

# LỜI CAM ĐOAN

Tôi xin cam đoan: Luận văn thạc sỹ chuyên ngành Khoa học máy tính, tên đề tài “Nghiên cứu hỗ trợ chuẩn đoán bệnh lao dựa vào học máy” là công trình nghiên cứu, tìm hiểu và trình bày do tôi thực hiện dưới sự hướng dẫn khoa học của **PGS.TS. Đỗ Năng Toàn**, Viện Công nghệ Thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam.

Kết quả tìm hiểu, nghiên cứu trong luận văn là hoàn toàn trung thực, không vi phạm bất cứ điều gì trong luật sở hữu trí tuệ và pháp luật Việt Nam. Nếu sai, tôi hoàn toàn chịu trách nhiệm trước pháp luật.

Tất cả các tài liệu, bài báo, khóa luận, công cụ phần mềm của các tác giả khác được sử dụng lại trong luận văn này đều được chỉ dẫn tường minh về tác giả và đều có trong danh mục tài liệu tham khảo.

*Thái Nguyên, ngày 30 tháng 6 năm 2022.*

Tác giả luận văn  
**Nguyễn Hữu Khánh**

# LỜI CẢM ƠN

Tác giả xin chân thành cảm ơn **PGS.TS. Đỗ Năng Toàn**, Viện Công nghệ Thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam, là giáo viên hướng dẫn khoa học đã hướng dẫn tác giả hoàn thành luận văn này, xin được cảm ơn các thầy, cô giáo trường Đại học công nghệ thông tin và truyền thông nơi tác giả theo học và hoàn thành chương trình cao học đã nhiệt tình giảng dạy và giúp đỡ.

Xin cảm ơn Trung tâm Đào tạo Từ xa - Đại học Thái Nguyên nơi tác giả công tác đã tạo mọi điều kiện thuận lợi để tác giả có thời gian, tâm trí để hoàn thành nhiệm vụ nghiên cứu và chương trình học tập.

Và cuối cùng xin cảm ơn gia đình, bạn bè, đồng nghiệp đã động viên, giúp đỡ tác giả trong suốt thời gian học tập, nghiên cứu và hoàn thành luận văn này.

Xin chân thành cảm ơn.

*Thái Nguyên, ngày 30 tháng 06 năm 2022*

Tác giả luận văn  
**Nguyễn Hữu Khánh**

# Mục lục

Lời cam đoan.....	i
Lời cảm ơn .....	ii
Mở đầu.....	1
Chương 1. Khái quát về CNN và bài toán chuẩn đoán bệnh lao..	3
1.1. Khái quát về CNN .....	3
1.1.1. Giới thiệu.....	3
1.1.2. Kiến trúc mạng CNN .....	4
1.1.3. Ứng dụng CNN trong phân loại ảnh.....	9
Chương 2. Một số mô hình hỗ trợ chuẩn đoán.....	11
Chương 3. Chương trình thử nghiệm.....	12
Tài liệu tham khảo.....	12

# MỞ ĐẦU

Theo báo cáo của Tổ chức Y tế thế giới (TCYTTG - WHO Report 2020 - Global Tuberculosis Control) [1], mặc dù đã đạt được một số thành tựu đáng kể trong công tác chống lao trong thời gian qua, bệnh lao vẫn đang tiếp tục là một trong các vấn đề sức khỏe cộng đồng chính trên toàn cầu. TCYTTG ước tính năm 2019 trên toàn cầu có khoảng 10 triệu người hiện mắc lao, một con số đã giảm rất chậm trong những năm gần đây; 8,2% trong số mắc lao có đồng nhiễm HIV. Bệnh lao là nguyên nhân gây tử vong đứng hàng thứ hai trong các bệnh nhiễm trùng với khoảng 1,2 triệu người tử vong do lao và khoảng 208.000 người chết do lao trong số những người nhiễm HIV. Số tử vong này làm cho lao là một trong các bệnh gây tử vong hàng đầu ở nữ giới. WHO đã công bố kết quả của mô hình đánh giá tác động ngắn hạn của đại dịch Covid-19 lên số ca tử vong do lao trong năm 2020. Kết quả cho thấy rằng tử vong do lao có thể tăng đáng kể trong năm 2020 và sẽ ảnh hưởng đến nhóm bệnh nhân lao dễ bị tổn thương nhất, tăng khoảng 200.000 – 400.000 ca tử vong, nếu như các dịch vụ chẩn đoán và phát hiện bệnh nhân trên toàn cầu bị ngưng trệ và giảm từ 25 – 50% trong khoảng 3 tháng. Con số tử vong sẽ tương ứng với mức tử vong toàn cầu do lao vào năm 2015, một bước lùi nghiêm trọng trong quá trình hướng tới mục tiêu của Hội nghị Cấp Cao Liên Hợp Quốc về Lao và Chiến lược thanh toán bệnh lao của WHO.

Về Việt nam, hiện chúng ta vẫn là nước có gánh nặng bệnh lao cao, đứng thứ 11 trong 30 nước có số người bệnh lao cao nhất trên toàn cầu, đồng thời đứng thứ 11 trong số 30 nước có gánh nặng bệnh lao kháng đa thuốc cao nhất thế giới (báo cáo WHO 2020). Bệnh lao vẫn là một trong những bệnh

truyền nhiễm phổ biến ở Việt Nam. Hàng năm, ước tính có 17.000 trường hợp tử vong do lao tại Việt Nam, cao hơn gấp hai lần so với con số tử vong do tai nạn giao thông. Mỗi năm ước tính có 180.000 người có bệnh lao hoạt động; 5.000 trường hợp trong số đó được xác định nhiễm lao kháng đa thuốc. Chẩn đoán bệnh lao không thật sự khó trong đa số các trường hợp. Điều đáng chú ý là làm sao chẩn đoán sớm và chẩn đoán đúng để khởi động điều trị sớm nhằm giảm các tổn thương cũng như biến chứng của lao gây ra. Để làm được điều trên, việc ứng dụng công nghệ thông tin vào quá trình chuẩn đoán là thực sự cần thiết, đặc biệt là áp dụng những tiến bộ của học máy để xây dựng lên hệ thống hỗ trợ chuẩn đoán bệnh lao.

**Đối tượng nghiên cứu:** Ảnh X-quang lồng ngực trong y tế thu nhận bởi các máy chiếu, chụp chuyên dụng.

**Phạm vi nghiên cứu:** Ảnh đa mức xám chụp phổi thẳng thường quy ( tư thế sau - trước), chụp phổi nghiêng thường quy và chụp đỉnh phổi tư thế ưỡn ngực.

**Những nội dung nghiên cứu chính:** Dự kiến nội dung báo cáo của luận văn gồm: phần mở đầu, 3 chương chính, phần kết luận, tài liệu tham khảo, phụ lục. Bố cục được trình bày như sau:

**Phần mở đầu:** Nêu lý do chọn đề tài và hướng nghiên cứu chính

**Chương 1:** Khái quát về CNN và bài toán chuẩn đoán bệnh lao.

**Chương 2:** Một số mô hình hỗ trợ chuẩn đoán.

**Chương 3:** Chương trình thử nghiệm.

Mặc dù đã có cố gắng nỗ lực, song luận văn không tránh khỏi những thiếu sót do năng lực và thời gian hạn chế. Em chân thành mong muốn lắng nghe những đóng góp, góp ý của thầy, cô, bạn bè, đồng nghiệp để luận văn được cải thiện tốt hơn.

Em xin chân thành cảm ơn.

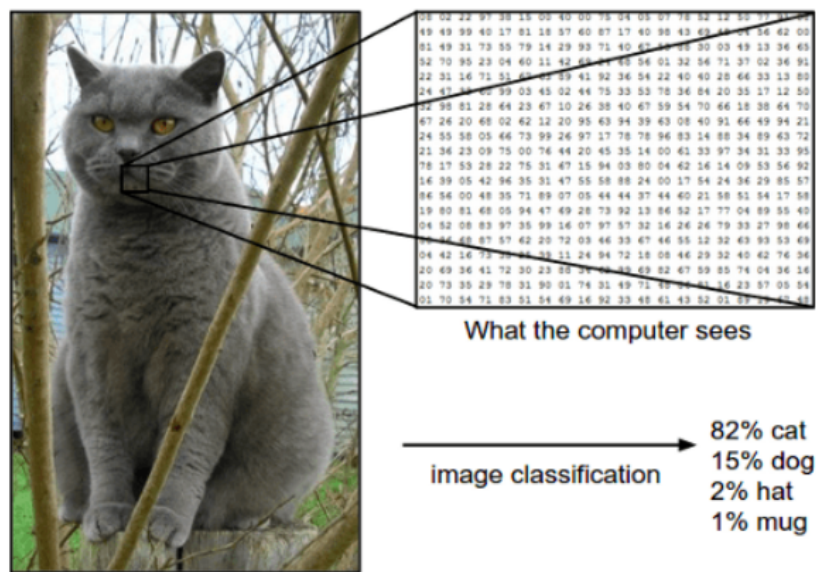
# CHƯƠNG 1

## Khái quát về CNN và bài toán chuẩn đoán bệnh lao.

### 1.1. Khái quát về CNN

#### 1.1.1. Giới thiệu

Tương tự như việc trẻ em học cách nhận diện đối tượng, chúng ta cần cho thuật toán học rất nhiều hình ảnh trước khi nó có thể đưa ra phân loại cho hình ảnh đầu vào mà nó chưa từng thấy [2].



Hình 1.1: Cách máy tính "nhìn" một hình.

Máy tính “nhìn” theo cách khác con người. Trong thế giới máy tính chỉ có những con số. Mỗi hình ảnh có thể được biểu diễn dưới dạng mảng 2 chiều những con số được gọi là các pixel.

Mặc dù máy tính nhìn nhận theo cách khác con người, chúng ta vẫn có thể dạy máy tính nhận diện các mẫu như con người. Điều quan trọng là chúng ta cần nghĩ về hình ảnh theo một cách khác đi.

Để dạy thuật toán nhận diện đối tượng trong hình ảnh, ta sử dụng một loại mạng ANN, đó là CNN. Tên của nó được dựa trên phép tính quan trọng được sử dụng trong mạng- tích chập.

Mạng CNN lấy cảm hứng từ não người. Nghiên cứu trong những thập niên

1950 và 1960 của D.H Hubel và T.N Wiesel trên não của động vật đã đề xuất một mô hình mới cho việc cách mà động vật nhìn nhận thế giới. Trong báo cáo, hai ông đã diễn tả 2 loại tế bào nơ-ron trong não và cách hoạt động khác nhau: tế bào đơn giản (simple cell – S cell) và tế bào phức tạp (complex cell – C cell).

Các tế bào đơn giản được kích hoạt khi nhận diện các hình dáng đơn giản như đường nằm trong một khu vực cố định và một góc cạnh của nó. Các tế bào phức tạp có vùng tiếp nhận lớn hơn và đầu ra của nó không nhạy cảm với những vị trí cố định trong vùng.

Trong thị giác, vùng tiếp nhận của một nơ-ron tương ứng với một vùng trên võng mạc nơi mà sẽ kích hoạt nơ-ron tương ứng.

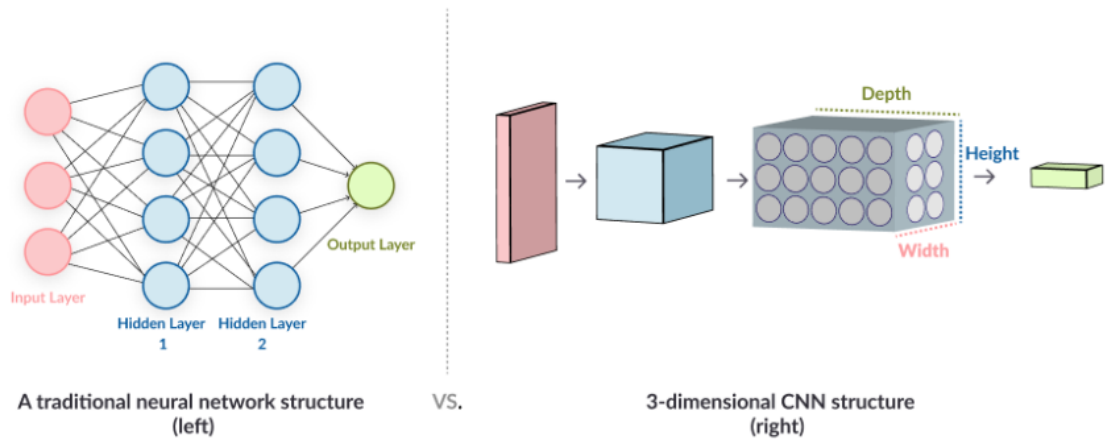
Năm 1980, Fukushima đề xuất mô hình mạng nơ-ron có cấp bậc gọi là neocognitron. Mô hình này dựa trên khái niệm về S cell và C cell. Mạng neocognitron có thể nhận diện mẫu dựa trên việc học hình dáng của đối tượng.

Sau đó vào năm 1998, mạng CNN được giới thiệu bởi Bengio, Le Cun, Bottou và Haffner. Mô hình đầu tiên của họ được gọi tên là LeNet-5. Mô hình này có thể nhận diện chữ số viết tay.

### *1.1.2. Kiến trúc mạng CNN*

Mạng CNN có kiến trúc khác với Mạng Nơ-ron thông thường. Mạng ANN bình thường chuyển đổi đầu vào thông qua hàng loạt các tầng ẩn. Mỗi tầng là một tập các nơ-ron và các tầng được liên kết đầy đủ với các nơ-ron ở tầng trước đó. Và ở tầng cuối cùng sẽ là tầng kết quả đại diện cho dự đoán của mạng.

Đầu tiên, mạng CNN được chia thành 3 chiều: rộng, cao, và sâu. Kế đến, các nơ-ron trong mạng không liên kết hoàn toàn với toàn bộ nơ-ron lớp sau mà chỉ liên kết tới một vùng nhỏ. Cuối cùng, một tầng đầu ra được tối giản thành véc-tơ của giá trị xác suất (hình 1.2).

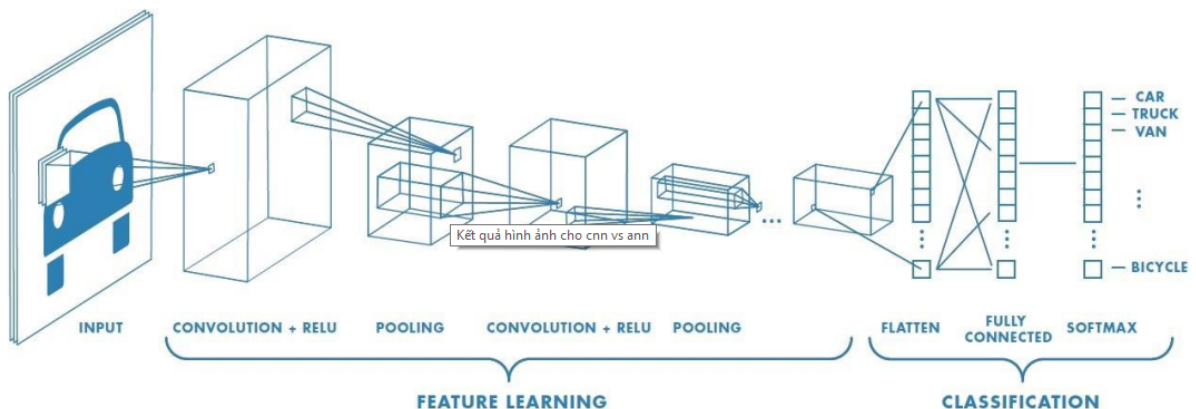


Hình 1.2: Mạng Neuron thông thường (trái) và CNN (phải).

Mạng CNN gồm hai thành phần:

**Phần tầng ẩn hay phần rút trích đặc trưng:** trong phần này, mạng sẽ tiến hành tính toán hàng loạt phép tích chập và phép hợp nhất (pooling) để phát hiện các đặc trưng. Ví dụ: nếu ta có hình ảnh con ngựa vằn, thì trong phần này mạng sẽ nhận diện các sọc vằn, hai tai, và bốn chân của nó.

**Phần phân lớp:** tại phần này, một lớp với các liên kết đầy đủ sẽ đóng vai trò như một bộ phân lớp các đặc trưng đã rút trích được trước đó. Tầng này sẽ đưa ra xác suất của một đối tượng trong hình 1.3.



Hình 1.3: Kiến trúc mạng CNN.

### 1.1.2.1. Trích rút đặc trưng

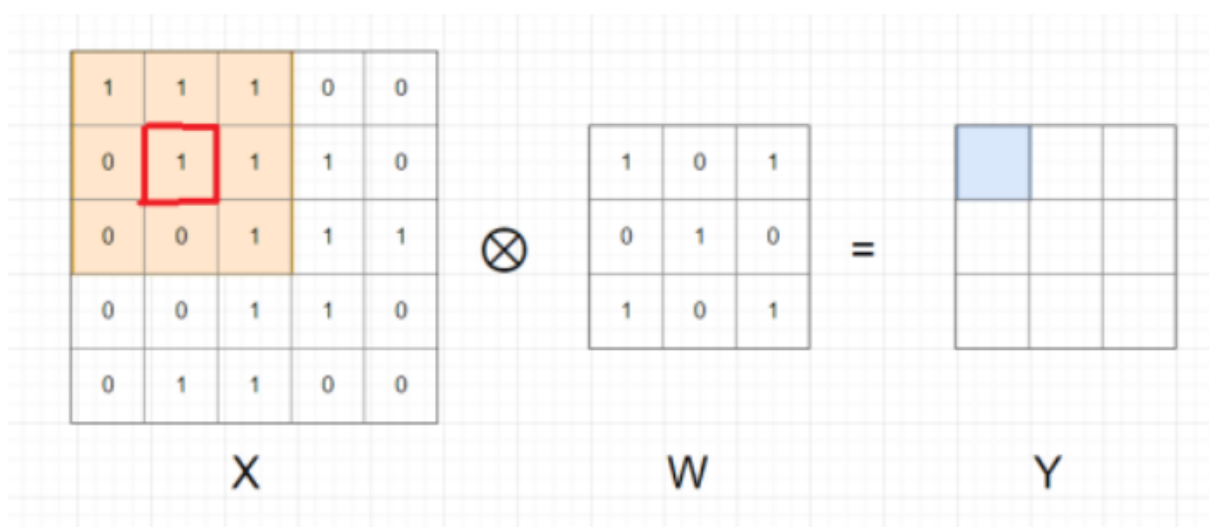
**1.1.2.1.1. Lớp tích chập** Tích chập là một khối quan trọng trong CNN. Thuật ngữ tích chập được dựa trên một phép hợp nhất toán học của hai hàm tạo



thành hàm thứ ba. Phép toán này kết hợp hai tập thông tin khác nhau.

Trong trường hợp CNN, tích chập được thực hiện trên giá trị đầu vào của dữ liệu và kernel/filter (thuật ngữ này được sử dụng khác nhau tùy tình huống) để tạo ra một bản đồ đặc trưng (feature map).

Ta thực hiện phép tích chập bằng cách trượt kernel/filter theo dữ liệu đầu vào. Tại mỗi vị trí, ta tiến hành phép nhân ma trận và tính tổng các giá trị để đưa vào bản đồ đặc trưng. Thao tác này đã được minh họa cụ thể trong hình 1.4



Hình 1.4: Minh họa phép tích chập.

Trong thực tế, tích chập được thực hiện hiện trên không gian 3 chiều. Vì mỗi hình ảnh được biểu diễn dưới dạng 3 chiều: rộng, cao, và sâu. Chiều sâu ở đây chính là giá trị màu sắc của hình (RGB).

Ta thực hiện phép tích chập trên đầu vào nhiều lần khác nhau. Mỗi lần sử dụng một kernel/filter khác nhau. Kết quả ta sẽ thu được những bản đồ đặc trưng khác nhau. Cuối cùng, ta kết hợp toàn bộ bản đồ đặc trưng này thành kết quả cuối cùng của tầng tích chập.

**1.1.2.1.2. Lớp ReLU** Tương tự như mạng nơ-ron thông thường, ta sử dụng một hàm kích hoạt (activate function) để có đầu ra dưới dạng phi tuyến. Trong trường hợp CNN, đầu ra của phép tích chập sẽ đi qua hàm kích hoạt nào đó ví dụ như hàm tinh chỉnh các đơn vị tuyến tính (Rectified linear units

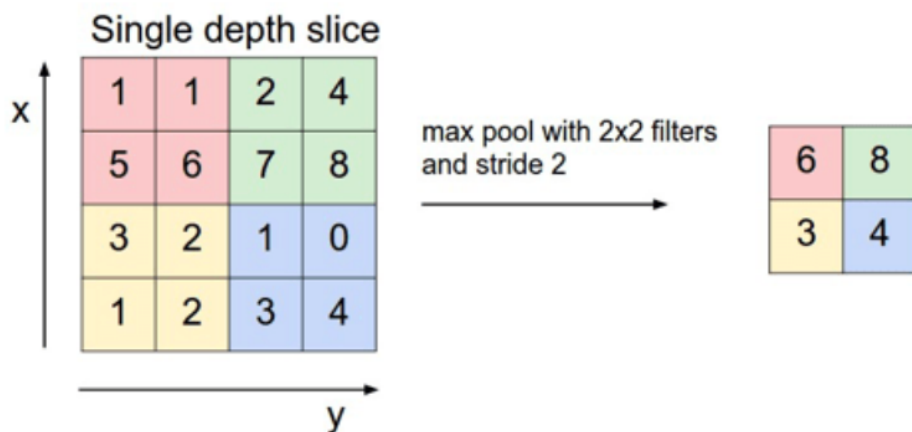
- ReLU).

Trong quá trình trượt kernel/filter trên dữ liệu đầu vào, ta sẽ quy định một bước nhảy (stride) với mỗi lần di chuyển. Thông thường ta lựa chọn thường chọn bước nhảy là 1. Nếu kích thước bước nhảy tăng, kernel/filter sẽ có ít ô trùng lặp.

Bởi vì kích thước đầu ra luôn nhỏ hơn đầu vào nên ta cần một phép xử lý đầu vào để đầu ra không bị co giãn. Đơn giản ta chỉ cần thêm một lề nhỏ vào đầu vào. Một lề (padding) với giá trị 0 sẽ được thêm vào xung quanh đầu vào trước khi thực hiện phép tích chập.

**1.1.2.1.3. Lớp pooling** Thông thường, sau mỗi tầng tích chập, ta sẽ cho kết quả đi qua một tầng hợp nhất (pooling layer). Mục đích của tầng này là để nhanh chóng giảm số chiều. Việc này giúp giảm thời gian học và hạn chế việc overfitting.

Một phép hợp nhất đơn giản thường được dùng đó là max pooling, phép này lấy giá trị lớn nhất của một vùng để đại diện cho vùng đó. Kích thước của vùng sẽ được xác định trước để giảm kích thước của bản đồ đặc trưng nhanh chóng nhưng vẫn giữ được thông tin cần thiết (hình 1.5).



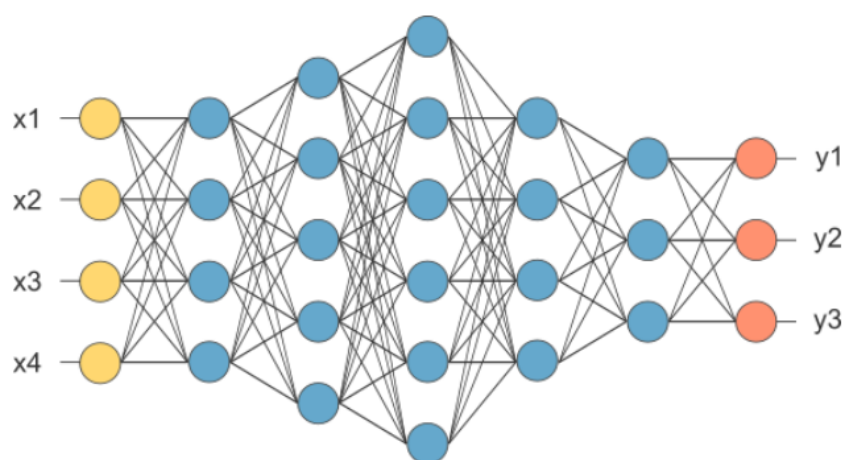
Hình 1.5: Max pooling kích thước 2x2

Như vậy, khi thiết kế phần rút trích đặc trưng của mạng CNN, ta cần chú ý đến 4 siêu tham số quan trọng là: Kích thước kernel/filter, Số lượng kernel/filter, Kích thước bước nhảy (stride), Kích thước lề (padding).

### 1.1.2.2. Phân lớp

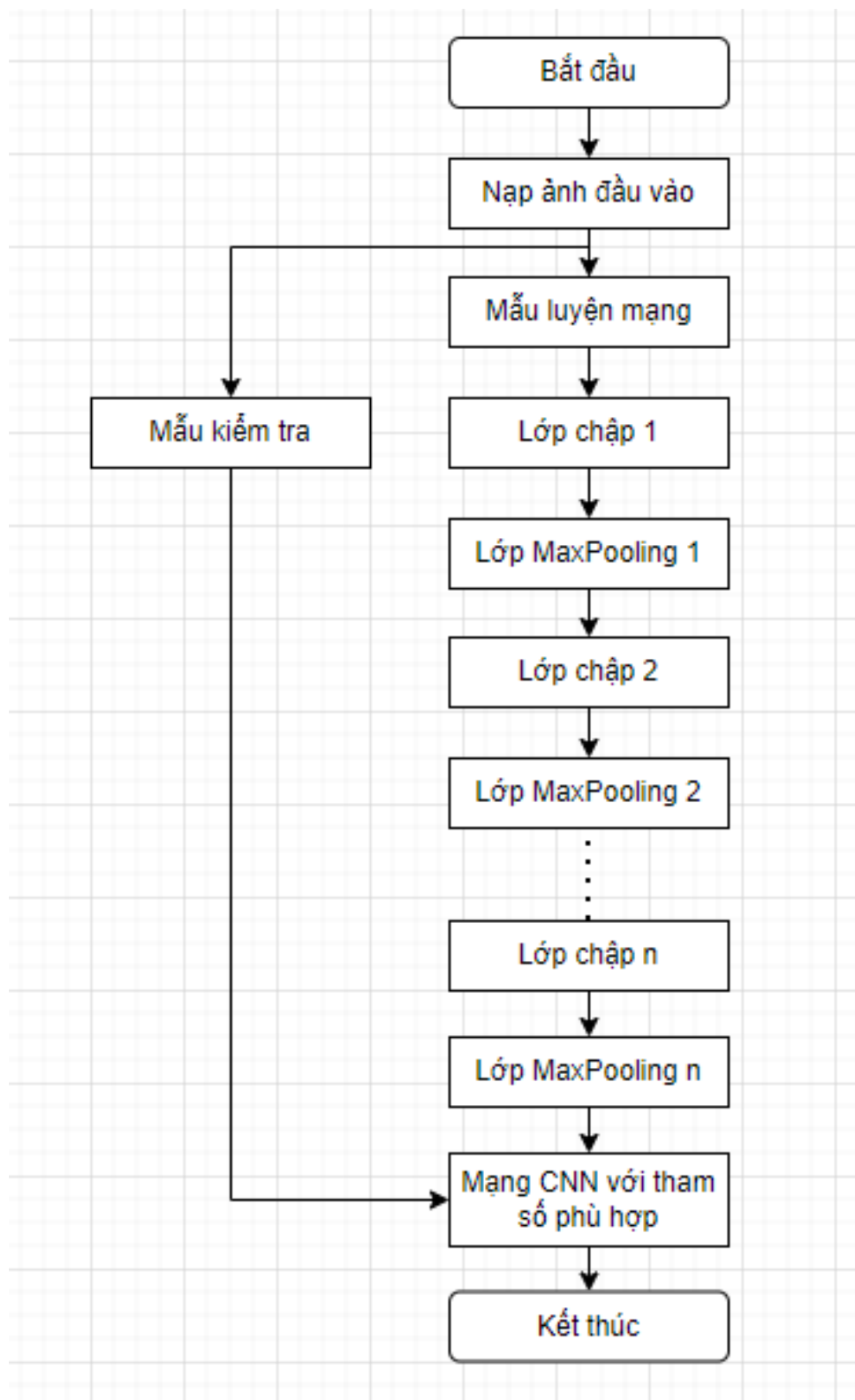
Trong phần phân lớp, ta sử dụng một vài tầng với kết nối đầy đủ để xử lý kết quả của phần tích chập. Vì đầu vào của mạng liên kết đầy đủ là 1 chiều, ta cần làm phẳng đầu vào trước khi phân lớp. Tầng cuối cùng trong mạng CNN là một tầng liên kết đầy đủ, phần này hoạt động tương tự như mạng nơ-ron thông thường.

Kết quả thu được cuối cùng cũng sẽ là một véc-tơ với các giá trị xác suất cho việc dự đoán như mạng nơ-ron thông thường.



Hình 1.6: Lớp kết nối đầy đủ

### 1.1.2.3. Ứng dụng CNN trong phân loại ảnh



Hình 1.7: Max pooling kích thước  $2 \times 2$

Các bước để thực hiện phân loại hình ảnh dựa trên mạng CNN được mô tả trong Hình 1.7. Đầu tiên, kho dữ liệu ảnh đầu vào được nạp. Ảnh này

được chia làm hai phần, một phần dành cho luyện mạng và một phần cho kiểm tra. Trước tiên, ta phải lựa chọn cấu trúc mạng CNN bao gồm số lượng lớp ẩn, các tham số trong mỗi lớp ẩn như kích thước trường tiếp nhận cục bộ, stride, padding. Ảnh luyện mạng sau đó được đưa vào lớp chập 1 để thực hiện tích chập trên ảnh và thực hiện hàm ReLU. Sau đó, kết quả được đưa đến quá trình thực hiện pooling với tham số pooling size phù hợp để giảm kích cỡ ảnh. Ảnh sẽ tiếp tục được đưa thêm qua các lớp tích chập nữa cho đến khi đạt được kết quả mong muốn. Kết quả này được dàn phẳng và đưa vào lớp kết nối đầy đủ. Cuối cùng là quá trình thực hiện các activation function và phân loại ảnh. Quá trình luyện mạng sẽ kết thúc sau khi tổng sai số nhỏ hơn một ngưỡng cho phép hoặc sau một số thế hệ cho trước (điều kiện hội tụ). Kết thúc của quá trình luyện mạng là cấu trúc mạng CNN với các tham số phù hợp. Để kiểm tra, các mẫu ảnh kiểm tra được đưa qua mạng CNN rồi thực hiện đánh giá sai số.

### *1.1.3. Xây dựng mạng CNN cho phân loại ảnh*

## CHƯƠNG 2

Một số mô hình hỗ trợ chuẩn đoán.

## CHƯƠNG 3

### Chương trình thử nghiệm.

# Tài liệu tham khảo

- [1] WHO *Global tuberculosis report 2020*, báo cáo tại <https://www.who.int/publications/i/item/9789240013131>, 2020
- [2] Nguyễn Thanh Tuấn *Bài 3: Neural network*, <https://nttuan8.com/bai-3-neural-network>, 2019