

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

KHOA CÔNG NGHỆ THÔNG TIN



MAI NGỌC TÚ – HUỲNH NHỰT QUANG

PHÂN LOẠI ẢNH Y KHOA DỰA VÀO ĐỘ BẤT THƯỜNG ĐẬM ĐỘ

**KHÓA LUẬN TỐT NGHIỆP CỬ NHÂN
CÔNG NGHỆ THÔNG TIN**

CHƯƠNG TRÌNH CHÍNH QUY

TP.Hồ Chí Minh, tháng 07/2022

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

KHOA CÔNG NGHỆ THÔNG TIN



MAI NGỌC TÚ – HUỲNH NHỰT QUANG

PHÂN LOẠI ẢNH Y KHOA DỰA VÀO ĐỘ BẤT THƯỜNG ĐẬM ĐỘ

KHÓA LUẬN TỐT NGHIỆP CỬ NHÂN

CÔNG NGHỆ THÔNG TIN

CHƯƠNG TRÌNH CHÍNH QUY

GIÁO VIÊN HƯỚNG DẪN

PGS.TS Lê Hoàng Thái

TP.Hồ Chí Minh, tháng 07/2022

Lời cảm ơn

Đầu tiên, chúng tôi xin gửi lời cảm ơn sâu sắc và chân thành nhất đối với PGS.TS Lê Hoàng Thái, thầy đã trực tiếp hướng dẫn, phân tích cũng như hỗ trợ nhóm tận tâm trong suốt thời gian và quá trình thực hiện khóa luận. Với sự tận tâm và hỗ trợ của thầy, chúng tôi mới có thể hoàn thành khóa luận này một cách trọn vẹn nhất có thể.

Bên cạnh ấy, chúng tôi chân thành cảm ơn Khoa Công Nghệ Thông Tin, Trường Đại học Khoa học Tự Nhiên, Đại học Quốc gia TP.Hồ Chí Minh đã tạo ra môi trường học tập, nguồn kiến thức bổ ích và tạo điều kiện thuận lợi để chúng tôi trao dồi bản thân, kiến thức chuyên ngành và thực hiện khóa luận trong suốt 4 năm.

Chúng tôi cũng xin chân thành cảm ơn tất cả quý thầy cô thuộc trường Đại học Khoa học Tự nhiên nói chung và các quý thầy cô Khoa Công Nghệ Thông Tin nói riêng. Tất cả kiến thức cả nền tảng và chuyên ngành, cùng những tư duy sáng tạo mà các quý thầy cô đã tận tâm giảng dạy và truyền đạt cho chúng tôi để không chỉ phục vụ cho việc thực hiện khóa luận này, mà còn là cả hành trang to lớn được trang bị cho công việc tương lai.

Và cuối cùng, sẽ là một thiếu sót to lớn nếu không nhắc đến tất cả người thân và bạn bè, những người đã và đang bên cạnh chúng tôi, luôn cổ vũ, giúp đỡ và động viên chúng tôi trong cả quá trình học tập cũng như thời gian thực hiện khóa luận. Chúng tôi xin chân thành cảm ơn tất cả mọi người.

Trải qua thời gian tìm hiểu, nghiên cứu và thực hiện khóa luận với tất cả sự cố gắng và phạm vi khả năng của bản thân chúng tôi. Tuy nhiên, chắc chắn sẽ tồn đọng những thiếu sót cũng như sơ sót trong khóa luận, kính mong quý thầy cô sẽ sẵn sàng chia sẻ và đóng góp tận tình để khóa luận trở nên trọn vẹn hơn.

Chúng tôi xin chân thành cảm ơn và biết ơn mong được sự chỉ bảo của quý thầy cô để khóa luận hoàn thành một cách tốt đẹp.

Thành phố Hồ Chí Minh, tháng 02 năm 2022

Nhóm sinh viên thực hiện

Mai Ngọc Tú – Huỳnh Nhật Quang

Đề cương chi tiết

Tên đề tài: Phân loại ảnh y khoa dựa vào độ bất thường đậm độ
Giáo viên hướng dẫn: Lê Hoàng Thái
Thời gian thực hiện: 01/2022 – 07/2022 (6 tháng)
Sinh viên thực hiện: Mai Ngọc Tú (MSSV: 18120253) Huỳnh Nhật Quang (MSSV: 18120228)
Loại đề tài: Nghiên cứu
Nội dung đề tài: <p>Đề tài là một dạng nghiên cứu lí thuyết. Mục đích chính là tìm ra phương pháp học máy đạt hiệu quả tốt nhất trong bài toán phân loại ảnh X-quang phổi của bệnh nhân. Chúng tôi dùng ba mô hình học sâu được đề xuất để huấn luyện dữ liệu. Cuối cùng dùng phương pháp Ensemble kết hợp kết quả ba mô hình trên lại với nhau, từ đó đưa ra dự đoán xem phổi của bệnh là bình thường (normal) hay bất thường (abnormal). Đây có thể không là một đề tài quá mới lạ, nhưng ý nghĩa thực tiễn của nó lại vô cùng to lớn.</p>
Kế hoạch thực hiện: <p>Ban đầu, tất cả thành viên phải hình dung được vấn đề mình đang tìm hiểu, tuy liên quan đến vấn đề y khoa nhưng về bản chất đây vẫn là một bài toán phân loại ảnh dựa vào thuộc tính là đậm độ. Trong quá trình làm việc, phải thường xuyên trao đổi, thảo luận cũng như chia sẻ cho nhau những kiến thức và vấn đề liên quan. Quá trình thực hiện khóa luận có thể chia thành 2 phần cụ thể như sau:</p> <ul style="list-style-type: none">• Tìm hiểu đề tài và tìm hiểu các phương pháp khả thi cho đề tài.• Thống nhất và triển khai phương pháp dùng 3 mô hình và kết hợp (Ensemble) kết quả của chúng lại với. <p>Thời gian thực hiện cụ thể:</p>

- 01/01/2022 – 31/01/2022:
 - Tìm hiểu bài toán nhận dạng ảnh
 - Tìm hiểu các mô hình đạt hiệu quả cao trong nhận dạng ảnh
- 01/02/2022 – 28/02/2022:
 - Tìm kiếm phương pháp phù hợp cho đề tài
- 01/03/2022 – 31/05/2022:
 - Tìm kiếm ba mô hình *Deep Learning* đạt hiệu quả cao với bộ dữ liệu ảnh Shenzhen
 - Kết hợp kết quả của ba mô hình *Deep Learning* trên
 - Phân tích và đánh giá kết quả
- 01/06/2022 – 23/06/2022:
 - Viết cuốn luận
 - Làm slide thuyết trình

Xác nhận của GVHD



Lê Hoàng Thái

**TP.Hồ Chí Minh, ngày 7 tháng 6
năm 2022**

**NHÓM SINH VIÊN THỰC HIỆN
KHÓA LUẬN
(Ký và ghi rõ họ tên)**



Mai Ngọc Tú



Huỳnh Nhựt Quang

Mục lục

Lời cảm ơn	3
Đề cương chi tiết	5
Mục lục.....	7
DANH MỤC HÌNH ẢNH	9
DANH MỤC BẢNG BIỂU	11
Chương 1 Giới thiệu đề tài.....	12
1.1 Lý do nghiên cứu	12
1.2 Mục đích nghiên cứu	13
1.3 Cách tiếp cận.....	13
1.4 Đóng góp của đề tài	15
Chương 2 Phát biểu bài toán	16
2.1 Tổng quan về bài toán phân lớp.....	16
2.1.1 Khái niệm về bài toán phân lớp.....	16
2.1.2 Bài toán phân lớp ảnh.....	17
2.1.3 Bài toán phân loại ảnh X-quang phổi dựa vào độ bất thường đậm độ.....	18
2.2 Các công trình nghiên cứu trong cùng lĩnh vực trong và ngoài nước	21
2.2.1 Trong nước	21
2.2.2 Ngoài nước	21
Chương 3 Cơ sở lý thuyết	23
3.1 Neural Network.....	23
3.2 VGG-16	26
3.3 ResNet50.....	30

3.4 DenseNet121.....	34
3.5 Ensemble.....	37
3.6 Ảnh X-quang.....	39
3.7 Transfer Learning	40
3.8 Fine tuning	41
3.9 Data Augmentation	42
Chương 4 Phương pháp đề nghị.....	45
4.1 Ý tưởng mô hình.....	45
4.2 Workflow	45
4.3 Kết luận và tổng quan.....	48
Chương 5 Thực nghiệm và kết quả.....	49
5.1 Về tập dữ liệu.....	49
Chương 6 Kết luận và tương lai phát triển.....	53

DANH MỤC HÌNH ẢNH

Hình 1-1: Hình minh họa một ảnh trong dữ liệu ShenZhen	14
Hình 2-1: Minh họa bài toán phân lớp	16
Hình 2-2: Minh họa diagram bài toán phân lớp ảnh	17
Hình 2-3: Ảnh chụp CT não (nguồn: Scientists Say: CT scan Science News for Students).....	19
Hình 2-4: Ảnh X-quang tay (nguồn: http://pctu.edu.vn).....	19
Hình 2-5: Hình minh họa diagram bài toán phân lớp ảnh X-quang phổi	20
Hình 3-1: Ảnh minh họa Neural Network (nguồn: IBM)	24
Hình 3-2: Hình minh họa các loại mạng Nơ-ron hiện nay (nguồn: trên ảnh).....	25
Hình 3-3: Ảnh minh họa kiến trúc CNN thông thường (nguồn: Convolutional Neural Network: An Overview).....	26
Hình 3-4: Hình minh họa kiến trúc VGG (Nguồn: VGG16 - Convolutional Network for Classification and Detection)	27
Hình 3-5: Hình minh họa kiến trúc VGG-16 (nguồn: trên ảnh)	28
Hình 3-6: Hình minh họa Skip Connection.....	31
Hình 3-7: Hình minh họa kiến trúc VGG-19 và ResNet34.....	32
Hình 3-8: Hình minh họa về ResNet50 (nguồn: Keras Implementation of ResNet-50 (Residual Networks) Architecture from Scratch - MLK - Machine Learning Knowledge)	33
Hình 3-9: Hình minh họa kiến trúc DenseNet và ResNet.....	35
Hình 3-10: Hình minh họa cho DenseBlock và Transition Layers(nguồn: Densely Connected Convolutional Networks).....	35
Hình 3-11: Hình minh họa toàn bộ kiến trúc DenseNet (nguồn: Pablo R).....	36
Hình 3-12: Hình minh họa phương pháp Bagging và Boosting	38
Hình 3-13: Hình minh họa phương pháp stacking	38
Hình 3-14: Hình minh họa ảnh X-quang thuộc Shenzhen Dataset	39
Hình 3-15: Hình minh họa Data Augmentation.....	42
Hình 3-16: Kỹ thuật Plip	43

Hình 3-17: Kỹ thuật Crop.....	43
Hình 3-18: Kỹ thuật Rotation.....	44
Hình 3-19: Kỹ thuật Scale	44
Hình 4-1: Workflow phương pháp Ensemble đề xuất. $Acc = Accuracy$; $w(i)$ là trọng số của mô hình thứ i , $p_j(i)$ là điểm xác suất cho mẫu thứ j của mô hình thứ i , ens_j là điểm xác suất ensemble của mẫu thứ j và hàm $argmax$ trả về giá trị lớn nhất của m	47
Hình 5-1: Ảnh X-quang phổi bình thường trong bộ dữ liệu Shenzhen.....	51
Hình 5-2: Ảnh minh họa phổi bất thường trong bộ dữ liệu Shenzhen.....	51

DANH MỤC BẢNG BIỂU

Bảng 3-1: Bảng so sánh cấu trúc các kiến trúc VGG.....	29
Bảng 3-2: Bảng tham số của các kiến trúc DenseNet (nguồn : Densely Connected Convolutional Networks)	36
Bảng 5-1: Chi tiết bên trong file excel đi kèm	50

Chương 1 Giới thiệu đề tài

1.1 Lý do nghiên cứu

Trong thời đại hiện nay, khi mà sự phát triển vượt trội của nền văn minh loài người, lĩnh vực công nghệ thông tin có những bước tiến vô cùng to lớn. Chúng ta có thể dễ dàng bắt gặp được dấu ấn của công nghệ thông tin trong hầu hết mọi lĩnh vực, giúp ích cho nhu cầu cuộc sống, nâng cao mức sống và chất lượng sống của chúng ta. Một điều cũng dễ nhận thấy sự tiến bộ vượt trội của công nghệ thông tin là khối lượng ảnh số cũng ngày một trở nên dày đặc và không ngừng phát triển trong mọi lĩnh vực. Với số lượng ảnh to lớn như vậy, phân loại và phân tích chúng dần trở thành một phương tiện và là một nhu cầu tất yếu trong cuộc sống để phục vụ cho các tác vụ nghiên cứu và giải quyết vấn đề trong cuộc sống. Đi cùng là sự phát triển của công nghệ chụp ảnh, thiết bị đồ họa và những thiết bị lưu trữ cũng góp phần tích cực thúc đẩy sự phát triển của lĩnh vực nghiên cứu phân tích, xử lý ảnh số và cả những ứng dụng của nó vào thực tiễn.

Với tất cả những phát triển đã nêu trên, ngành Thị giác máy tính (Computer Vision) và Khoa học máy tính (Computer Science) cũng xuất hiện như một điều tất yếu, và đã đóng góp to lớn vào nền công nghệ của nhân loại. Tiêu biểu có thể nói đến những ứng dụng tiêu biểu như nhận dạng (nhận dạng khuôn mặt, vân tay, giọng nói, ...), phát triển các thiết bị tự hành, tự động hóa, ... Và được các ông lớn công nghệ trên thế giới chú ý và đầu tư phát triển như Facebook, Microsoft, Google, ... Nhờ vào công nghệ và kiến thức từ 2 ngành trên ta cũng có thể thấy những ứng dụng phục vụ đời sống con người xuất hiện như các máy nhận dạng người, vật, hoặc nhận dạng một thứ gì đó từ ảnh số, đảm bảo cho an ninh và quản lý, và không thể thiếu là nhận dạng ảnh y khoa góp phần vào dự đoán cũng như điều trị trong y tế.

Ngoài các phương pháp thăm khám trực tiếp các cơ quan trong cơ thể, ảnh y khoa dường như là một bản ghi được xem như là đơn giản và dễ tiếp cận nhất trong việc khám và điều trị bệnh. Hình ảnh y khoa bao gồm các hình ảnh chụp X-quang, CT- cắt lớp , MRI cộng hưởng từ , PET, ảnh nội soi và ảnh siêu âm, nhằm tạo hình

ảnh trực quan về các cơ quan bên trong cơ thể mà mắt thường không nhìn thấy, từ đó có thể giúp các bác sĩ và chuyên gia có chuyên môn phân tích lâm sàng và đưa ra chẩn đoán cũng như có thể là can thiệp y tế. Có thể thấy việc chẩn đoán ảnh y khoa được xem là việc làm tuyến đầu của trong quá trình khám bệnh, nhằm đưa ra lộ trình điều trị chính xác và không lãng phí thời gian. Hiện nay, hầu hết các bệnh viện đều có một bộ phận chẩn đoán cùng với các trang thiết bị hỗ trợ và chẩn đoán ảnh y khoa cũng không ngoại lệ.

Ở trong luận văn này, chúng tôi sẽ nghiên cứu và tìm hiểu về bài toán nhận dạng ảnh y khoa từ đó phát hiện ra những sự bất thường góp phần vào việc chẩn đoán của y tế. Nếu chúng ta có một mô hình chẩn đoán ảnh y khoa với độ chính xác cao, cũng có nghĩa chúng ta sẽ hạn chế thời gian xét nghiệm và rút ngắn tiến trình điều trị của bác sĩ với bệnh nhân. Bằng những phương pháp học máy như kNN, SVM, hồi quy tuyến tính, ... và những mô hình Deep Learning hiện đại, sẽ giúp những chẩn đoán ảnh y khoa trở nên chính xác hơn và nhanh chóng đem lại hiệu quả. Vì vậy, chúng tôi xin đưa đề tài nghiên cứu và phát triển đề tài “ *Phân loại ảnh y khoa dựa vào độ bất thường đậm độ* “.

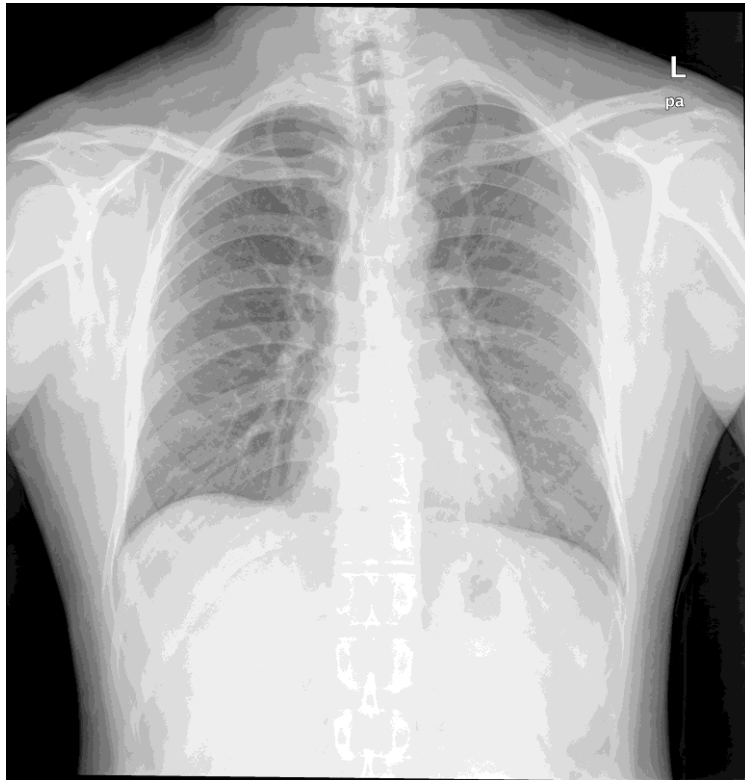
1.2 Mục đích nghiên cứu

Trong phạm vi hiểu biết và khả năng của chúng tôi, ở khóa luận này, mục tiêu nghiên cứu tìm ra phương pháp học máy đạt hiệu quả tốt trong bài toán phân loại ảnh X-quang phổi của bệnh nhân với bộ dữ liệu huấn luyện của Shenzhen.

1.3 Cách tiếp cận

Như đã được đề cập bên trên, ban đầu nhóm chúng tôi sẽ tìm hiểu về ảnh X-quang phổi. Sau đó thay vì dùng chỉ một thuật toán hay một mô hình cụ thể để huấn luyện và đưa ra kết quả, nhóm chúng tôi sẽ kết hợp 3 mô hình Deep Learning tốt nhất phù hợp với bộ dữ liệu đã qua so sánh kết quả của từng mô hình, từ đó dựa vào accuracy để quyết định.

Tập dữ liệu được sử dụng để huấn luyện là tập dữ liệu ảnh X-quang phổi ShenZhen bao gồm 662 ảnh X-quang phổi của bệnh nhân với các tình trạng bệnh khác nhau. Ảnh minh họa bên dưới:



Hình 1-1: Hình minh họa một ảnh trong dữ liệu ShenZhen

Đây là tập dữ liệu ảnh do bệnh viện Shenzhen công bố công khai, cho nên chúng tôi cam đoan về tính pháp lý. Và đây cũng là ảnh của người thật, sẽ gồm nhiều bệnh, và chắc chắn không được tạo nên từ máy tính hay bất kỳ phần mềm nhân tạo nào cả. Vì vậy, ở phạm vi nghiên cứu, chúng tôi thấy tập dữ liệu này khá phù hợp để triển khai huấn luyện.

Mặc khác, ở phạm vi khả năng có thể, chúng tôi khó mà có thể mở rộng tập dữ liệu, vì được biết, ảnh y khoa là tài liệu cá nhân và cần được sự cho phép mới có thể sử dụng để phân tích và nghiên cứu. Nhưng với số lượng ảnh ở mức 600+, chúng tôi vẫn nghĩ sẽ không có trường hợp bị overfitting.

1.4 Đóng góp của đề tài

Với kết quả nghiên cứu này, nhóm chúng tôi mong muốn có thể tạo ra một mô hình học sâu có thể phân loại được ảnh X-quang phổi dựa vào độ bất thường đậm độ. Từ đó, rút ngắn thời gian nhận định, chẩn đoán của bác sĩ và chuyên gia, thúc đẩy tiến trình khám và điều trị của bệnh nhân. Đồng thời, cũng khẳng định một lần nữa tác dụng và ảnh hưởng tích cực của ngành Khoa học máy tính và Thị giác máy tính lên ngành y khoa. Cuối cùng, mô hình học sâu được cài đặt và triển khai trong luận văn này có thể được phát triển để trở nên hoàn thiện hơn trong lĩnh vực phân loại ảnh y khoa.

Chương 2 Phát biểu bài toán

2.1 Tổng quan về bài toán phân lớp

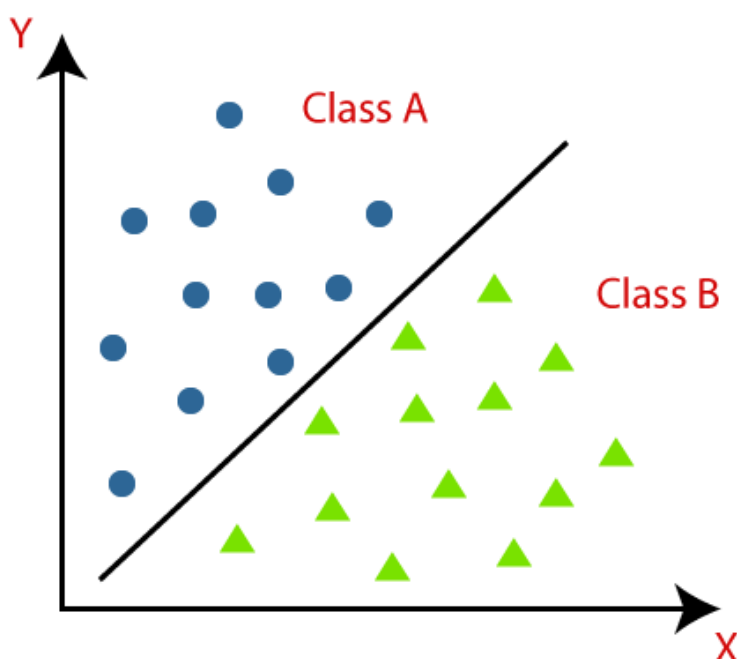
2.1.1 Khái niệm về bài toán phân lớp

Bên cạnh hai bài toán gom nhóm (Clustering) và hồi quy (Regression), bài toán phân lớp luôn đóng một vai trò quan trọng và phổ biến trong lĩnh vực Machine Learning.

Về cơ bản, bài toán phân lớp là quá trình phân lớp 1 nhóm đối tượng vào các lớp tương ứng (có thể là 1 hay nhiều lớp) cho trước thông qua một mô hình phân lớp.

Bài toán phân lớp là bài toán học có quan sát (Supervised learning), vì vậy chuẩn bị một bộ dữ liệu đã được gán nhãn là điều kiện tất yếu. Các bước để thực hiện việc phân lớp một bộ dữ liệu:

- Xây dựng mô hình được huấn luyện dựa trên bộ dữ liệu đã được gán nhãn.
- Dựa vào mô hình đã được huấn luyện ở trên, gán nhãn các đối tượng chưa được gán nhãn, quá trình này được gọi là quá trình phân lớp.



Hình 2-1: Minh họa bài toán phân lớp

Mục tiêu hướng đến của bài toán phân lớp là tìm ra được một mô hình hay một thuật toán có thể phân loại tốt dữ liệu và gán nhãn chúng chính xác. Thực tế, có khá nhiều loại bài toán phân lớp có thể kể đến như phân lớp nhị phân (Ví dụ: phân lớp mail có phải spam hay không), phân lớp đa lớp (VD: phân loại nhiều loại hoa), ...

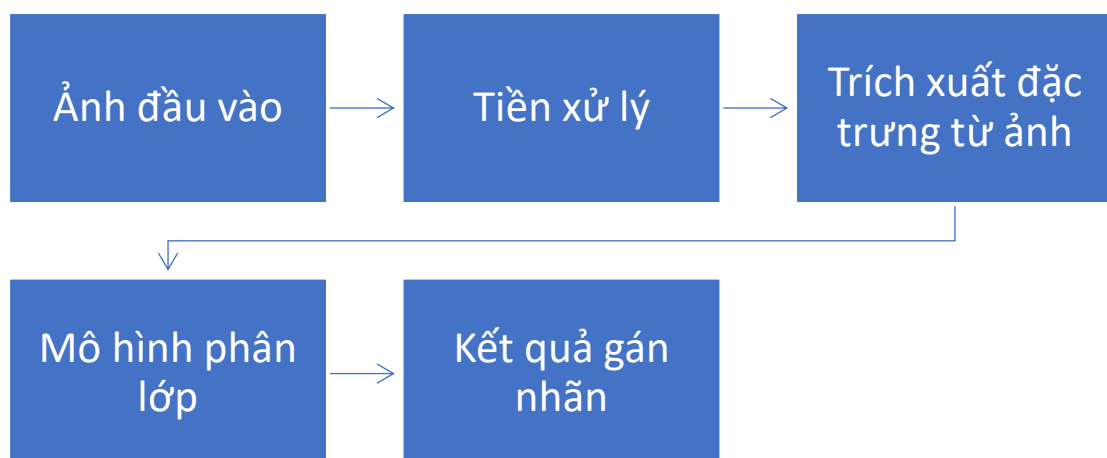
Phân lớp có thể dựa vào nhiều thuộc tính được cung cấp theo dạng table, dataframe, Và phân lớp ảnh cũng là một phần rất quan trọng trong bài toán phân lớp. Dựa vào ảnh, có thể trích xuất đặc trưng và gán nhãn.

2.1.2 Bài toán phân lớp ảnh

Bài toán phân lớp ảnh là một bài toán phổ biến trong lĩnh vực Computer Vision, góp phần giải quyết nhiều vấn đề liên quan đến cuộc sống thông qua ảnh số. Có thể kể đến một số bài toán phân lớp ảnh phổ biến như nhận dạng chữ viết tay MNIST, nhận dạng dấu vân tay, phân lớp động vật (chó hay mèo) thông qua ảnh, ...

Mục tiêu chính của bài toán phân lớp ảnh là từ một ảnh hay một bộ ảnh đầu vào (input) thông qua mô hình đã được xây dựng phân ra đầu ra (output) theo những nhãn cho trước. Các bước phân lớp ảnh được biểu diễn theo workflow bên dưới:

- Tiền xử lý: ở giai đoạn này, ta có thể resize cũng sử dụng một số kỹ thuật xử lý ảnh để làm lộ ra các đặc trưng cụ thể.

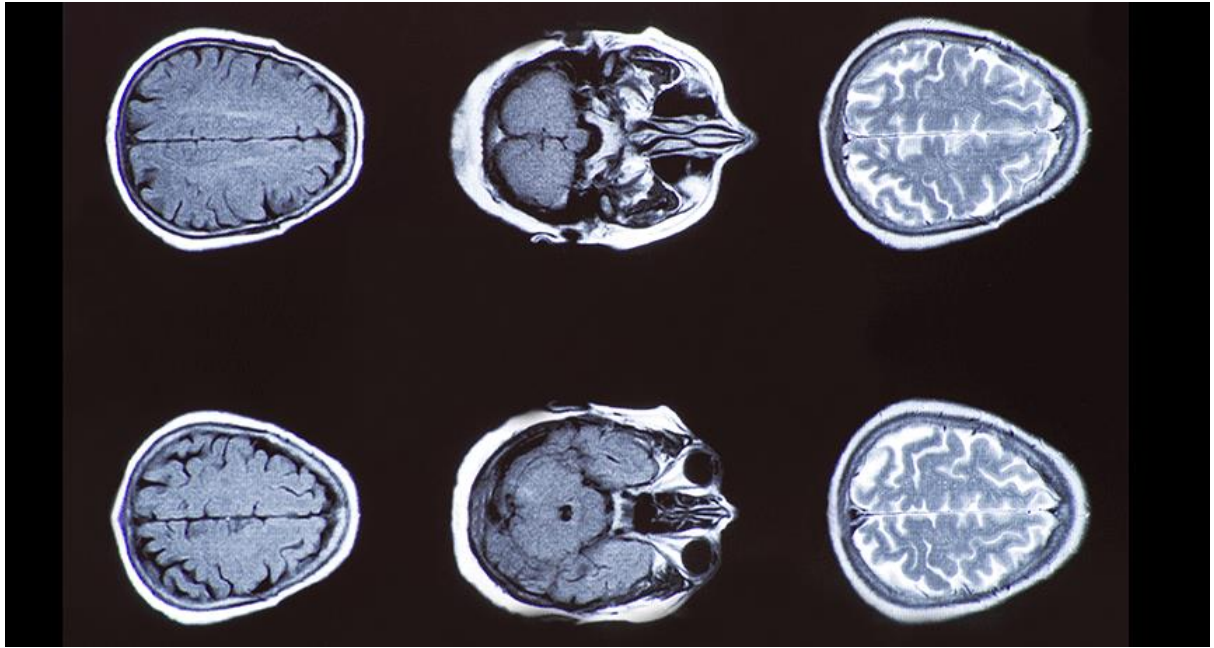


Hình 2-2: Minh họa diagram bài toán phân lớp ảnh

Bài toán phân lớp ảnh đã giải quyết nhiều vấn đề mà trước kia con người chưa thể dự đoán và phân loại được bằng mắt thường hay trực quan ảnh. Trong y tế, ảnh cũng đóng vai trò quan trọng trong chẩn đoán và điều trị. Việc phân tích và dự đoán được từ ảnh sẽ tạo ra bước tiến lớn trong lĩnh vực này. Ở đây, chúng tôi sẽ sử dụng ảnh chụp X-quang về phổi, dựa vào độ bất thường đậm độ của nó để phân lớp, bình thường (normal) hay bất thường (abnormal).

2.1.3 Bài toán phân loại ảnh X-quang phổi dựa vào độ bất thường đậm độ

Theo báo cáo của bộ y tế ngày 28 – 30/9/2020, hằng năm có đến hàng triệu người chết vì bệnh liên quan đến thuốc và môi trường. Các chuyên gia hô hấp cho biết, hầu hết các tác nhân trên đều gây ra các bệnh nguyên hiểm liên quan đến phổi, cũng chính là nguyên nhân gây tử vong đứng thứ 4 trên toàn thế giới. Trong bước đầu chẩn đoán về phổi ở hầu hết các nền y học trên thế giới và cả Việt Nam, ảnh X-quang và ảnh CT- cắt lớp luôn là lựa chọn ưu tiên hàng đầu. Ảnh X-quang sẽ cho ta duy nhất 1 góc nhìn tổng thể về phổi thông qua các tia X chiếu qua xương và mô của khoang ngực. Chi tiết hơn, ta có ảnh CT – cắt lớp , qua đó có thể mô phỏng lại chính xác chi tiết các bộ phận trong khoang ngực chúng ta. Có thể thấy, ảnh CT – cắt lớp sẽ cho chúng ta một cái nhìn dễ dàng và rõ ràng hơn về kích thước, vị trí, hình dạng và mật độ các nốt của phổi, điều mà ảnh X-quang chưa thể làm được chi tiết đến vậy. Tuy nhiên, khó khăn và rào cản lớn nhất ở đây là trang thiết bị để chụp khá tốn kém, vì vậy, cũng không phổ biến chụp CT – cắt lớp ở các bệnh viện nhỏ hoặc nông thôn, vùng có điều kiện khó khăn. Tuy không chi tiết như ảnh CT – cắt lớp, ảnh X-quang lại phổ biến hơn vì giá thành rẻ, trang thiết bị dễ dàng trang bị, tiết kiệm thời gian và phần nào là giảm bớt ảnh hưởng của tia X đối với cơ thể bệnh nhân trong quá trình lấy ảnh. Vậy, câu hỏi ở đây được đặt ra, làm thế nào từ ảnh X-quang thông thường bằng phương pháp học máy có thể phân lớp chính xác và hiệu quả tương đương như với ảnh CT – cắt lớp, và cụ thể ở đây là xác định phổi có bình thường hay bất thường.



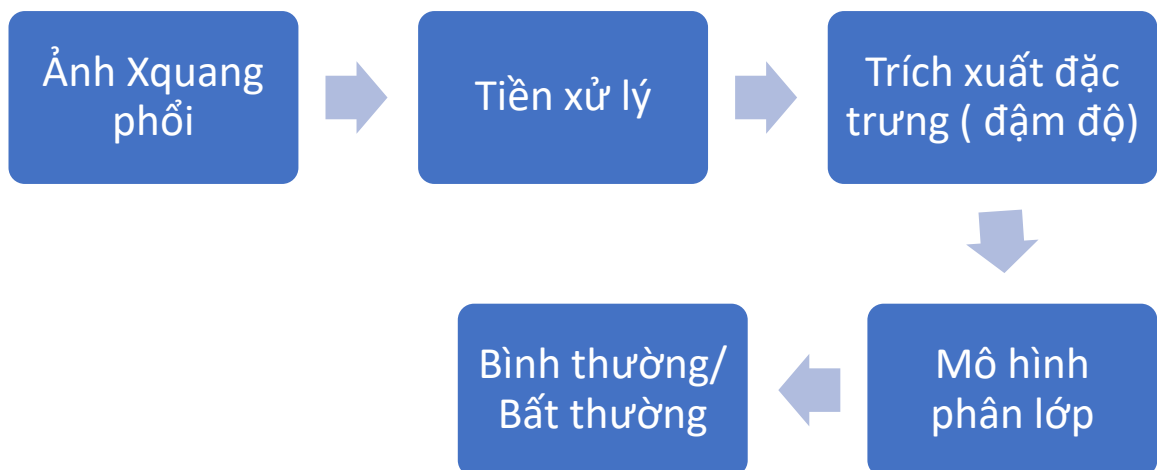
Hình 2-3: Ảnh chụp CT não (nguồn: [Scientists Say: CT scan | Science News for Students](#))



Hình 2-4: Ảnh X-quang tay (nguồn: <http://pctu.edu.vn>)

Bài toán phân loại ảnh X-quang dựa vào độ bất thường đậm độ trong thực tế lại có ý nghĩa vô cùng to lớn và hiệu quả rõ ràng trong quá trình chẩn đoán và phân loại bệnh của bệnh nhân. Nếu chúng ta có thể làm tốt bước phân loại này chỉ thông qua độ bất thường đậm độ của ảnh X-quang về phổi, ta sẽ rút ngắn giai đoạn chẩn đoán của bác sĩ chuyên gia và từ đó việc điều trị cũng trở nên nhanh chóng và hiệu quả đối với bệnh nhân. Thứ chúng ta cần để phân loại tốt ảnh X-quang này là một mô hình hay một phương pháp phân loại nào đó, dựa vào input đầu có thể phân lớp tốt chúng theo nhãn cho trước. Về cơ bản, có thể khái quát bài toán này là một bài toán

phân lớp ảnh nhị phân thông thường, chỉ đặc biệt ảnh đầu vào là ảnh Xquang và đầu ra là nhãn tương ứng (bình thường hoặc bất thường). Tương tự các bước thực hiện nhìn chung vẫn giống với bài toán phân lớp.



Hình 2-5: Hình minh họa diagram bài toán phân lớp ảnh X-quang phổi

Chúng tôi xin nói thêm một chút về nhãn đầu ra. Ở trong phạm vi nghiên cứu và phân tích của sinh viên, chúng tôi chỉ phân ra làm 2 lớp:

- Bình thường: Phổi ở trạng thái ổn định, hoặc các bất thường không đáng kể, không được xem là dấu hiệu của bệnh.
- Bất thường: Tất cả trạng thái khác đều được xem là bất thường, không phân biệt bệnh gì về phổi, và cũng không phân biệt phổi trái hay phổi phải hay cả hai phổi đều bất thường.

Trong vài năm gần đây, dịch bệnh Covid hoành hành, và phổi là cơ quan bị tác động nặng nề nhất đối với bệnh nhân. Vì vậy, việc tạo ra một mô hình chẩn đoán thông qua ảnh X-quang lại trở nên cần thiết và được nhiều tổ chức, cá nhân tham gia nghiên cứu sôi nổi và có nhiều đóng góp thiết thực cho nhân loại.

2.2 Các công trình nghiên cứu trong cùng lĩnh vực trong và ngoài nước

2.2.1 Trong nước

Ở Việt Nam, những năm gần đây, lĩnh vực Thị giác máy tính và Khoa học máy tính phát triển khá mạnh mẽ. Trong các trường đại học và sau đại học đều có những chuyên ngành về 2 ngành trên. Từ ấy, những bài báo, nghiên cứu cũng ra đời nhiều tạo nên sự sinh động và cơ hội phát triển cho lĩnh vực. Có thể nói đến một số nghiên cứu như sau:

- Luận văn Tiến sỹ của ông Vũ Mạnh Hùng: tác giả thực hiện giải quyết một lớp các bài toán nhận dạng chữ viết tay mà cụ thể là nhận dạng các ký tự đơn lẻ các chữ cái từ A đến Z với 2 mô hình mô hình học sâu (deep- 16 learning): MLP và RBM. Tác giả kết luận: Việc áp dụng mô hình RBM cho ra một kết quả tương đối tốt, với mô hình này kết quả thu được trên bộ dữ liệu chữ viết tay tuyển sinh là 99.69%, tốt hơn xấp xỉ khoảng 2% so với mô hình MLP (97,94%).

- Luận văn Tiến sỹ của ông Nguyễn Kiều Phát: tác giả thực hiện giải quyết bài toán phân lớp ảnh X-quang ở phổi với mạng CNN bằng cách chia phổi ra làm 2 phần, huấn luyện lần lượt với phổi trái, phải và tổng thể. Luận văn thu lại kết quả khả quan với độ chính xác xấp xỉ 92%.

- Gần đây, phân lớp ảnh cũng được tiếp cận rộng rãi với hầu hết người đam mê với lĩnh vực này qua các cuộc thi trên Kaggle về phân lớp ảnh và nhận dạng thông qua ảnh.

2.2.2 Ngoài nước

Trên thế giới, Thị giác máy tính và Khoa học máy tính nói chung, hay bài toán phân lớp ảnh nói riêng đang phát triển rất mạnh mẽ và được sự ủng hộ cũng như đầu tư của nhiều tổ chức trên thế giới. Rõ ràng và dễ nhận thấy nhất là các mô hình mạng neural network được xây dựng và sử dụng ngày càng nhiều. Nhiều cuộc thi được tổ chức thường niên để cập nhật tình hình cũng là sản phẩm cho những mô

hình học máy và học sâu triển vọng. Trong lĩnh vực phân lớp ảnh có thể nói đến The ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Cuộc thi ILSVRC có thể được xem là nơi bắt đầu của nhiều mô hình mạng Neural Network nổi tiếng hiện nay trong lĩnh vực phân lớp ảnh như AlexNet, GoogleNet, VGG,....

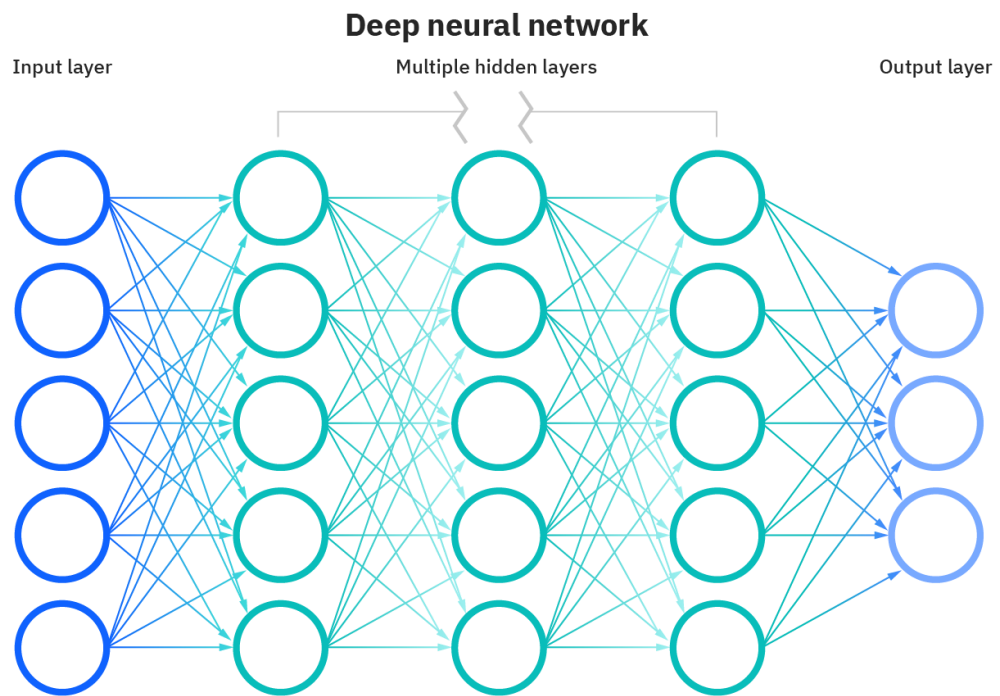
Chương 3 Cơ sở lý thuyết

3.1 Neural Network

Không sai khi nói “ Mạng Nơ-ron là một bước đột phá và là một bước tiến dài trong lĩnh vực Trí tuệ nhân tạo và học máy, tạo ra một lĩnh vực nghiên cứu mới – Học Sâu (Deep Learning). Mạng Nơ-ron xây dựng dựa trên mạng nơ-ron sinh học, các nơ-ron xử lý thông tin, kết nối với nhau và tính các giá trị.

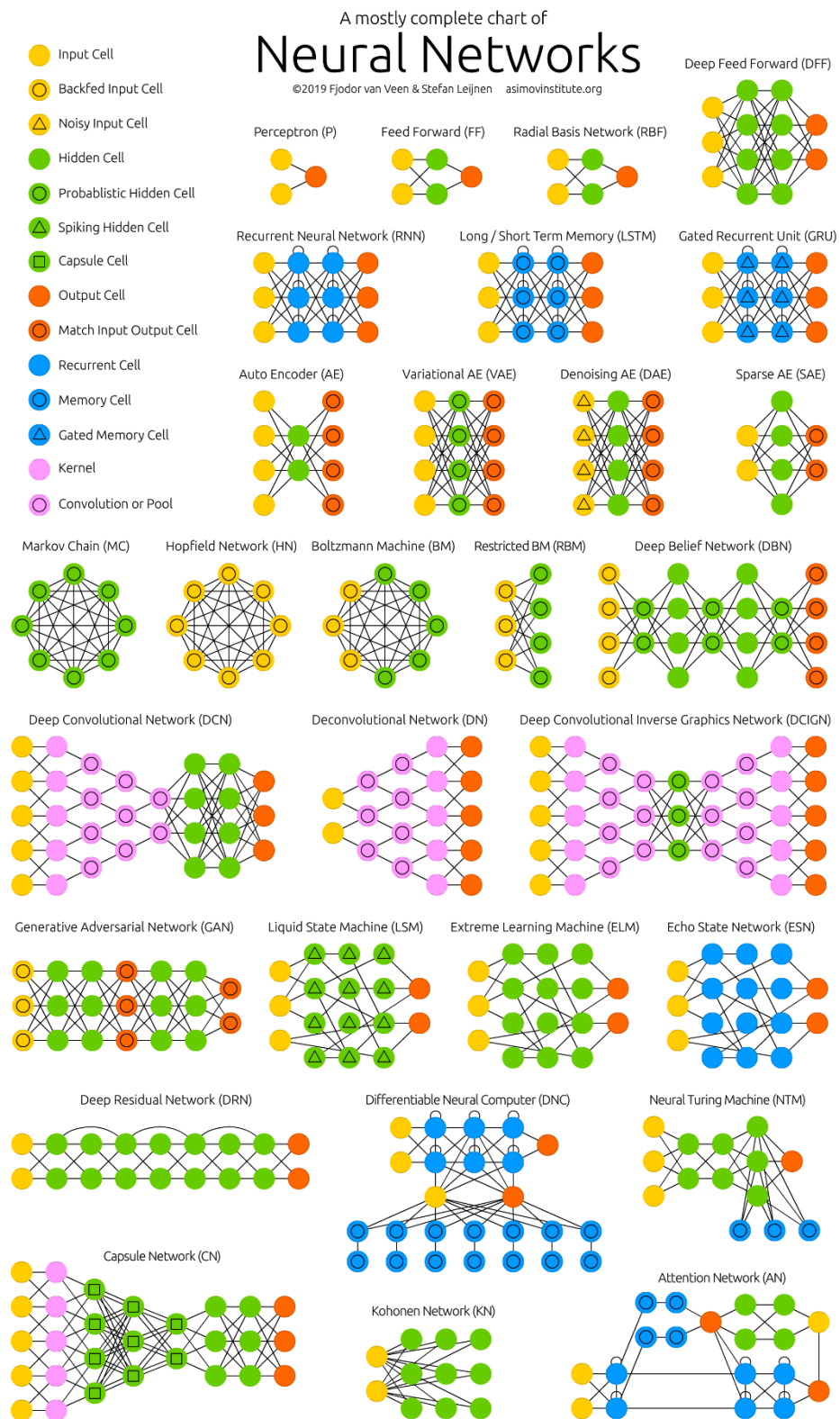
Về cơ bản, mạng nơ-ron có 3 lớp chính: Input, các lớp Hidden giữa, và Output. Mỗi nút trong mạng sẽ nhận thông tin dữ liệu đầu vào, tính toán dữ liệu đầu vào bằng cách tính tổng các input với trọng số tương ứng của đầu vào đó, sau đó dùng một hàm biến đổi phi tuyến tính cho phép biến đổi này để tính toán trạng thái trung gian. Hàm biến đổi được gọi là Activation Function. Output của lớp trước sẽ là input của lớp sau. Tùy vào loại mạng sẽ có các loại kết nối giữa các lớp khác nhau.

Thông tin dữ liệu từ đầu vào, thông qua nhiều lớp và các hàm biến đổi phi tuyến sẽ kết hợp lại tạo ra một dự đoán. Quá trình ấy cũng được xem là quá trình học của một mạng Nơ-ron. Mạng nơ-ron học thông qua tín hiệu lỗi do sự khác biệt của dự đoán so với giá trị mong muốn, sau đó cập nhật lại các bias và trọng số trong activation function, từ đó nâng cao tính chính xác cho mạng.



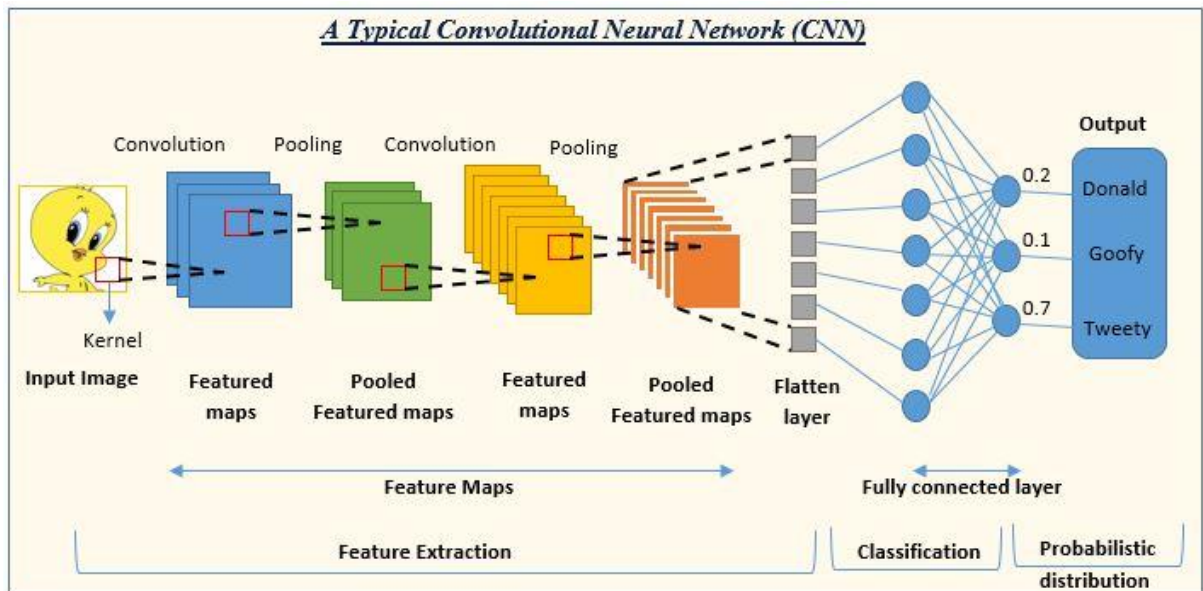
Hình 3-1: Ảnh minh họa Neural Network (nguồn: IBM)

Mạng nơ-ron có rất nhiều biến thể và kiến trúc khác nhau mà trong khóa luận này chúng tôi không thể trình bày hết được, chỉ có thể khái quát các loại mạng Nơ-ron bằng bảng dưới đây, mong sau này có cơ hội trình chi tiết hơn.



Hình 3-2: Hình minh họa các loại mạng Nơ-ron hiện nay (nguồn: trên ảnh)

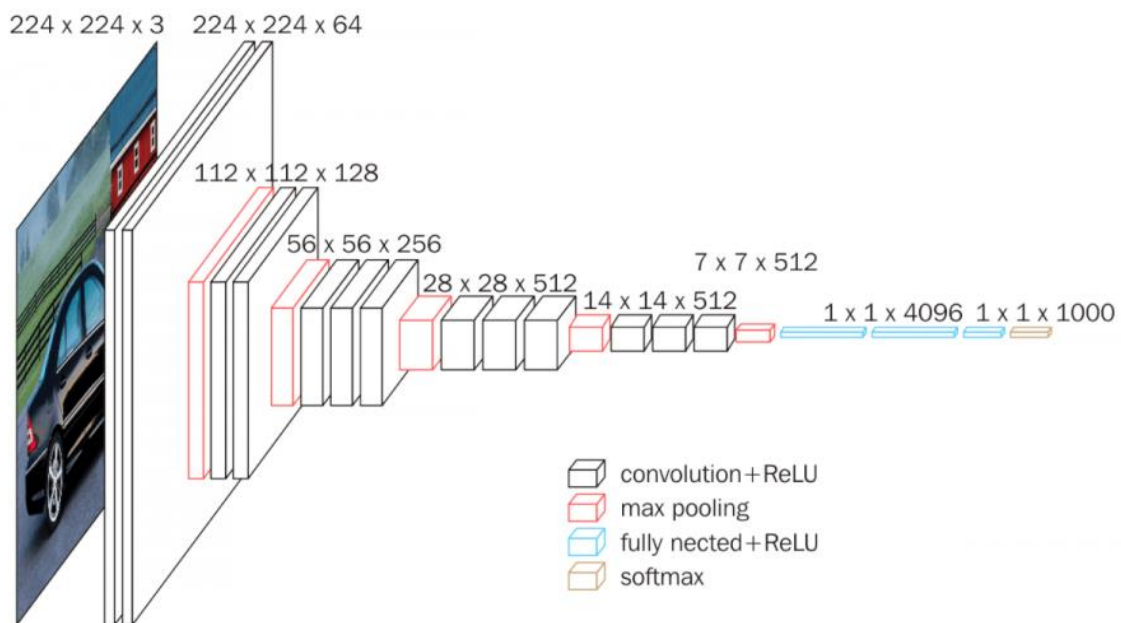
Trong khóa luận, mô hình phân lớp của chúng tôi dựa vào 3 mô hình mạng Convolutional Neural Network(CNN). Kiến trúc CNN sẽ dùng các ma trận thay thế hình ảnh từ đó sẽ đưa ra đặc trưng.



Hình 3-3: Ảnh minh họa kiến trúc CNN thông thường (nguồn: [Convolutional Neural Network: An Overview](#))

3.2 VGG-16

VGG16(hay còn gọi là ConvNet) là một mô hình mạng nơ-ron tích tụ được đề xuất bởi K. Simonyan và A. Zisserman từ Đại học Oxford trong bài báo “Mạng lưới kết hợp rất sâu để nhận dạng hình ảnh quy mô lớn”. Mô hình này đạt độ chính xác trong bài kiểm tra top 5 là 92,7% trong ImageNet, đây là tập dữ liệu của hơn 14 triệu hình ảnh thuộc 1000 lớp. Đây là một trong những mô hình nổi tiếng được đề trình lên ILSVRC-2014. Nó tạo ra sự cải tiến so với AlexNet bằng cách thay thế các bộ lọc có kích thước hạt nhân lớn (11 và 5 trong lớp chập đầu tiên và thứ hai, tương ứng) bằng nhiều bộ lọc kích thước hạt nhân 3×3 lần lượt. VGG16 đã được đào tạo trong nhiều tuần và đang sử dụng GPU NVIDIA Titan Black.

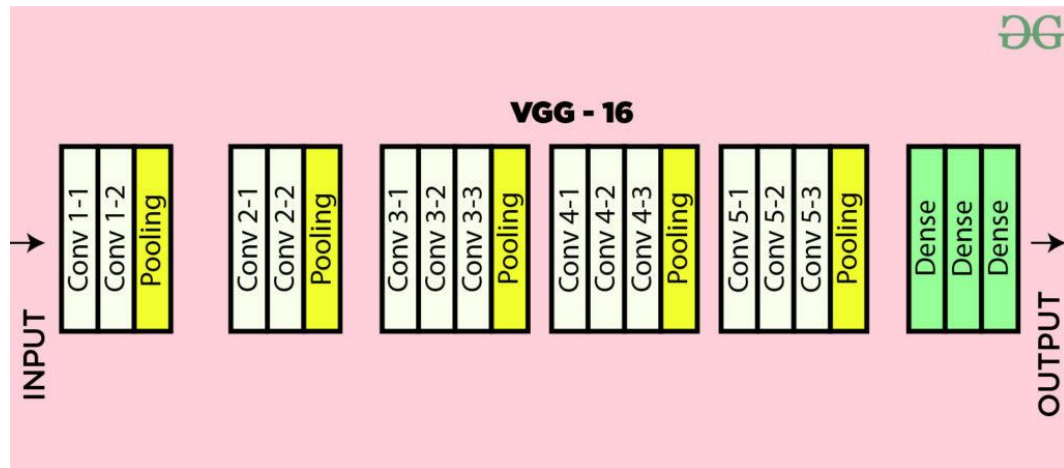


Hình 3-4: Hình minh họa kiến trúc VGG (Nguồn: [VGG16 - Convolutional Network for Classification and Detection](#))

Đầu vào của mạng là một hình ảnh có kích thước (224, 224, 3). Hai lớp đầu tiên có 64 kênh với kích thước bộ lọc 3×3 và đệm giống nhau. Sau đó, sau khi lớp hồ bơi tối đa của sai phân (2, 2), hai lớp có các lớp tích chập của kích thước bộ lọc 256 và kích thước bộ lọc (3, 3). Tiếp theo là lớp sai phân tổng hợp tối đa (2, 2) giống như lớp trước đó. Sau đó, có 2 lớp tích chập kích thước bộ lọc (3, 3) và 256 bộ lọc. Sau đó, có 2 bộ gồm 3 lớp chập và một lớp nhóm tối đa. Mỗi bộ lọc có 512 bộ lọc có kích thước (3, 3) với cùng một phần đệm. Hình ảnh này sau đó được chuyển đến ngăn xếp của hai lớp tích chập. Trong các lớp tích hợp và tổng hợp tối đa này, các bộ lọc chúng tôi sử dụng có kích thước 3×3 thay vì 11×11 trong AlexNet và 7×7 trong ZF-Net. Trong một số lớp, nó cũng sử dụng pixel 1×1 được sử dụng để thao tác số lượng kênh đầu vào. Có một phần đệm 1 pixel (đệm giống nhau) được thực hiện sau mỗi lớp tích chập để ngăn chặn đặc điểm không gian của hình ảnh.

Sau lớp tích lũy và lớp gộp tối đa, chúng tôi có một bản đồ đối tượng (7, 7, 512). Chúng tôi làm phẳng đầu ra này để biến nó thành một vectơ đặc trưng (1, 25088). Sau đó, có 3 lớp được kết nối đầy đủ, lớp đầu tiên lấy đầu vào từ vectơ đặc

trung cuối cùng và xuất ra vector (1, 4096), lớp thứ hai cũng xuất ra vector có kích thước (1, 4096) nhưng lớp thứ ba xuất ra 1000 kênh cho 1000 lớp của thử thách ILSVRC tức là lớp thứ 3 được kết nối đầy đủ được sử dụng để triển khai chức năng softmax để phân loại 1000 lớp. Tất cả các lớp ẩn sử dụng ReLU làm chức năng kích hoạt của nó. ReLU hiệu quả hơn về mặt tính toán vì nó giúp học tập nhanh hơn và nó cũng làm giảm khả năng biến mất các vấn đề về độ dốc.



Hình 3-5: Hình minh họa kiến trúc VGG-16 (nguồn: trên ảnh)

Bảng dưới đây liệt kê các kiến trúc VGG khác nhau. Chúng ta có thể thấy rằng có 2 phiên bản VGG-16 (C và D). Không có nhiều sự khác biệt giữa chúng ngoại trừ một điều là ngoại trừ một số lớp tích chập, (3, 3) tích chập kích thước bộ lọc được sử dụng thay vì (1, 1). Hai thứ này lần lượt chứa 134 triệu và 138 triệu tham số.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224×224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64	conv3-64	conv3-64	conv3-64
maxpool					
conv3-128	conv3-128	conv3-128	conv3-128	conv3-128	conv3-128
maxpool					
conv3-256	conv3-256	conv3-256	conv3-256	conv3-256	conv3-256
conv3-256	conv3-256	conv3-256	conv3-256	conv3-256	conv3-256
maxpool					
conv3-512	conv3-512	conv3-512	conv3-512	conv3-512	conv3-512
conv3-512	conv3-512	conv3-512	conv3-512	conv3-512	conv3-512
maxpool					
conv3-512	conv3-512	conv3-512	conv3-512	conv3-512	conv3-512
conv3-512	conv3-512	conv3-512	conv3-512	conv3-512	conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Bảng 3-1: Bảng so sánh cấu trúc các kiến trúc VGG

Localization đối tượng trong hình ảnh: Để perform localization, chúng ta cần thay thế điểm lớp bằng tọa độ vị trí hộp giới hạn. Vị trí hộp giới hạn được biểu thị bằng vector 4-D (tọa độ tâm (x, y), chiều cao, chiều rộng). Có hai phiên bản của kiến trúc bản địa hóa, một là hộp giới hạn được chia sẻ giữa các ứng viên khác nhau (đầu ra là 4 vector tham số) và phiên bản kia là hộp giới hạn dành riêng cho lớp (đầu ra là 4000 vector tham số). Bài báo đã thử nghiệm với cả hai cách tiếp cận trên kiến trúc VGG -16 (D). Ở đây, chúng ta cũng cần thay đổi tổn thất từ tổn thất phân loại sang các hàm tổn thất hồi quy (chẳng hạn như MSE) xử phạt sự sai lệch của tổn thất dự đoán so với sự thật cơ bản.

Kết quả: VGG-16 là một trong những kiến trúc hoạt động tốt nhất trong thử thách ILSVRC năm 2014, đứng thứ nhất trong nhiệm vụ phân loại với lỗi phân loại top 5 là 7,32% (chỉ sau GoogLeNet với lỗi phân loại 6,66%). Nó cũng là người chiến thắng trong nhiệm vụ bản địa hóa với 25,32% lỗi localization.

Hạn chế của VGG 16:

Nó rất chậm để đào tạo (mô hình VGG ban đầu được đào tạo trên GPU Nvidia Titan trong 2-3 tuần).

Kích thước của trọng lượng imageNet được đào tạo VGG-16 là 528 MB. Vì vậy, nó chiếm khá nhiều dung lượng ổ đĩa và băng thông khiến nó hoạt động kém hiệu quả.

138 triệu tham số dẫn đến vấn đề độ dốc bùng nổ.

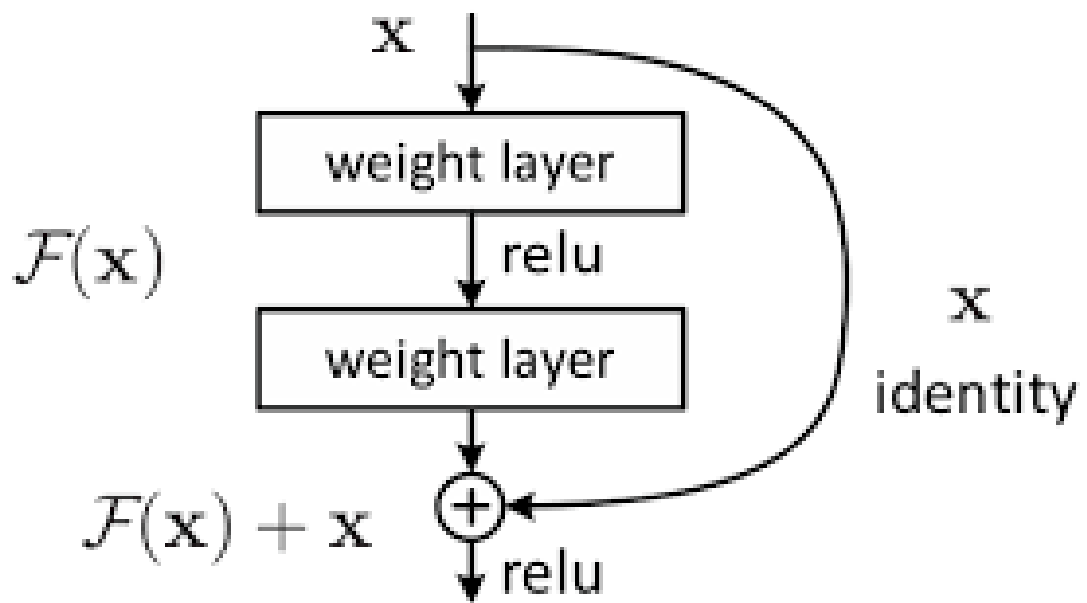
3.3 ResNet50

Residual Network(hay còn gọi là ResNet) là một mạng neural nhân tạo được giới thiệu lần đầu bởi Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun trong một bài báo nghiên cứu khoa học về thị giác máy tính tên “Deep Residual Learning for Image Recognition” vào năm 2015. Và ngay trong năm ấy, ResNet chứng tỏ sức mạnh và sự thành công của mình bằng những giải thưởng và thứ hạng đạt được trong cuộc thi ILSRVC. ResNet đạt vị trí thứ nhất trong phần thi phân lớp với độ lỗi chỉ 3.57%. Bên cạnh đó, ResNet cũng giành giải nhất trong các phần thi về nhận dạng, phân vùng của ImageNet và COCO.

ResNet có rất nhiều biến thể tùy thuộc vào số lớp của chúng. ResNet50 mà chúng tôi đang sử dụng có 50 lớp mạng nơ-ron.

Mạng ResNet ra đời giải quyết vấn đề Vanishing Gradient bằng một giải pháp được gọi là skip-connection (kết nối tắt). ResNet34 là kiến trúc ResNet đầu tiên, lấy cảm hứng từ kiến trúc VGG(VGG-16, VGG-19) với các filter tích chập 3x3. Tuy nhiên, ResNet có ít filter hơn và độ phức tạp thấp hơn.

Kết nối tắt trong ResNet (Skip Connections)



Hình 3-6: Hình minh họa Skip Connection

Các kết nối bỏ qua này hoạt động theo hai cách. Đầu tiên, họ làm giảm bớt vấn đề về độ dốc biến mất bằng cách thiết lập một phím tắt thay thế cho độ dốc đi qua. Ngoài ra, chúng cho phép mô hình học một chức năng nhận dạng. Điều này đảm bảo rằng các lớp cao hơn của mô hình không hoạt động kém hơn các lớp thấp hơn.

Tóm lại, các residual blocks giúp các lớp học các chức năng nhận dạng dễ dàng hơn đáng kể. Kết quả là, ResNet cải thiện hiệu quả của mạng nơ-ron sâu với nhiều lớp nơ-ron hơn trong khi giảm thiểu tỷ lệ lỗi. Nói cách khác, các kết nối bỏ qua thêm đầu ra từ các lớp trước vào đầu ra của các lớp xếp chồng lên nhau, giúp cho việc đào tạo các mạng sâu hơn nhiều so với trước đây có thể.

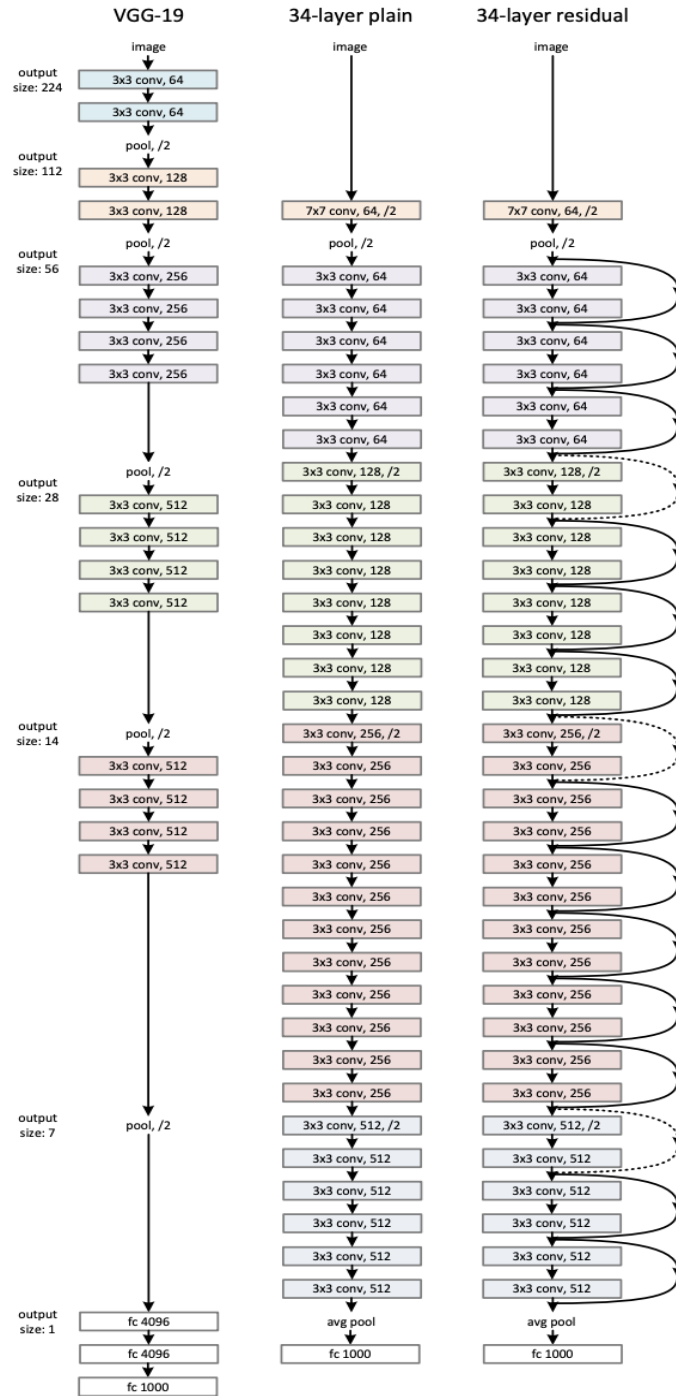


Figure 3. Example network architectures for ImageNet. **Left:** the VGG-19 model [41] (19.6 billion FLOPs) as a reference. **Middle:** a plain network with 34 parameter layers (3.6 billion FLOPs). **Right:** a residual network with 34 parameter layers (3.6 billion FLOPs). The dotted shortcuts increase dimensions. **Table 1** shows more details and other variants.

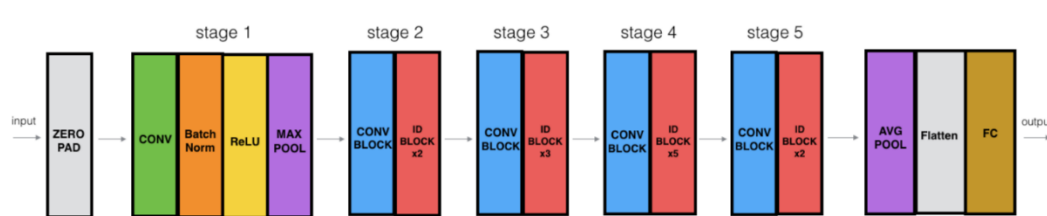
Hình 3-7: Hình minh họa kiến trúc VGG-19 và ResNet34

Nó cũng tuân theo hai quy tắc thiết kế đơn giản - các lớp có cùng số lượng bộ lọc cho cùng kích thước bản đồ đối tượng địa lý đầu ra và số bộ lọc tăng gấp đôi trong trường hợp kích thước bản đồ đối tượng địa lý giảm đi một nửa để duy trì độ phức tạp về thời gian cho mỗi lớp. Nó bao gồm 34 lớp có trọng số.

Các kết nối lồi tắt đã được thêm vào mạng đơn giản này. Mặc dù kích thước đầu vào và đầu ra giống nhau, nhưng các phép tắt nhận dạng đã được sử dụng trực tiếp. Với sự gia tăng về kích thước, có hai lựa chọn được xem xét. Đầu tiên là phép tắt sẽ vẫn thực hiện ánh xạ danh tính trong khi các mục nhập số 0 bổ sung sẽ được đệm để tăng kích thước. Tùy chọn khác là sử dụng phép tắt chiếu để khớp với kích thước.

ResNet50 cũng tương tự mô hình ResNet34 tiền nhiệm. Thay đổi thêm thiết kế nút thắt cổ chai từ building blocks do vấn đề về thời gian huấn luyện. Vì vậy sử dụng 3 lớp chồng thay vì 2 lớp. Sự thay đổi này đã làm tăng tính chính xác cho mô hình ResNet50. ResNet50 đạt 3.8 tỉ vòng FLOP.

Dưới đây là diagram ResNet50:



Hình 3-8: Hình minh họa về ResNet50 (nguồn: [Keras Implementation of ResNet-50 \(Residual Networks\) Architecture from Scratch - MLK - Machine Learning Knowledge](#))

- Zero-padding : Input với (3,3)
- Stage 1 : Tích chập (Conv1) với 64 filters với shape(7,7), sử dụng stride (2,2). BatchNorm, MaxPooling (3,3).
- Stage 2 : Convolutional block sử dụng 3 filter với size 64x64x256, f=3, s=1. Có 2 Identity blocks với filter size 64x64x256, f=3.

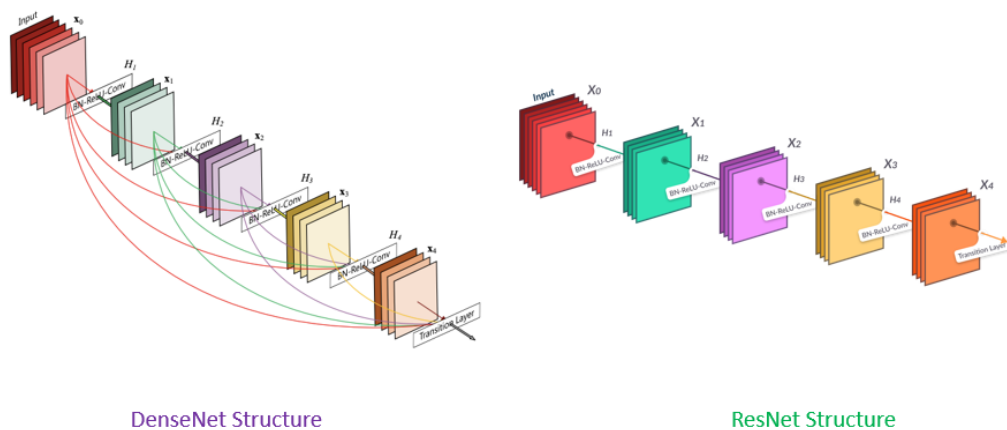
- Stage 3 : Convolutional sử dụng 3 filter size $128 \times 128 \times 512$, $f=3, s=2$. Có 3 Identity blocks với filter size $128 \times 128 \times 512$, $f=3$.
- Stage 4 : Convolutional sử dụng 3 filter size $256 \times 256 \times 1024$, $f=3, s=2$. Có 5 Identity blocks với filter size $256 \times 256 \times 1024$, $f=3$.
- Stage 5 : Convolutional sử dụng 3 filter size $512 \times 512 \times 2048$, $f=3, s=2$. Có 2 Identity blocks với filter size $512 \times 512 \times 2048$, $f=3$.
- The 2D Average Pooling : sử dụng với kích thước (2,2).
- The Flatten.
- Fully Connected (Dense) : sử dụng softmax activation.

3.4 DenseNet121

Densely Connected Convolutional Networks (hay còn gọi là DenseNet) được nghiên cứu và công bố bởi Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger trong bài báo “ Densely Connected Convolutional Networks” vào năm 2016. Cho đến nay, DenseNet vẫn được xem là một trong những mạng neural network mới, phục vụ cho phân lớp và nhận dạng ảnh. Giống với mạng ResNet, DenseNet vẫn giải quyết vấn đề về vanishing gradient.

DenseNet có cấu trúc đặc trưng với 2 phần denseblock và transition layers. Với mạng CNN truyền thống, nếu có L lớp thì sẽ có L kết nối, tuy nhiên với DenseNet thì sẽ có $L(L+1)/2$ kết nối.

Hình bên dưới đây mô phỏng các kết nối của DenseNet và ResNet, ta có thể thấy, output của lớp trước sẽ input của các lớp phía sau thông qua một hàm được gọi là composite function operation. Composite function operation này bao gồm các lớp tích chập, lớp pooling, batch normalization và hàm kích hoạt phi tuyến.



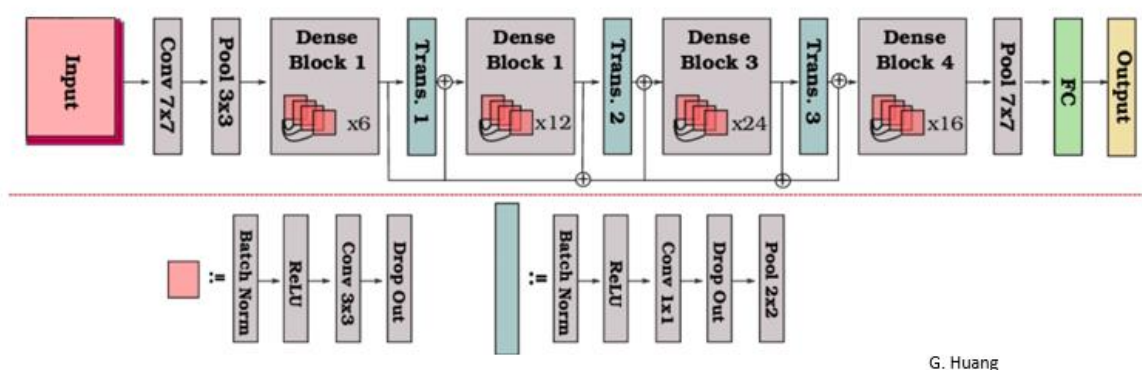
$$a^{[l]} = g([a^{[0]}, a^{[1]}, a^{[2]}, \dots, a^{[l-1]}])$$

$$a^{[l]} = g(z^{[l+1]} + a^{[l]})$$

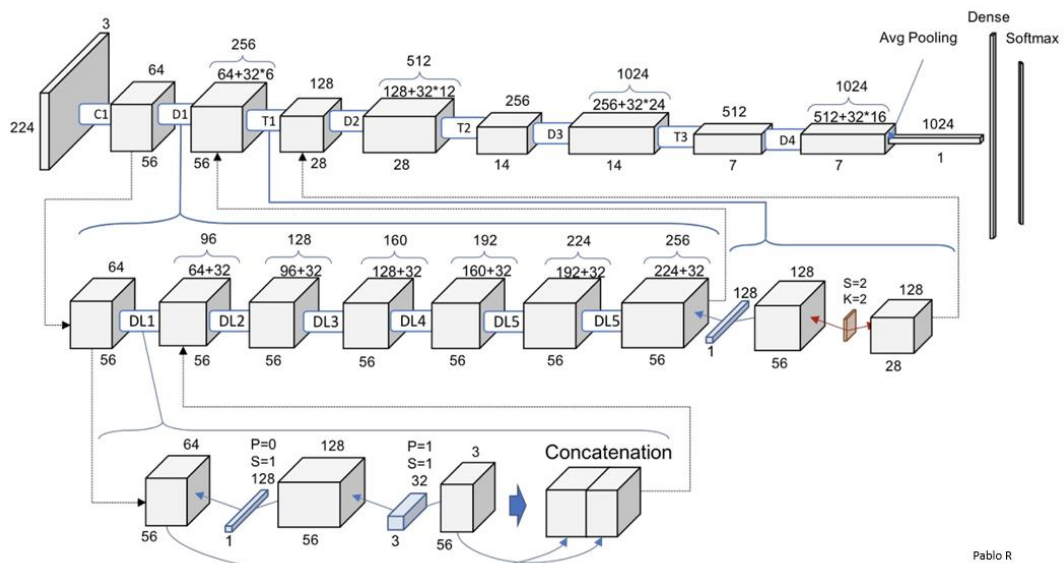
Hình 3-9: Hình minh họa kiến trúc DenseNet và ResNet

DenseNet có rất nhiều phiên bản khác nhau, tương tự như với ResNet, tùy thuộc vào số lượng các lớp sẽ có DenseNet 121, DenseNet 160,...

DenseNet chia các DenseBlock với số lượng filter khác nhau nhưng kích thước thì vẫn giống nhau. Transition Layer áp dụng batch normalization bằng cách downsampling, là một bước quan trọng trong CNN. Hình bên dưới minh họa cho cơ chế vừa nêu.



Hình 3-10: Hình minh họa cho DenseBlock và Transition Layers(nguồn: Densely Connected Convolutional Networks)



Hình 3-11: Hình minh họa toàn bộ kiến trúc DenseNet (nguồn: Pablo R)

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112×112	7×7 conv, stride 2			
Pooling	56×56	3×3 max pool, stride 2			
Dense Block (1)	56×56	$\left[\begin{matrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{matrix} \right] \times 6$	$\left[\begin{matrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{matrix} \right] \times 6$	$\left[\begin{matrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{matrix} \right] \times 6$	$\left[\begin{matrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{matrix} \right] \times 6$
Transition Layer (1)	56×56	1×1 conv			
	28×28	2×2 average pool, stride 2			
Dense Block (2)	28×28	$\left[\begin{matrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{matrix} \right] \times 12$	$\left[\begin{matrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{matrix} \right] \times 12$	$\left[\begin{matrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{matrix} \right] \times 12$	$\left[\begin{matrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{matrix} \right] \times 12$
Transition Layer (2)	28×28	1×1 conv			
	14×14	2×2 average pool, stride 2			
Dense Block (3)	14×14	$\left[\begin{matrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{matrix} \right] \times 24$	$\left[\begin{matrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{matrix} \right] \times 32$	$\left[\begin{matrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{matrix} \right] \times 48$	$\left[\begin{matrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{matrix} \right] \times 64$
Transition Layer (3)	14×14	1×1 conv			
	7×7	2×2 average pool, stride 2			
Dense Block (4)	7×7	$\left[\begin{matrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{matrix} \right] \times 16$	$\left[\begin{matrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{matrix} \right] \times 32$	$\left[\begin{matrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{matrix} \right] \times 32$	$\left[\begin{matrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{matrix} \right] \times 48$
Classification Layer	1×1	7×7 global average pool			
		1000D fully-connected, softmax			

Bảng 3-2: Bảng tham số của các kiến trúc DenseNet (nguồn : Densely Connected Convolutional Networks)

Dense cũng đạt được nhiều ưu điểm:

- Accuracy tương đương với ResNet dùng tham số giảm gần một nửa trên bộ dữ liệu ImageNet classification.
- Giảm tình trạng overfitting và vanish gradient.

3.5 Ensemble

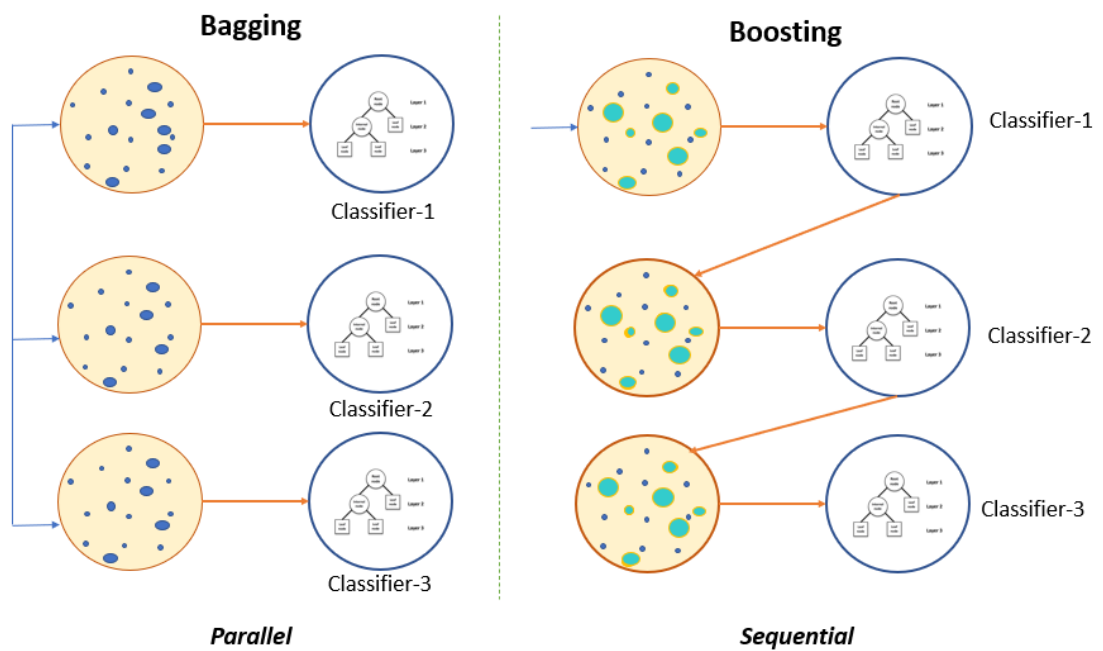
Trong Machine Learning, tồn tại một định lý “ Không có thuật toán hay mô hình học máy nào là hoàn hảo cho tất cả dạng bài toán hay bộ dữ liệu”. Thật vậy, tùy vào tính chất dữ liệu và ứng dụng bài toán sẽ có những thuật toán phù hợp. Hầu hết các thuật toán và mô hình đều dựa vào một ý tưởng giả thiết nào đó mà cài đặt. Nếu chỉ sử dụng một mô hình thuật toán, ta phải tốn thời gian, công sức để thử tất cả các thuật toán, mô hình để tìm ra cái phù hợp nhất cho nhu cầu bài toán và dữ liệu, hay phải tinh chỉnh tham số để đạt độ chính xác cao nhất có thể. Thay vì vậy, chúng ta có thể “kết hợp” hai hay nhiều mô hình lại với nhau, từ đó rút ra kết quả dự đoán cuối cùng. Phương pháp này gọi là ensemble learning. Ý tưởng để thực hiện phương pháp này là chúng ta sẽ kết hợp nhiều mô hình có triển vọng hoặc thu được độ chính xác cao ở mức chấp nhận được, tổng hợp các kết quả chúng lại và tạo ra một mô hình kết hợp đạt hiệu suất cao hơn. Phương pháp ensemble learning chia thành 3 loại:

- Bagging (đóng bao):

Những mô hình sẽ được huấn luyện riêng biệt trên cùng một bộ dữ liệu một cách song song, sau đó dùng kết quả của các mô hình để đưa ra kết quả dự đoán cuối cùng. Kết quả của từng mô hình có thể được xem là một bộ thông số cố định để đưa ra quyết định cuối cùng.

- Boosting (tăng cường) :

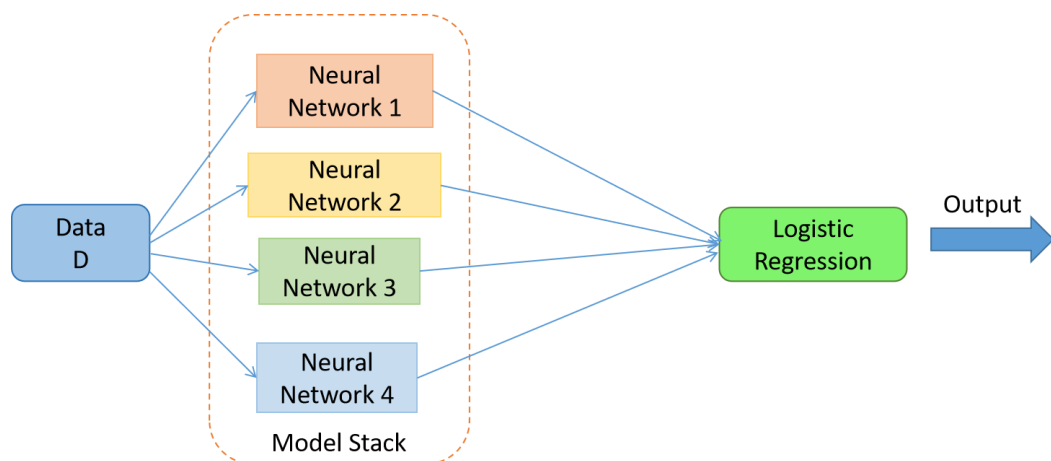
Các mô hình sẽ được huấn luyện lần lượt, mô hình đầu tiên sẽ huấn luyện với toàn bộ dữ liệu, mô hình thứ 2 sẽ huấn luyện các phần từ mà mô hình thứ nhất dự đoán sai. Lần lượt như thế tương ứng với n mô hình. Các mô hình sẽ huấn luyện xâu chuỗi với nhau.



Hình 3-12: Hình minh họa phương pháp Bagging và Boosting

- Stacking :

Gần tương tự với phương pháp Bagging, tuy nhiên, thay vì chỉ dựa vào kết quả các mô hình và chọn ra mô hình tốt nhất thì sẽ có một phương pháp để kết hợp tất cả các mô hình được sử dụng. Có thể là hồi quy hoặc tính toán có trọng số dựa vào độ chính xác của các mô hình



Hình 3-13: Hình minh họa phương pháp stacking

3.6 Ảnh X-quang

Ảnh X-quang(hay còn gọi là Xray) là một loại ảnh y khoa, dùng tia X có bức xạ cao, các tia X đi xuyên qua các mô mềm, xương và thành phần dịch trong cơ thể, từ đó tái tạo hình các cơ quan và vùng trong cơ thể con người được chụp.

Đậm độ là số vật chất một khoảng không gian nhất định. Đối với các mô có mật độ đặc dày như xương thì X sẽ ít đi xuyên qua nhiều. Ngược lại, không khí và các chất như nước, dung dịch sẽ được tia X xuyên qua nhiều. Vùng trên ảnh X-quang càng đen thì chứng tỏ mô ở đó ít. Việc xác định đậm độ dựa trên tính bất thường là điều mà chúng tôi hướng tới trong khóa luận.

Như đã giới thiệu ban đầu, ảnh X-quang đóng vai trò quan trọng, là bước đầu trong quá trình chẩn đoán và điều trị bệnh của đa số các bệnh viện và bác sĩ chuyên gia.



Hình 3-14: Hình minh họa ảnh X-quang thuộc Shenzhen Dataset

3.7 Transfer Learning

Các kỹ thuật học máy và khai thác dữ liệu đã được sử dụng rộng rãi trong nhiều lĩnh vực. Các phương pháp học máy truyền thống lấy các dữ liệu huấn luyện và dữ liệu thử nghiệm của cùng một nguồn để đạt sự tương đồng của phân phối dữ liệu và không gian đặc trưng. Tuy nhiên, trong thực tế sẽ có nhiều trường hợp bộ dữ liệu không đạt tiêu chí trên do số lượng dữ liệu ít hoặc dữ liệu bị mất cân bằng. Điều này sẽ dẫn đến kết quả dự đoán của mô hình bị giảm sút.

Do đó, cần tạo ra các mô hình đạt hiệu suất cao cho một mục tiêu nhất định mặc dù huấn luyện với nhiều dữ liệu từ nhiều nguồn khác nhau nhằm giải quyết vấn đề dữ liệu trong tương lai. Đây chính là động lực để phát triển kỹ thuật Transfer learning.

Transfer learning là kỹ thuật sử dụng những tri thức đã có từ việc học một vấn đề này sang ứng dụng thực hiện một vấn đề khác có liên quan. Ví dụ trong đời sống về hai người muốn học chơi piano. Một người không có kinh nghiệm về âm nhạc và một người đã có kiến thức về âm nhạc thông qua việc chơi guitar. Có thể dễ dàng thấy được, người có kiến thức âm nhạc sẽ học chơi piano nhanh hơn người không có kiến thức. Xem xét trong lĩnh vực máy học, các mô hình pretrained có chất lượng tốt và đạt độ chính xác cao trong bài toán phân loại sẽ được tận dụng lại và cải thiện trên các tác vụ phân loại tương tự.

Việc tận dụng lại các trọng số trong mô hình pretrained đạt kết quả tốt cho các bài toán tương tự sẽ giúp giảm thời gian huấn luyện trong khi kết quả dự đoán vẫn đảm bảo sẽ đạt độ chính xác cao do các trọng số đã được scale về giá trị tối ưu. Trong trường hợp bộ dữ liệu có kích thước nhỏ, các mô hình pretrained sẽ giúp mô hình dự đoán tốt hơn do mô hình được học trên hai nguồn tri thức bao gồm dữ liệu huấn luyện và dữ liệu đã được học trước đó.

Kiến trúc mô hình sử dụng Transfer learning gồm 2 phần:

- **Base network:** dùng để trích xuất đặc trưng, được lấy từ một phần hoặc toàn bộ của mô hình pretrained.
- **Các lớp Fully Connected Layers:** giảm chiều dữ liệu và tính toán xác suất của output.

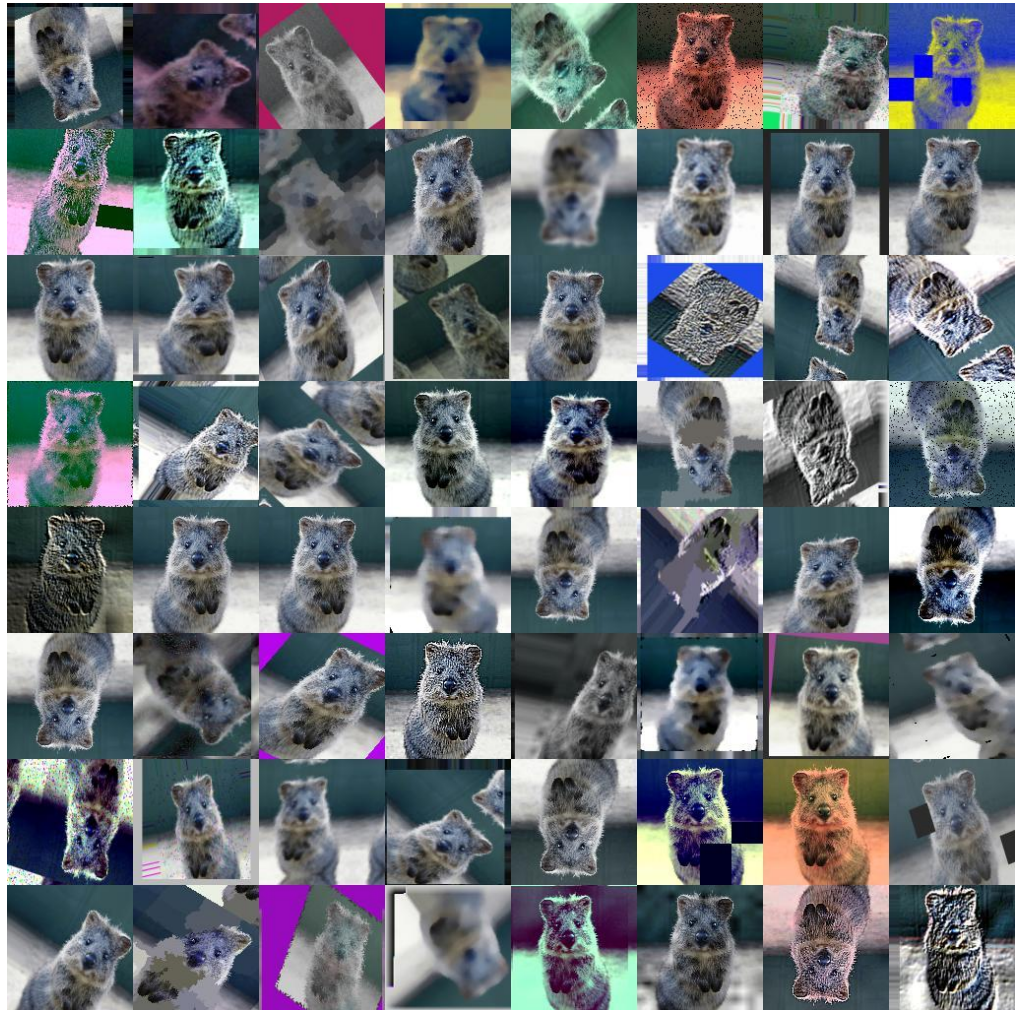
3.8 Fine tuning

Fine-tuning là một dạng phổ biến của kỹ thuật Transfer learning. Fine-tuning sử dụng lại một phần hoặc toàn bộ các layer trong mô hình pretrained kết hợp với các thao tác thêm/xóa/sửa 1 vài layer để tạo ra một mô hình mới.

Các layer đầu tiên của mô hình thường sẽ được đóng băng (freeze) – các trọng số sẽ không được cập nhật trong quá trình huấn luyện do các layer đầu tiên có khả năng trích xuất đặc trưng mức trừu tượng thấp.

Sử dụng phương pháp Fine-tuning giúp quá trình training diễn ra nhanh chóng do chỉ cập nhật các trọng số ở các lớp cuối trong khi vẫn đảm bảo độ chính xác của mô hình đạt hiệu suất cao.

3.9 Data Augmentation



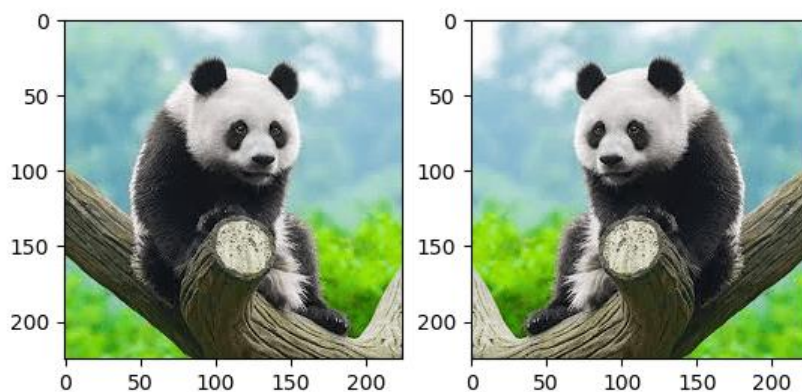
Hình 3-15: Hình minh họa Data Augmentation

Một số bài toán, lĩnh vực có ít dữ liệu cho việc train model. Điều này dẫn đến vấn đề về overfitting và khó tạo ra được kết quả tốt trong việc dự đoán của các mô hình học máy Deep Learning. Do đó, các kỹ thuật Data augmentation hay tăng cường dữ liệu được phát triển để giải quyết vấn đề trên.

Data augmentation là phương pháp phát sinh dữ liệu dựa trên các dữ liệu đã có. Mục đích là để tăng số lượng dữ liệu huấn luyện giúp cho các mô hình Deep Learning cập nhật trọng số tốt hơn, từ đó đạt hiệu suất cao trong việc dự đoán.

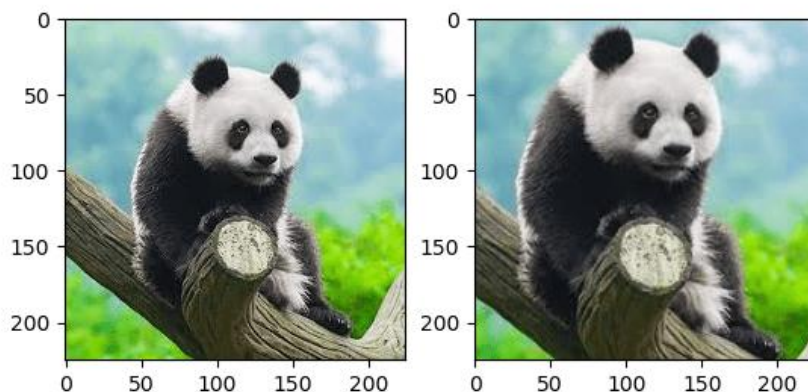
Các phương pháp Data Augmentation vô cùng đa dạng và tùy theo yêu cầu bài toán hoặc đặc điểm của bộ dữ liệu, ta có thể lựa chọn các phương pháp phù hợp để triển khai. Một số phương pháp Data augmentation phổ biến:

- Flip (lật): lật ảnh theo chiều dọc hoặc chiều ngang. Tuy nhiên, khi thực hiện phương pháp này phải đảm bảo rằng nhãn của ảnh vẫn được giữ nguyên. Ví dụ, lật ảnh số 6 thì sẽ thành số 9, như vậy là sai nhãn với ảnh và khiến mô hình huấn luyện sai.



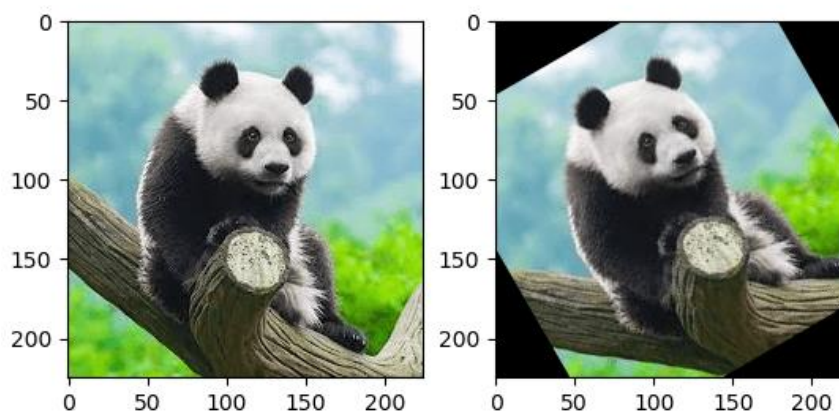
Hình 3-16: Kỹ thuật Plip

- Crop (cắt): cắt một phần trong ảnh và resize vùng ảnh đó về kích thước ảnh ban đầu. Tuy nhiên, phần cắt phải chứa chủ thể chính của bức ảnh cần được dự đoán để đảm bảo hiệu suất huấn luyện cho mô hình.



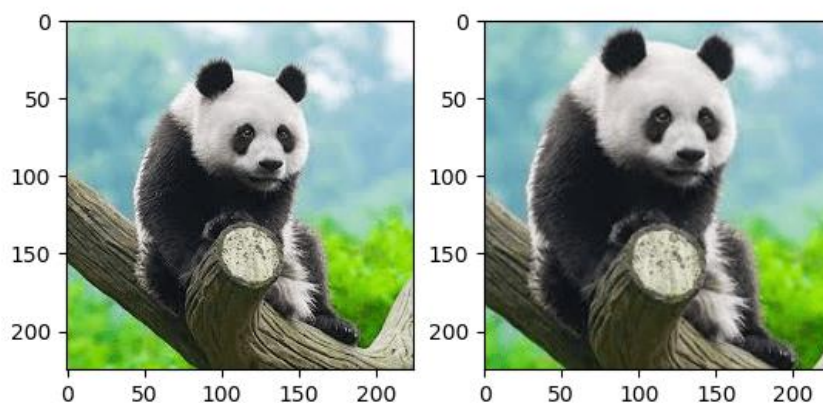
Hình 3-17: Kỹ thuật Crop

- Color shift (chuyển màu): chuyển màu của ảnh dữ liệu bằng cách thay đổi giá trị 3 kênh màu RGB.
- Contrast change (đổi độ tương phản): thay đổi độ tương phản của ảnh.
- Rotation (phép quay): quay ảnh theo nhiều góc khác nhau.



Hình 3-18: Kỹ thuật Rotation

- Scale: phóng to hoặc thu nhỏ ảnh.



Hình 3-19: Kỹ thuật Scale

Trong thư viện keras có hỗ trợ class ImageDataGenerator cho phép tạo thêm dữ liệu dựa trên các phương pháp Data augmentation phổ biến.

Chương 4 Phương pháp đề nghị

4.1 Ý tưởng mô hình

Với một bài toán phân lớp ảnh, việc tìm hiểu dữ liệu và tìm kiếm mô hình phù hợp là vô cùng cần thiết. Trên thực tế có rất nhiều phương pháp cũng như mô hình xử lý ảnh, tuy nhiên ở mức độ sinh viên, tìm được một mô hình phù hợp với trình độ cũng như hiểu biết khá là khó khăn. Những mô hình dùng để phân lớp đặc trưng cho phổi như ChestNet thì lại khá phức tạp để cài đặt. Vì vậy, sau khi đã thử nghiệm cũng như cài đặt một số mô hình riêng biệt nhưng kết quả lại không khả thi. Với cảm hứng từ một bài báo trên mạng về phân lớp phổi do tác động Covid 19, và ý tưởng “ Góp gió thành bão” , chúng tôi xin đề xuất phương pháp phân lớp ảnh X-quang về phổi bằng việc kết hợp 3 mô hình GoogleNet, ResNet50 và DenseNet121. Tuy nhiên, chúng tôi không dùng lại các mô hình mà các thư viện cung cấp như Keras, Tensorflow,... Trong lúc cài đặt các mô hình, chúng tôi đã áp dụng kỹ thuật fine-tuning để nâng cao độ chính xác. Thông qua kết quả huấn luyện từ 3 mô hình đưa về thông số Accuracy. Sau đó dùng phương pháp Ensemble Learning để đưa ra kết quả cuối cùng

4.2 Workflow

Đầu tiên, bộ dữ liệu huấn luyện sau khi tiền xử lý sẽ lần lượt đi qua 3 mô hình lựa chọn để cập nhật trọng số tối ưu cho các mô hình.

Sau khi huấn luyện, tiến hành chạy tập test trên từng mô hình để dự đoán và ta sẽ thu được độ accuracy (acc) của từng mô hình. acc được dùng để tính trọng số w cho từng mô hình.

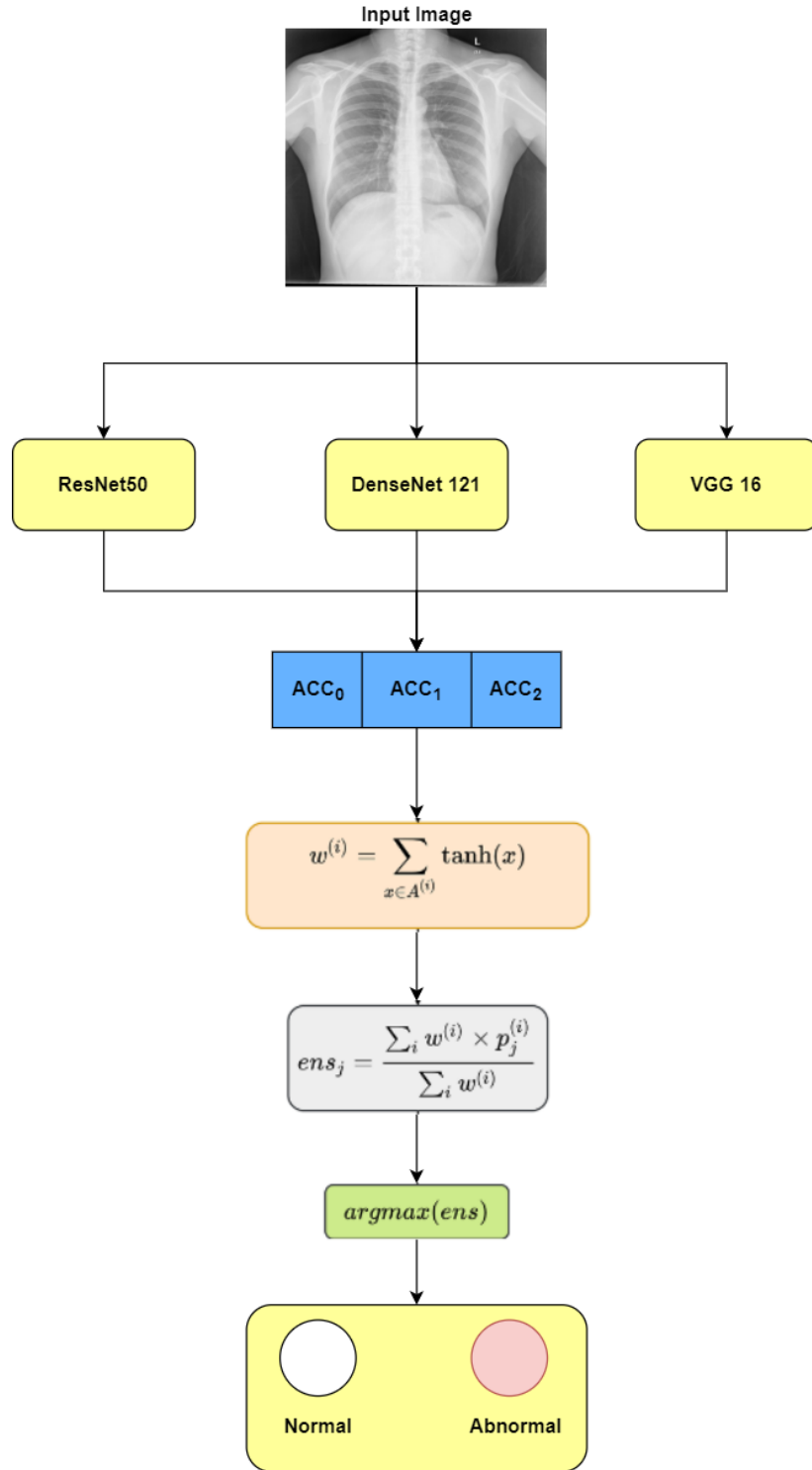
Sử dụng hàm tanh để tính trọng số, trọng số này sẽ quyết định mức độ phụ thuộc của kết quả đối với ba mô hình. Mô hình có độ chính xác càng cao thì trọng số càng lớn, chứng tỏ mức độ quan trọng của mô hình đó đến kết quả cuối cùng của toàn phương pháp càng cao. Hàm tanh được sử dụng vì giá trị của acc nằm trong

khoảng $[0,1]$, vì vậy nếu giá trị acc càng cao, hàm tanh sẽ tính ra trọng số càng lớn và ngược lại.

$$w^{(i)} = \sum_{x \in A_i} \tanh(x) = \sum_{x \in A_i} \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad \text{với } A_i = \{Acc_0, Acc_1, Acc_2\}$$

Sau khi tính được trọng số w , ta sẽ tính giá trị xác suất ensemble (ens) của từng mẫu dữ liệu. Ở phạm vi bài toán, ta chỉ có hai lớp là Normal và Abnormal. Vì vậy khi đưa tập dữ liệu thử nghiệm qua mỗi mô hình, ta sẽ thu được hai giá trị xác suất đại diện cho xác suất của mẫu dữ liệu thuộc về từng lớp là bao nhiêu. Gọi $p_j^{(i)}$ là mảng xác suất của mẫu thứ j đi qua mô hình thứ i , ta có: $p_j^{(i)} = \{a, 1 - a\}$. Sau khi đi qua hàm tính ens_j theo công thức trên, ta sẽ thu được mảng xác suất mới là: $ens_j = \{b, 1 - b\}$. Đây là mảng xác suất của mẫu thứ j sau khi kết hợp giá trị xác suất của ba kết quả thuộc ba mô hình tương ứng.

Cuối cùng, mảng xác suất đi qua hàm $argmax$ để tìm xác suất lớn nhất tương ứng với lớp mà mẫu thuộc về.



Hình 4-1: Workflow phương pháp Ensemble đề xuất. Acc = Accuracy; $w^{(i)}$ là trọng số của mô hình thứ i, $p_j^{(i)}$ là điểm xác suất cho mẫu thứ j của mô hình thứ i, ens_j là điểm xác suất ensemble của mẫu thứ j và hàm $argmax$ trả về giá trị lớn nhất của m

4.3 Kết luận và tổng quan

Trong khóa luận này, phương pháp của chúng tôi đã thử nghiệm nhiều mô hình trên bộ dữ liệu, đi kèm là những tinh chỉnh và áp dụng nhiều kỹ thuật trong học sâu như transfer learning và fine tuning. Bộ ba mô hình ResNet50, DenseNet 121, VGG 16 cho kết quả khả quan nhất khi được huấn luyện kiểm tra riêng biệt. Ba mô hình đều là mạng CNN và nhận input đầu vào là ảnh có kích thước 224x224.

Với mục đích nâng cao tính chính xác và hiệu quả, chúng tôi đã dùng phương pháp ensemble learning, kết hợp kết quả của ba mô hình lại dưới hình thức stacking. Kết quả mang lại đúng như dự định. Độ chính xác và hiệu quả cũng tăng cao lên.

Xét về tổng quan, từ ý tưởng cho đến triển khai cài đặt và kết quả đúng như kế hoạch ban đầu, đây được xem như một thành công. Chương tiếp theo sẽ cho chúng ta thấy rõ hơn về thực nghiệm và kết quả được.

Chương 5 Thực nghiệm và kết quả

5.1 Về tập dữ liệu

Trong một bài nghiên cứu, phần thực nghiệm như một bằng chứng rõ ràng nhất để đánh giá tính hiệu quả và thiết thực của một phương pháp. Để thực nghiệm với bài toán này, việc tìm kiếm tập dữ liệu là vô cùng quan trọng và cần độ cẩn thận. Vì tư liệu về y khoa và hình ảnh X-quang của bệnh nhân là vô cùng nhạy cảm và bảo mật, được bảo vệ về cả mặt pháp lí và đạo đức. Sau thời gian, tìm kiếm cũng như đánh giá mức độ đáng tin cậy và khả thi của dữ liệu, nhóm chúng tôi quyết định dùng bộ dữ liệu về ảnh X-quang của bệnh viện Shenzhen để thực nghiệm. Bộ dữ liệu này được công bố và cho phép sử dụng để nghiên cứu một cách công khai và không vi phạm các yếu tố về pháp lý cũng như đạo đức.

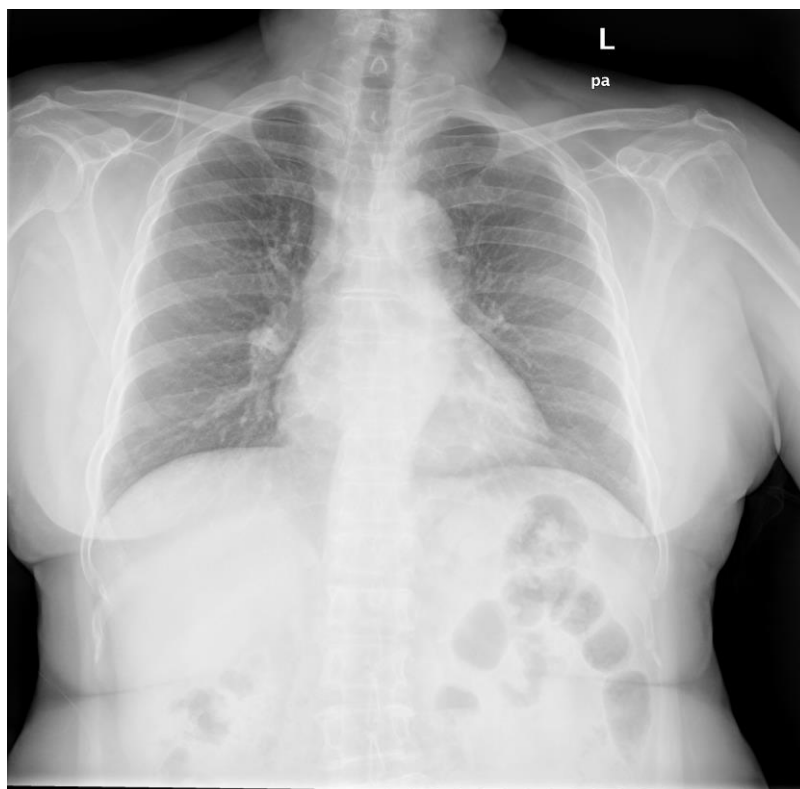
Bộ dữ liệu ảnh X-quang này được thu thập tại bệnh viện nhân dân số 3 Thẩm quyền (Shenzhen No.3 People's Hospital) thuộc tỉnh Quảng Đông, Trung Quốc. Ảnh được lấy trong quá trình khám và chữa bệnh thường quy tại bệnh viện, được tích góp qua thời gian. Bộ dữ liệu có tổng cộng 662 ảnh X-quang về phổi của cả người lớn và trẻ em, với 326 ảnh X-quang phổi bình thường và 336 ảnh X-quang phổi được chẩn đoán là có mắc bệnh. Tất cả ảnh X-quang được định dạng PNG format. Kích thước của ảnh không cố định, có ảnh lên đến 3Kx3K pixels. Ảnh được đặt tên theo template như sau: CHNCXR_####_X.png với:

- #### là số định danh của ảnh bệnh nhân.
- X là 1 nếu phổi bất thường (abnormal), 0 nếu phổi bình thường (normal).

Một file excel đi kèm, bên trong miêu tả số định dạng với tuổi và bệnh tương ứng (nếu có).

study_id	sex	age	findings
CHNCXR_0001_0.png	Male	45	normal
CHNCXR_0002_0.png	Male	63	normal
CHNCXR_0003_0.png	Female	48	normal
CHNCXR_0004_0.png	Male	58	normal
CHNCXR_0005_0.png	Male	28	normal
CHNCXR_0006_0.png	Male	60	normal
CHNCXR_0007_0.png	Female	53	normal
CHNCXR_0008_0.png	Male	48	normal
CHNCXR_0009_0.png	Female	71	normal
CHNCXR_0010_0.png	Female	59	normal
CHNCXR_0011_0.png	Male	43	normal
CHNCXR_0012_0.png	Male	39	normal
CHNCXR_0013_0.png	Male	41	normal
CHNCXR_0014_0.png	Male	39	normal
CHNCXR_0015_0.png	Female	28	normal
CHNCXR_0016_0.png	Female	57	normal
CHNCXR_0017_0.png	Male	35	normal
CHNCXR_0018_0.png	Male	42	normal
CHNCXR_0019_0.png	Male	27	normal
CHNCXR_0020_0.png	Male	32	normal
CHNCXR_0021_0.png	Female	38	normal

Bảng 5-1: Chi tiết bên trong file excel đi kèm



Hình 5-1: Ảnh X-quang phổi bình thường trong bộ dữ liệu Shenzhen



Hình 5-2: Ảnh minh họa phổi bất thường trong bộ dữ liệu Shenzhen

Bên cạnh dữ liệu gốc của sẵn trong bộ dữ liệu, để tăng thêm tính đa dạng, chúng tôi đã dùng kỹ thuật tăng dữ liệu với các phương pháp xử lý ảnh như scale, rotate, Được nêu bên trên phần cơ sở lý thuyết với mục đích làm đa dạng bộ dữ liệu.

5.2 Kết quả và đánh giá

Khi thực nghiệm trên tập dữ liệu Shenzhen đã nêu ở 5.1, tiền xử lý đưa ảnh input đầu vào về kích thước 224x224, đáp ứng yêu cầu của 3 mô hình CNN đề xuất. Kết quả về accuracy của 3 mô hình như sau:

- VGG 16: 0.8671875
- ResNet50: 0.8671875
- DenseNet121: 0.84375

Và sau khi tính toán trọng số, độ chính xác và ensemble 3 mô hình lại kết quả cuối cùng thu được là 0.8571428571428571, tương đương xấp xỉ 86%. Với độ chính xác ở mức hơn 86%, mô hình nhóm thực hiện được xem ở mức chấp nhận được, nhưng sẽ cần được cải tiến trong tương lai.

Từ kết quả thực nghiệm có thể đánh giá, tuy mô hình DenseNet121 được phát triển sau nhưng lại đạt kết quả thấp hơn 2 mô hình còn lại. Ta có thể thấy việc dữ liệu phù hợp với mô hình huấn luyện là vô cùng quan trọng. Việc tổng hợp ba mô hình sẽ cho ta một cái nhìn tổng quát hơn và độ chính xác lại không được đảm bảo.

Chương 6 Kết luận và tương lai phát triển

Nhìn vào thực tế, bài toán phân lớp ảnh đóng vai trò khá quan trọng trong lĩnh vực Khoa học máy tính và Thị giác máy tính, giúp chúng ta giải quyết được nhiều vấn đề trong cuộc sống và nâng cao mức sống. Trong y khoa, bài toán phân lớp ảnh dựa vào ảnh y khoa rút ngắn thời gian chẩn đoán và nâng cao hiệu quả khám chữa bệnh cho bệnh nhân. Nhận thấy được ý nghĩa thực tiễn và tiềm năng phát triển, nhóm chúng tôi cũng muốn đóng góp chút sức của bản thân mình vào chủ đề này.

Với ý tưởng nhận được từ giáo viên hướng dẫn PGS.TS Lê Hoàng Thái, chúng tôi đã lên ý tưởng và kế hoạch để tạo ra một phương pháp, một mô hình phân lớp ảnh X-quang phổi từ đặc trưng quan trọng nhất là độ bất thường của đậm độ. Điểm độc đáo từ phương pháp của nhóm chúng tôi ở việc kết hợp 3 mô hình CNN bao gồm VGG-16, ResNet50 và DenseNet121 lại với nhau bằng phương pháp stacking. Do kết hợp độ chính xác của 3 mô hình sau khi qua tinh chỉnh, kết quả thu lại được khá cao và chấp nhận được ở mức độ là một phương pháp mới.

Với mục tiêu nâng cao độ chính xác và rút ngắn thời gian của mô hình, chúng tôi mong từ ý tưởng này có thể phát triển mô hình hoàn thiện nhất có thể bằng công sức bỏ ra. Đối với nhiều bộ dữ liệu khác nhau, kết quả thu lại có thể trên lệch nhưng sẽ không vượt mức chấp nhận được. Từ phương pháp nhóm đề xuất, chúng tôi mong có thể góp một chút sức mình vào công tác chẩn đoán y khoa và thêm phần sinh động cũng như thú vị cho lĩnh Thị giác Máy tính và Khoa học Máy tính.

Trong tương lai, phương pháp đã được đề xuất có thể phát triển và hoàn thiện hơn nhờ vào việc các mô hình học máy và học sâu tiên tiến hơn ra đời. Thật không bất ngờ nếu có một ngày sẽ có một bộ mô hình có thể đạt độ chính xác và thay thế bác sĩ chuyên gia ở khâu chẩn đoán.

Cuối khóa luận, chúng tôi xin cảm ơn mọi người đã dành thời gian quý báu để đón nhận công trình nghiên cứu của nhóm. Mong tương lai, lĩnh vực và bài toán này sẽ trở nên phổ biến, phát triển và đóng góp to lớn cho cuộc sống và sức khỏe con người.

Tài liệu tham khảo

Link các trang web tham khảo:

<https://www.stdio.vn/ai-ml/bai-toan-phan-lop-trong-machine-learning-classification-in-machine-learning-515OIh>

<https://www.geeksforgeeks.org/understanding-googlenet-model-cnn-architecture/>

<https://viblo.asia/p/gioi-thieu-mang-resnet-vyDZOa7R5wj>

<https://viso.ai/deep-learning/resnet-residual-neural-network/>

<https://svcuong.github.io/post/ensemble-learning/>

<https://www.pluralsight.com/guides/introduction-to-densenet-with-tensorflow>

<https://dlapplications.github.io/2018-07-06->

[CNN/#:~:text=Densenet\(Dense%20connected%20convolutional%20network,block%20v%C3%A0%20c%C3%A1c%20transition%20layers.](https://dlapplications.github.io/2018-07-06-CNN/#:~:text=Densenet(Dense%20connected%20convolutional%20network,block%20v%C3%A0%20c%C3%A1c%20transition%20layers.)

<https://viblo.asia/p/fine-tuning-pre-trained-model-trong-pytorch-va-ap-dung-va-visual-saliency-prediction-4P856LY1ZY3>

<https://viblo.asia/p/tang-cuong-du-lieu-trong-deep-learning-oOVIYe4nl8W>

[VGG-16 | CNN model - GeeksforGeeks](#)

Các bài báo tham khảo:

Densely Connected Convolutional Networks - Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger

A survey of transfer learning - Karl Weiss, Taghi M. Khoshgoftaar & DingDing Wang

Very Deep Convolutional Networks for Large-Scale Image Recognition - Karen Simonyan, Andrew Zisserman