạt động tích chập trong học sâu sử dụng hạt nhân 1x1. Hạt nhân này lặp qua từng điểm trong dữ liệu đầu vào, thực sự áp dụng phép biến đổi tuyến tính cho từng pixel một cách độc lập trên toàn bộ chiều sâu của đầu vào. Chiều sâu của hạt nhân bằng với số kênh trong hình ảnh đầu vào

A diagram of a computer

AI-generated content may be incorrect.

Hình 6. sự khác biệt giữa standard CNN và kiến trúc MobileNet

Khối tích chập tiêu chuẩn thực hiện lọc không gian và lọc kênh đồng thời trong một bước.

Khối Depthwise Separable Convolution tách quá trình này thành hai bước:

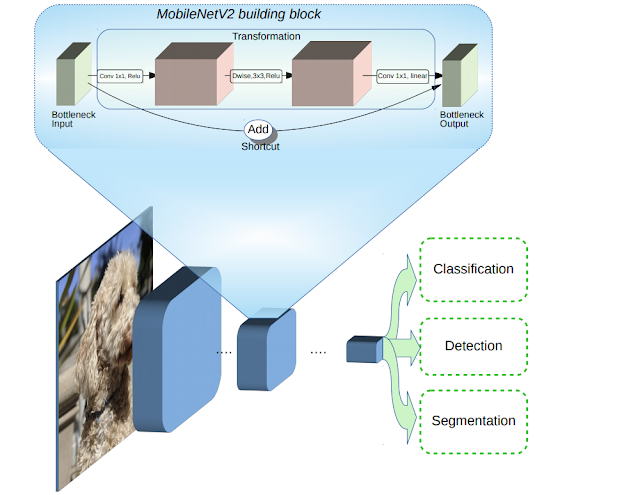
* Depthwise Convolution: Lọc không gian độc lập trên từng kênh.
* Pointwise Convolution: Kết hợp thông tin giữa các kênh.

## 2.4 MobileNetV2

MobileNetV2 là kiến trúc mạng nơ-ron tích chập được tối ưu hóa cho các ứng dụng thị giác di động và nhúng. Nó cải thiện MobileNet ban đầu bằng cách đưa vào các khối Inverted Residual và linear bottleneck, mang lại độ chính xác và tốc độ cao hơn trong khi vẫn duy trì chi phí tính toán thấp.

MobileNet v2 tiếp tục sử dụng Depthwise Separable Convolutions, ngoài ra còn đề xuất thêm:

* Linear bottlenecks
* Inverted Residual Block (shortcut connections giữa các bottlenecks)



Hình 7. Kiến trúc Inverted Residual

# CHƯƠNG 3:

# MÔ HÌNH THỰC NGHIỆM

## 3.1. Tổng quan dữ liệu

Trong bài toán phân loại bệnh về mắt, chất lượng và đặc điểm của bộ dữ liệu đóng vai trò then chốt trong việc quyết định hiệu quả của mô hình. Bộ dữ liệu được sử dụng trong dự án này là tập hợp các hình ảnh bệnh lý về mắt được thu thập và xây dựng lại từ các nguồn khác nhau được công khai trên Internet, đặc biệt là từ các kho dữ liệu y khoa.

Link nguồn cụ thể:

Nguồn 1: Mendeley Data: https://data.mendeley.com/datasets/s9bfhswzjb/1

Nguồn 2: https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification

Bộ dữ liệu bao gồm 4 loại bệnh về mắt phổ biến:

* Cataract (Đục thủy tinh thể)
* Diabetic Retinopathy (Bệnh võng mạc do tiểu đường)
* Macular Scar (Sẹo hoàng điểm)
* Myopia (cận thị)

A close-up of a human eye

AI-generated content may be incorrect.

Hình 8. Cataract

A close-up of a lens

AI-generated content may be incorrect.

Hình 9. Diabetic Retinopathy

A close-up of a red blood vessels

AI-generated content may be incorrect.

Hình 10. Macular Scar

A close-up of a red blood vessels

AI-generated content may be incorrect.

Hình 11. Myopia

Tổng cộng, bộ dữ liệu có khoảng 4000 hình ảnh, được phân bố đồng đều giữa các lớp. Sự phân bố đồng đều dữ liệu giữa các loại bệnh giúp giảm thiểu tình trạng mất cân bằng dữ liệu, một yếu tố có thể làm sai lệch quá trình học của mô hình. Các hình ảnh có độ phân giải khác nhau nên cần được xử lý đồng bộ trước khi đưa vào mô hình huấn luyện. Do đó, việc xử lý dữ liệu là bước không thể thiếu nhằm cải thiện chất lượng đầu vào cho mô hình học sâu.

A green and blue circle with white text

AI-generated content may be incorrect.

Hình 12. Biểu đồ Pie Chart thể hiện phân bố dữ liệu

## 3.2. Tiền xử lý dữ liệu

Trước khi đưa vào mô hình huấn luyện, dữ liệu hình ảnh cần trải qua nhiều bước tiền xử lý để đảm bảo tính nhất quán và hỗ trợ mô hình học được các đặc trưng quan trọng.

### 3.2.1. Tiền xử lý hình ảnh

Đầu tiên, toàn bộ ảnh được resize về kích thước cố định 256x256 pixel để phù hợp với kiến trúc của MobileNetV2.

### 3.2.2. Tăng cường dữ liệu (Data Augmentation)

Để cải thiện hiệu quả của mô hình, kỹ thuật tăng cường dữ liệu (data augmentation) được áp dụng để cải thiện khả năng tổng quát của mô hình. Các phép biến đổi như xoay ảnh, dịch ảnh, lật ngang, zoom, thay đổi độ sáng, và cắt ảnh ngẫu nhiên được sử dụng. Điều này không chỉ làm tăng số lượng mẫu dữ liệu mà còn giúp mô hình học được sự đa dạng trong biểu hiện của các loại bệnh.

### 3.2.3. Phân chia tập dữ liệu

Dữ liệu được chia thành hai phần:

Tập huấn luyện (train): 80% (3200 hình ảnh)

Tập kiểm định (validation): 20% (800 hình ảnh)

A blue and orange rectangular chart

AI-generated content may be incorrect.

Hình 13. Biểu đồ Bar Chart thể hiện phân chia dữ liệu huấn luyện

Việc phân chia dữ liệu được thực hiện ngẫu nhiên nhưng đảm bảo duy trì tỷ lệ cân bằng giữa các lớp trong cả tập huấn luyện và tập kiểm định.

## 3.3. Huấn luyện mô hình

Mô hình được lựa chọn trong dự án này là MobileNetV2, một kiến trúc mạng nơ-ron tích chập nhẹ, được thiết kế tối ưu cho các thiết bị di động hoặc hệ thống có tài nguyên hạn chế. MobileNetV2 có đặc điểm sử dụng Inverted Residuals và Linear Bottleneck kết hợp với depthwise separable convolutions, giúp giảm số lượng tham số mà vẫn duy trì hiệu quả nhận diện cao.

### 3.3.1. Xây dựng mô hình

Trong thực nghiệm này, mô hình MobileNetV2 được sử dụng như một mạng nền (base model) với các trọng số được đào tạo từ tập dữ liệu ImageNet. Các lớp đầu của mô hình được giữ nguyên nhằm khai thác các đặc trưng thị giác cơ bản như cạnh, họa tiết, màu sắc,... đã được học từ ImageNet.

Lớp phân loại (classifier) của mô hình được thiết kế lại để phù hợp với bài toán phân loại 4 loại bệnh về mắt. Cấu trúc cụ thể bao gồm:

* Một lớp GlobalAveragePooling2D dùng để giảm chiều đầu ra từ MobileNetV2 và làm nổi bật các đặc trưng quan trọng.
* Tiếp theo là một lớp Dense 512 neuron với hàm kích hoạt ReLU, kèm theo kỹ thuật chuẩn hóa L2 Regularization (hệ số 0.01) nhằm giảm hiện tượng overfitting.
* Sau đó, một lớp Dropout 0.2 được thêm vào để ngẫu nhiên bỏ bớt 20% số node trong quá trình huấn luyện, giúp tăng khả năng tổng quát hóa của mô hình.
* Cuối cùng, lớp Dense đầu ra gồm 4 node tương ứng với 4 lớp bệnh, sử dụng hàm kích hoạt softmax để đưa ra xác suất cho từng lớp.

A computer code with text

AI-generated content may be incorrect.

Hình 14. Xây dựng mô hình

### 3.3.2. Huấn luyện mô hình

Trước khi huấn luyện, các lớp được "freeze" để giúp giữ nguyên trọng số đã học từ ImageNet, chỉ huấn luyện phần lớp phân loại mới được thêm vào phía sau (classification head).

Tham số huấn luyện gồm:

* Optimizer: Adam
* Learning rate: 0.0001
* Batch size: 32
* Epochs: 15
* Loss function: categorical crossentropy

Trong quá trình huấn luyện, mô hình sử dụng thêm các cơ chế:

EarlyStopping: Dừng sớm nếu độ chính xác trên tập validation không cải thiện sau mỗi 5 epoch nhất định.

ModelCheckpoint: Tự động lưu trọng số mô hình tốt nhất dựa trên độ chính xác cao nhất trên tập validation.

CSVLogger: Lưu lại kết quả quá trình huấn luyện để trực quan kết quả.

ReduceLROnPlateau: Giảm learning\_rate nếu val\_loss không giảm sau mỗi 3 epoch.

### 3.3.3. Môi trường thực nghiệm

Việc huấn luyện được thực hiện trên nền tảng Google Colab với GPU miễn phí (Tesla T4).

## 3.4. Đánh giá mô hình

Đánh giá mô hình là bước quan trọng để xác định mức độ hiệu quả của thuật toán trong việc phân loại các loại bệnh về mắt. Để đánh giá tổng quan mô hình, ta sử dụng các chỉ số:

### 3.4.1 Quá trình huấn luyện

A graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of

AI-generated content may be incorrect.

Hình 15. Biểu đồ Accuracy/Loss

Biểu đồ Accuracy/Loss cho thấy mô hình đạt độ chính xác tốt dần theo thời gian huấn luyện. Tuy nhiên, mô hình vẫn chưa đạt mức tối ưu hết mức, có thể do sự tương đồng giữa hình ảnh của các bệnh lý khác nhau hoặc do kiến trúc mạng chưa đủ phức tạp để phân biệt chi tiết hơn.

### 3.4.2 Kết quả trên tập test

Trên tập kiểm định, mô hình đạt accuracy trung bình khoảng 82%, trong đó các lớp như Cataract, Diabetic Retinopathy đạt độ chính xác cao, trong khi các lớp như Macular Scar, Myopia thấp hơn do đặc trưng hình ảnh kém rõ ràng hơn.

### 3.4.3 Ma trận nhầm lẫn (Confusion Matrix)

A blue squares with white text

AI-generated content may be incorrect.

Hình 16. Confusion Matrix

Confusion matrix cho thấy mô hình có xu hướng nhầm lẫn giữa một số bệnh có biểu hiện tương tự, đặc biệt là giữa Myopia và Sẹo Hoàng Điểm Macular Scar. Điều này chỉ ra rằng mô hình có thể cần thêm dữ liệu hoặc điều chỉnh kiến trúc để cải thiện hiệu suất.

# KẾT LUẬN VÀ HƯỚNG CẢI TIẾN

## 1. Kết Luận

Phân loại bệnh về mắt thông qua hình ảnh y tế là một bài toán quan trọng và mang tính ứng dụng cao trong lĩnh vực chăm sóc sức khỏe. Với sự phát triển mạnh mẽ của học sâu (deep learning), việc ứng dụng các mô hình mạng nơ-ron tích chập đã mở ra hướng tiếp cận mới cho việc chẩn đoán bệnh lý từ ảnh. Trong khuôn khổ bài tiểu luận này, nhóm em đã triển khai mô hình MobileNetV2 để nhận diện và phân loại 4 loại bệnh về mắt, đồng thời đánh giá hiệu quả của mô hình thông qua quá trình huấn luyện và kiểm thử.

Kết quả thực nghiệm cho thấy mô hình đạt được độ chính xác tốt trên tập huấn luyện và có khả năng tổng quát tương đối trên tập kiểm thử. Tuy nhiên, một số khó khăn vẫn còn tồn tại, như sự tương đồng cao giữa các bệnh lý khiến mô hình khó phân biệt. Dù vậy, kết quả đạt được đã phần nào chứng minh tính khả thi trong việc ứng dụng mô hình học sâu vào hỗ trợ chẩn đoán bệnh về mắt, mở ra nhiều tiềm năng phát triển trong tương lai.

## 2. Hướng Cải Tiến

Trong tương lai, hệ thống có thể được mở rộng để phân loại nhiều loại bệnh về mắt hơn, hoặc kết hợp thêm thông tin lâm sàng để tăng độ chính xác chẩn đoán. Việc xây dựng bộ dữ liệu hoàn thiện, chất lượng hơn cũng giúp mô hình học được đặc trưng của từng loại bệnh một cách khách quan và hiệu quả hơn. Ngoài ra, mô hình có thể được tích hợp vào các hệ thống quản lý y tế, hỗ trợ bác sĩ trong việc lưu trữ, truy xuất và đánh giá tình trạng bệnh nhân nhanh chóng. Điều này không chỉ giúp tối ưu hóa quy trình làm việc của đội ngũ y tế mà còn nâng cao chất lượng chăm sóc sức khỏe bệnh nhân.

Bên cạnh đó, việc mở rộng mô hình để phân loại thêm các dạng bệnh lý khác trong lĩnh vực y khoa nói chung sẽ góp phần nâng cao khả năng ứng dụng thực tiễn của hệ thống. Những cải tiến này sẽ không chỉ nâng cao hiệu quả trong việc chẩn đoán và điều trị mà còn đóng góp vào sự phát triển của y học nước nhà.

# TÀI LIỆU THAM KHẢO

1. Nguyễn Đức Nghĩa. (2021). Xử lý ảnh và ứng dụng trong Y tế. Nhà xuất bản Khoa học và Kỹ thuật.
2. Trần Văn Lợi, Nguyễn Hữu Đức. (2020). Trí tuệ nhân tạo: Nguyên lý và ứng dụng. Nhà xuất bản Đại học Quốc gia TP. Hồ Chí Minh.
3. Bùi Thị Minh Hằng. (2022). Nghiên cứu áp dụng học sâu trong chẩn đoán bệnh võng mạc tiểu đường từ ảnh đáy mắt. Luận văn Thạc sĩ, Trường Đại học Bách Khoa Hà Nội.
4. Bộ môn Khoa học Máy tính – Trường Đại học Công nghệ, ĐHQGHN. (2023). Giáo trình Học máy và Mô hình trí tuệ nhân tạo (Bản nội bộ giảng dạy).
5. Tạp chí Khoa học và Công nghệ – Đại học Cần Thơ. (2021). Ứng dụng học sâu trong nhận dạng hình ảnh y tế. Tập 57, Số 2, trang 45–56.
6. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv preprint arXiv:1704.04861.
7. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4510–4520.
8. Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1251–1258.
9. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In 2009 IEEE Conference on Computer Vision and Pattern Recognition (pp. 248-255). IEEE.
10. Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. Journal of Big Data, 6(1), 60. https://doi.org/10.1186/s40537-019-0197-0