|  |
| --- |
| **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** |

# 

# 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.

# NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN

*TPHCM, Ngày …… tháng …… năm 2024*

**Giáo viên hướng dẫn**

(Ký tên)

# MỤC LỤC

[LỜI CẢM ƠN 3](#_Toc195463390)

[LỜI MỞ ĐẦU 4](#_Toc195463391)

[NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN 5](#_Toc195463392)

[MỤC LỤC 6](#_Toc195463393)

[CHƯƠNG 1: 8](#_Toc195463394)

[GIỚI THIỆU CƠ QUAN THỰC TẬP VÀ ĐỀ TÀI 8](#_Toc195463395)

[1.1 Giới thiệu đề tài 8](#_Toc195463396)

[1.2 Lý do chọn đề tài 8](#_Toc195463397)

[1.3 Phương pháp và mục tiêu 9](#_Toc195463398)

[1.3.1 Phương pháp 9](#_Toc195463399)

[1.3.2 Mục tiêu 10](#_Toc195463400)

[CHƯƠNG 2: 11](#_Toc195463401)

[CƠ SỞ LÝ THUYẾT 11](#_Toc195463402)

[2.1 MobileNetV2 11](#_Toc195463403)

[2.1.1 Inverted Residuals (Đảo ngược dư thừa): 12](#_Toc195463404)

[2.1.2 Depthwise Separable Convolutions: 12](#_Toc195463405)

[2.1.3 Hiệu suất cao và nhẹ: 12](#_Toc195463406)

[2.1.4 Chuyển giao học (Transfer Learning): 12](#_Toc195463407)

[2.1.5 Ứng dụng trên nhiều nền tảng: 13](#_Toc195463408)

[CHƯƠNG 3: 14](#_Toc195463409)

[MÔ HÌNH THỰC NGHIỆM 14](#_Toc195463410)

[3.1. Tổng quan dữ liệu 14](#_Toc195463411)

[3.2. Xử lý dữ liệu 15](#_Toc195463412)

[3.2.1. Tiền xử lý hình ảnh 15](#_Toc195463413)

[3.2.2. Tăng cường dữ liệu (Data Augmentation) 15](#_Toc195463414)

[3.2.3. Phân chia tập dữ liệu 15](#_Toc195463415)

[3.3. Huấn luyện mô hình 16](#_Toc195463416)

[3.3.1. Kiến trúc mô hình 16](#_Toc195463417)

[3.3.2. Tối ưu mô hình 17](#_Toc195463418)

[3.3.3. Môi trường thực nghiệm 17](#_Toc195463419)

[3.4. Đánh giá mô hình 17](#_Toc195463420)

[3.4.1 Quá trình huấn luyện 18](#_Toc195463421)

[3.4.2 Kết quả trên tập test 18](#_Toc195463422)

[3.4.3 Ma trận nhầm lẫn (Confusion Matrix) 19](#_Toc195463423)

[CHƯƠNG 4: 20](#_Toc195463424)

[KẾT LUẬN 20](#_Toc195463425)

[DANH MỤC TÀI LIỆU THAM KHẢO 21](#_Toc195463426)

# CHƯƠNG 1:

# GIỚI THIỆU CƠ QUAN THỰC TẬP VÀ 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 6 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 MobileNetV1**

MobileNet v1 là một kiến trúc mạng nơ-ron tích chập (CNN) gọn nhẹ, được thiết kế đặc biệt cho các ứng dụng thị giác máy tính trên thiết bị di động và nhúng. Kỹ thuật này giúp giảm đáng kể số lượng tham số và chi phí tính toán so với các lớp tích chập tiêu chuẩn, trong khi vẫn duy trì được độ chính xác tương đương.

**2.2.1 Kiến trúc cốt lõi của MobileNetV1:**

MobileNetV1 đạt được hiệu quả bằng cách sử dụng một kỹ thuật gọi là tích chập tách rời theo chiều sâu (depthwise separable convolution). Kỹ thuật này thay thế các lớp tích chập tiêu chuẩn bằng hai lớp nhỏ hơn:

1. Tích chập theo chiều sâu (Depthwise Convolution): Áp dụng một bộ lọc tích chập riêng biệt cho từng kênh đầu vào. Điều này giúp học các đặc trưng không gian trong mỗi kênh một cách hiệu quả.
2. Tích chập theo điểm (Pointwise Convolution hay 1x1 Convolution): Sử dụng các bộ lọc 1x1 để kết hợp thông tin giữa các kênh đầu ra từ lớp tích chập theo chiều sâu. Điều này giúp học các mối tương quan giữa các đặc trưng trên các kênh khác nhau.

**Ưu điểm của tích chập tách rời:**

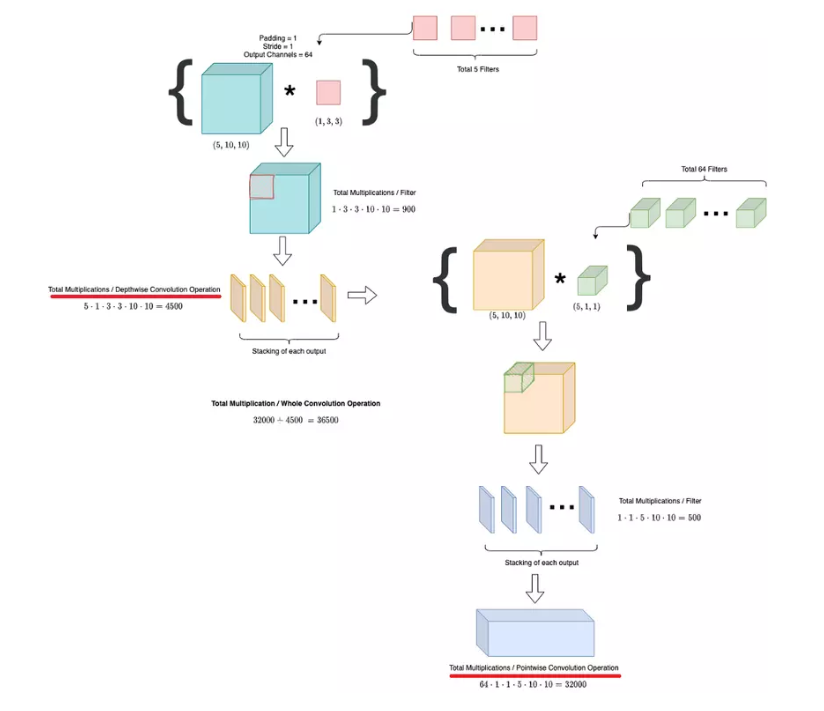
* Giảm số lượng tham số: So với tích chập tiêu chuẩn, tích chập tách rời sử dụng ít tham số hơn đáng kể để thực hiện một phép biến đổi tương đương.
* Giảm chi phí tính toán: Ít tham số hơn đồng nghĩa với việc cần ít phép tính toán hơn, dẫn đến tốc độ suy luận nhanh hơn và tiêu thụ ít năng lượng hơn.

**2.2.2 Cấu trúc chung của MobileNetV1**:

Mạng MobileNetV1 thường bao gồm một lớp tích chập tiêu chuẩn ban đầu, sau đó là một chuỗi các khối tích chập tách rời, cuối cùng là các lớp gộp trung bình toàn cục (global average pooling) và một lớp kết nối đầy đủ để thực hiện phân loại.

Các yếu tố ảnh hưởng đến kích thước và hiệu suất của MobileNetV1:

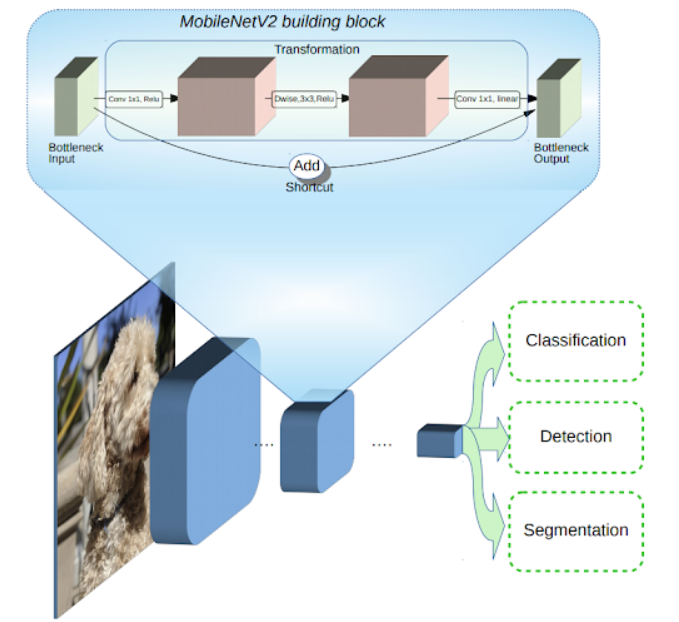
* Hệ số độ rộng (Width Multiplier - α): Một tham số toàn cục có thể được sử dụng để làm mỏng mạng một cách đồng đều bằng cách nhân số lượng kênh trong mỗi lớp với hệ số này (0 < α ≤ 1). Điều này cho phép tạo ra các mô hình nhỏ hơn và nhanh hơn với sự đánh đổi về độ chính xác.
* Độ phân giải đầu vào (Input Resolution): Kích thước của ảnh đầu vào cũng ảnh hưởng đến hiệu suất và kích thước của mô hình.



Hình ảnh này cho thấy rõ ràng rằng việc sử dụng tích chập tách rời theo chiều sâu giúp giảm đáng kể số lượng phép nhân cần thiết (từ hàng triệu xuống hàng trăm nghìn trong ví dụ này) để thực hiện một phép biến đổi tương tự (tạo ra đầu ra có cùng kích thước). Đây là lý do chính tại sao MobileNet và các kiến trúc tương tự sử dụng tích chập tách rời để đạt được hiệu quả tính toán cao hơn, phù hợp cho các ứng dụng trên thiết bị di động và nhúng với nguồn lực hạn chế.

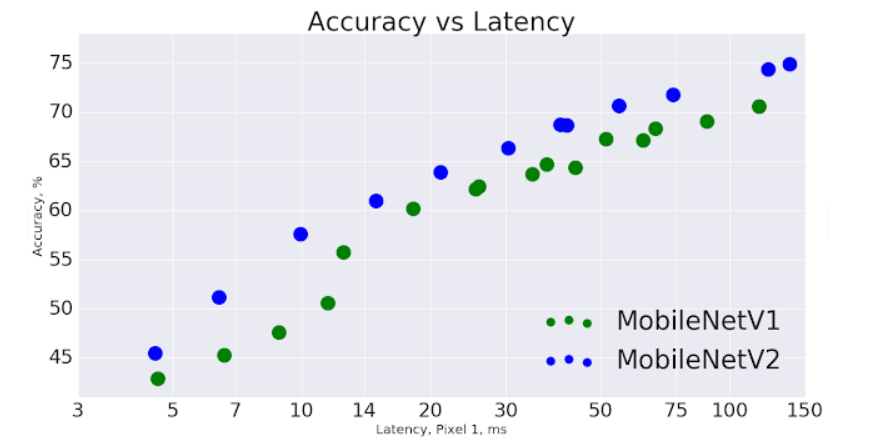
## 2.3 MobileNetV2

MobileNetV2 là phiên bản cải tiến của MobileNet, với các thay đổi quan trọng trong cấu trúc giúp tăng cường hiệu suất và khả năng phân loại mà vẫn giữ được độ nhẹ và nhanh.

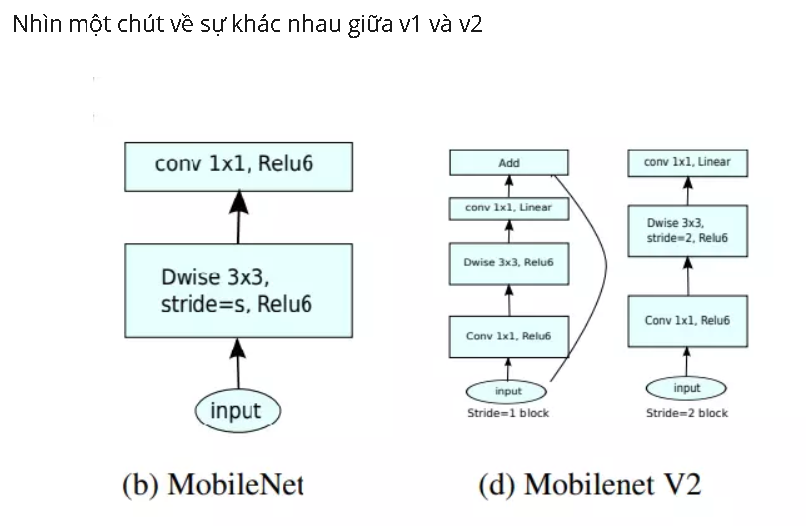


Tổng quan về Kiến trúc MobileNetV2. Các khối màu xanh biểu diễn các khối xây dựng tích chập tổng hợp như được hiển thị ở trên.

Các mô hình MobileNetV2 nhanh hơn với cùng độ chính xác trên toàn bộ phổ độ trễ. Đặc biệt, các mô hình mới sử dụng ít hơn 2 lần các hoạt động, cần ít hơn 30% các tham số và nhanh hơn khoảng 30-40% trên điện thoại Google Pixel so với các mô hình MobileNetV1, tất cả trong khi vẫn đạt được độ chính xác cao hơn.



Bảng so sánh độ chính xác và độ trễ giữa mobileNetV1 và mobileNetV2



Các đặc điểm nổi bật của MobileNetV2:

### 2.3.1 Inverted Residuals (Đảo ngược dư thừa):

MobileNetV2 giới thiệu một khái niệm mới gọi là inverted residuals (dư thừa đảo ngược), giúp giảm số lượng tham số mà vẫn duy trì hiệu suất cao. Điều này giúp mô hình hoạt động nhanh hơn và sử dụng ít tài nguyên hơn, rất phù hợp với các thiết bị di động hoặc hệ thống không có nhiều bộ nhớ hoặc sức mạnh tính toán.

Thay vì sử dụng các phép toán tích chập truyền thống, MobileNetV2 sử dụng một loại depthwise separable convolution kết hợp với một lớp bottleneck để giảm số lượng phép toán cần thiết.

### 2.3.2 Depthwise Separable Convolutions:

MobileNetV2 sử dụng một kỹ thuật gọi là depthwise separable convolution, nơi các phép toán tích chập được phân tách thành hai bước: bước đầu tiên là tích chập riêng biệt cho mỗi kênh đầu vào, và bước thứ hai là một phép tích chập 1x1 để kết hợp các kết quả. Kỹ thuật này giúp giảm số lượng tham số và tính toán, làm cho mô hình nhẹ hơn và nhanh hơn so với các mô hình CNN truyền thống.

### 2.3.3 Hiệu suất cao và nhẹ:

Với thiết kế nhẹ nhàng và tối ưu hóa tài nguyên, MobileNetV2 cho phép đạt được hiệu suất phân loại cao mà không cần quá nhiều tài nguyên tính toán. Điều này khiến MobileNetV2 trở thành một lựa chọn lý tưởng cho các ứng dụng di động, các thiết bị nhúng hoặc các hệ thống không có phần cứng mạnh mẽ như các máy chủ với GPU.

### 2.3.4 Chuyển giao học (Transfer Learning):

MobileNetV2 có thể được sử dụng với kỹ thuật transfer learning, tức là sử dụng các trọng số đã được huấn luyện trước trên các bộ dữ liệu lớn và sau đó tinh chỉnh mô hình cho các tác vụ cụ thể, chẳng hạn như phân loại hình ảnh bệnh lý mắt. Điều này giúp tăng cường khả năng phân loại mà không cần huấn luyện từ đầu, tiết kiệm thời gian và tài nguyên.

### 2.3.5 Ứng dụng trên nhiều nền tảng:

MobileNetV2 được thiết kế để chạy trên nhiều nền tảng, bao gồm các thiết bị di động (smartphone, tablet), máy tính cá nhân và các hệ thống tính toán đám mây. Điều này giúp triển khai các mô hình học sâu vào thực tế một cách dễ dàng và hiệu quả, đặc biệt khi các thiết bị này có tài nguyên hạn chế.

**2.4 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.

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.

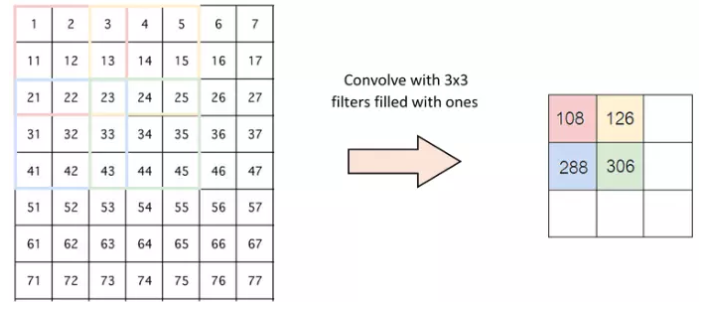
**Các lớp cơ bản của thuật toán CNN**

**2.4.1 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.

**2.4.2 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 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

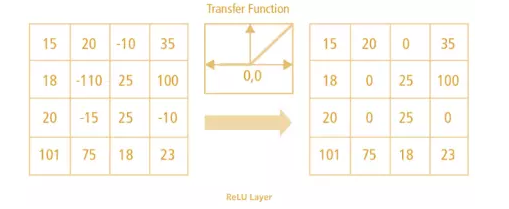
**2.4.3 Padding (Đường viền)**

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

**2.4.4 Hàm phi tuyến - ReLU**

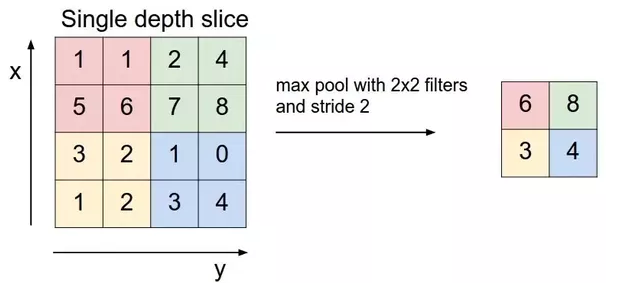
* ReLU viết tắt của Rectified Linear Unit, là 1 hàm phi tuyến. Với đầu ra là: ƒ (x) = max (0, x).
* Tại sao ReLU lại quan trọng: ReLU giới thiệu tính phi tuyến trong ConvNet. Vì dữ liệu trong thế giới mà chúng ta tìm hiểu là các giá trị tuyến tính không âm.



Hình 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 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.

# 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 6 loại bệnh về mắt phổ biến:

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

Tổng cộng, bộ dữ liệu có khoảng 6000 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 pie chart with different colored circles

AI-generated content may be incorrect.

## 3.2. 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. Sau đó, ảnh được chuẩn hóa bằng cách chia giá trị điểm ảnh cho 255 để đưa về khoảng [0, 1], giúp mô hình hội tụ nhanh hơn khi huấn luyện.

### 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, 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% (4800 hình ảnh)

Tập kiểm định (validation): 20% (1200 hình ảnh)

A blue and orange rectangular graph

AI-generated content may be incorrect.

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

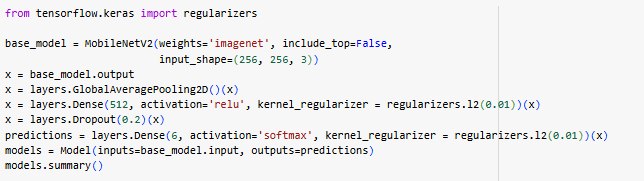
Mô hình được lựa chọn trong dự án này là MobileNetV2, một kiến trúc mạng nơ-ron tích chập nhẹ, được thiết kế tối ưu cho các thiết bị di động hoặc hệ thống có tài nguyên hạn chế. MobileNetV2 có đặc điểm sử dụng khối “inverted residuals” kết hợp với depthwise separable convolutions, giúp giảm số lượng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

### 3.3.1. Kiến trúc mô hình

Trong thực nghiệm này, mô hình MobileNetV2 được sử dụng như một mạng nền (base model) với các trọng số pretrained 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 6 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 6 node tương ứng với 6 lớp bệnh, sử dụng hàm kích hoạt softmax để đưa ra xác suất cho từng lớp.



### 3.3.2. Tối ưu 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). Đây là kỹ thuật transfer learning, thường dùng khi dữ liệu huấn luyện bị hạn chế và muốn tận dụng kiến thức từ mô hình đã học trước đó.

Tham số huấn luyện gồm:

* Optimizer: Adam
* Learning rate: 0.0001
* Batch size: 32
* Epochs: 20
* Loss function: categorical crossentropy

Trong quá trình huấn luyện, mô hình sử dụng thêm hai 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.

### 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.

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, 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 67%, trong đó các lớp như Cataract, Diabetic Retinopathy và Healthy đạt độ chính xác cao, trong khi các lớp như Macular Scar, Myopia và Glaucoma thấp hơn do đặc trưng hình ảnh kém rõ ràng hơn.

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

A graph of different types of diseases

AI-generated content may be incorrect.

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

# CHƯƠNG 4:

# KẾT LUẬN

# DANH MỤC TÀI LIỆU THAM KHẢO