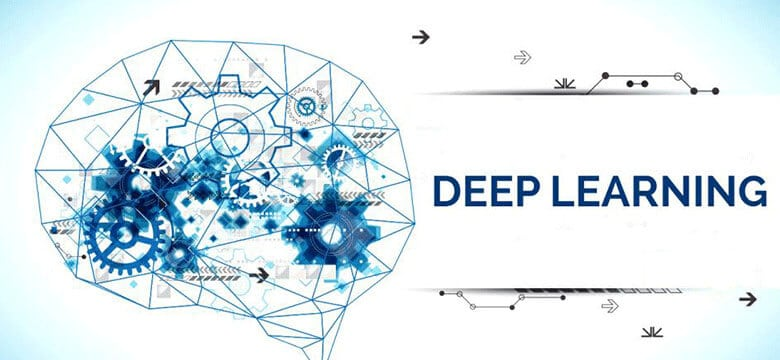
# NGHIÊN CỨU MỘT SỐ MÔ HÌNH DEEP LEANING

Deep Learning là một chức năng của trí tuệ nhân tạo (AI), bắt chước cách bộ não con người hoạt động để xử lí dữ liệu, tạo ra các mẫu để sử dụng cho việc đưa ra quyết định.

## 1.1 Tổng quan về Deeplearning

### 1.1.1 Deeplearning là gì?

Deep Learning (học sâu) có thể được xem là một lĩnh vực con của Machine Learning (học máy) – ở đó các máy tính sẽ học và cải thiện chính nó thông qua các thuật toán. Deep Learning được xây dựng dựa trên các khái niệm phức tạp hơn rất nhiều, chủ yếu hoạt động với các mạng nơ-ron nhân tạo để bắt chước khả năng tư duy và suy nghĩ của bộ não con người.

 Tổng quan về Deeplearning

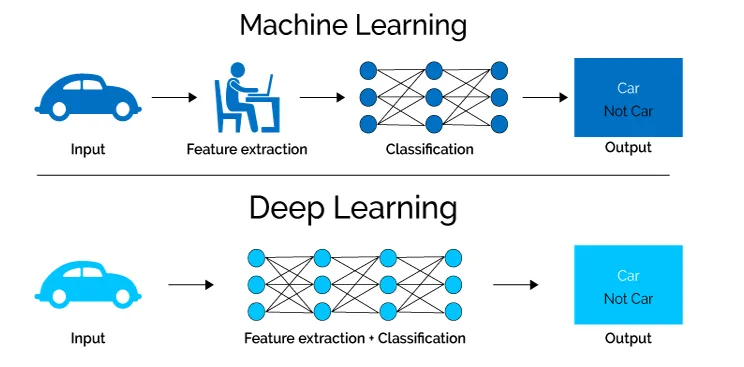
Mạng nơ-ron nhân tạo chính là động lực chính để phát triển Deep Learning. Các mạng nơ-ron sâu (DNN) bao gồm nhiều lớp nơ-ron khác nhau, có khả năng thực hiện các tính toán có độ phức tạp rất cao. Deep Learning hiện đang phát triển rất nhanh và được xem là một trong những bước đột phá lớn nhất trong Machine Learning.

### 1.1.2 Deeplearning hoạt động như thế nào?

Deep Learning là một phương pháp của Machine Learning. Mạng nơ-ron nhân tạo trong Deep Learning được xây dựng để mô phỏng khả năng tư duy của bộ não con người.

Một mạng nơ-ron bao gồm nhiều lớp (layer) khác nhau, số lượng layer càng nhiều thì mạng sẽ càng “sâu”. Trong mỗi layer là các nút mạng (node) và được liên kết với những lớp liền kề khác. Mỗi kết nối giữa các node sẽ có một trọng số tương ứng, trọng số càng cao thì ảnh hưởng của kết nối này đến mạng nơ-ron càng lớn.

Mỗi nơ-ron sẽ có một hàm kích hoạt, về cơ bản thì có nhiệm vụ “chuẩn hoá” đầu ra từ nơ-ron này. Dữ liệu được người dùng đưa vào mạng nơ-ron sẽ đi qua tất cả layer và trả về kết quả ở layer cuối cùng, gọi là output layer.

 Quy trình hoạt động của Deeplearning

Trong quá trình huấn luyện mô hình mạng nơ-ron, các trọng số sẽ được thay đổi và nhiệm vụ của mô hình là tìm ra bộ giá trị của trọng số sao cho phán đoán là tốt nhất.

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

### 1.1.3 Ưu nhược điểm của Deeplearning

1.1.3.1. Ưu điểm:

Deeplearning là một bước ngoặt to lớn trong lĩnh vực trí tuệ nhân tạo, cho phép khác nhà khoa học dữ liệu xây dựng nhiều mô hình có độ chính xác rất cao trong lĩnh vực nhận dạng ảnh, xử lý ngôn ngữ tự nhiên, xử lý giọng nói,… Một số ưu điểm vượt trội của Deeplearning gồm có:

#### - Kiến trúc mạng nơ-ron linh hoạt, có thể dễ dàng thay đổi để phù hợp với nhiều vấn đề khác nhau.

#### - Có khả năng giải quyết nhiều bài toán phức tạp với độ chính xác rất cao.

#### - Tính tự động hoá cao, có khả năng tự điều chỉnh và tự tối ưu.

#### - Có khả năng thực hiện tính toán song song, hiệu năng tốt, xử lý được lượng dữ liệu lớn.

1.1.3.2. Nhược điểm:

Bên cạnh những ưu điểm, mặt khác, hiện nay Deeplearning vẫn còn nhiều khó khăn và hạn chế, chẳng hạn như:

- Cần có khối lượng dữ liệu rất lớn để tận dụng tối đa khả năng của Deeplearning.

- Chi phí tính toán cao vì phải xử lý nhiều mô hình phức tạp.

- Chưa có nền tảng lý thuyết mạnh mẽ để lựa chọn các công cụ tối ưu cho Deeplearning.

## 1.2 Một số mô hình Deeplearning

### 1.2.1 Mạng nơ-ron cổ điển

Kiến trúc cổ điển của mạng nơ-ron là mạng kết nối đầy đủ, thường được xác định bằng các perceptron đa lớp. (Perceptron là một thuật toán đơn giản, cho phép tìm một ranh giới siêu phẳng cho các bài toán phân lớp nhị phân). Mạng nơ-ron cổ điển được thiết kế bởi Fran Rosenblatt vào năm 1958, chủ yếu được sử dụng cho các bài toán phân lớp nhị phân. Có hai loại hàm thường được sử dụng trong mô hình này là:

- Hàm tuyến tính.

- Hàm phi tuyến: gồm có hàm sigmoid, hàm tanh và hàm ReLU (Rectified Linear Unit).

Kiến trúc mạng nơ-ron cổ điển tương đối đơn giản, phù hợp nhất với các bộ dữ liệu có dạng bảng hoặc những bài toán phân loại, hồi quy có đầu vào là giá trị thực.

### 1.2.2 Mạng nơ-ron tích chập (CNN)

Mạng nơ-ron tích chập (Convolutional Neural Network – CNN) là một kiến trúc Neural Network nhân tạo nâng cao, được xây dựng để giải quyết các bài toán phức tạp, đặc biệt là liên quan đến xử lý hình ảnh.

Tích chập là một khái niệm trong xử lý tín hiệu số nhằm biến đổi thông tin đầu vào qua một phép tích chập với bộ lọc, nhằm trả về đầu ra là một tín hiệu mới. Tín hiệu này sẽ giảm bớt những đặc trưng mà bộ lọc không quan tâm, giữ lại những đặc trưng chính và quan trọng nhất.

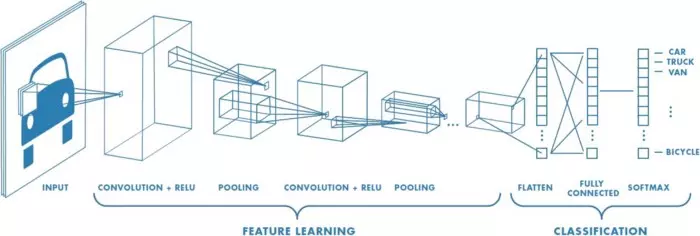
Bên cạnh input layer và output layer, mô hình CNN còn có thêm một sampling layer để giới hạn số lượng nơ-ron tham gia vào các layer tương ứng. Việc xây dựng mô hình trải qua ba giai đoạn chính:

- Quá trình tích chập (convolution): Thông qua các tích chập giữa ma trận đầu vào với bộ lọc để tạo thành các đơn vị trong một tầng mới. Quá trình này có thể diễn ra liên tục ở phần đầu của mạng và thường sử dụng kèm với hàm kích hoạt ReLU. Mục tiêu của tầng này là trích suất đặc trưng hai chiều.

- Quá trình tổng hợp (max pooling): Giảm kích thước khối ma trận đầu vào thông qua việc tìm ra 1 giá trị đại diện cho mỗi một vùng không gian mà bộ lọc đi qua sẽ không làm thay đổi các đường nét chính của bức ảnh nhưng lại giảm được kích thước của ảnh.

- Quá trình kết nối hoàn toàn (fully connected): Sau khi đã giảm kích thước đến một mức độ hợp lý, ma trận cần được trải phẳng (flatten) thành một vector và sử dụng các kết nối hoàn toàn giữa các tầng. Tầng kết nối hoàn toàn cuối cùng (fully connected layer) sẽ có số lượng đơn vị bằng với số lớp.

Dựa vào những đặc điểm của mình, các ứng dụng phổ biến nhất của mạng CNN gồm có: Nhận diện, phân tích và phân khúc hình ảnh, phân tích video, xử lý ngôn ngữ tự nhiên,…

 Mạng nơ-ron tích chập (CNN)

### 1.2.3 Mạng nơ-ron hồi quy (RNN)

Recurrent Neural Network (RNN) là một thuật toán nổi tiếng trong lĩnh vực xử lý ngôn ngữ tự nhiên. Trong các mô hình mạng nơ-ron truyền thống, đầu vào và đầu ra độc lập với nhau, tuy nhiên RNN thực hiện cùng một tác vụ cho tất cả phần tử của một chuỗi với đầu ra phụ thuộc vào cả các phép tính trước đó. Vì vậy mạng RNN có khả năng nhớ các thông tin được tính toán trước đó.

Có hai thiết kế chính của RNN:

- LSTM (Long Short-Term Memory): Được dùng để dự đoán dữ liệu dạng chuỗi thời gian, có khả năng bỏ đi hoặc thêm các thông tin cần thiết, được điều chỉnh bởi các nhóm được gọi là cổng (gate): Input, Output và Forget.

- Gated RNN: Cũng là một thiết kế phổ biến trong lĩnh vực dự đoán dữ liệu của chuỗi thời gian, có hai cổng là Update và Reset.

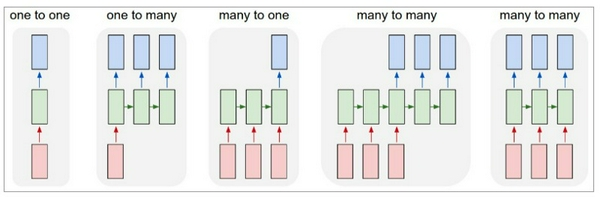
Các dạng bài toán RNN:

- One to one: Chỉ có một input kết nối với một output duy nhất, chẳng hạn như các bài toán phân loại hình ảnh.

- One to many: Một input liên kết với nhiều chuỗi output, phổ biến là các bài toán đặt caption cho ảnh.

- Many to One: Nhiều input nhưng chỉ có output, ví dụ phổ biến là bài toán phân loại cảm xúc.

- Many to many: Nhiều input và nhiều output, chẳng hạn như phân loại video.

 Mạng nơ-ron hồi quy (RNN)

### 1.2.4 Mạng sinh đối nghịch (GAN)

Generative Adversarial Networks (GAN) là lớp mô hình có mục tiêu tạo ra dữ liệu giả giống với thật, tên của mạng được dựa trên kiến trúc gồm hai mạng có mục tiêu đối nghịch nhau: Generator và Discriminator. Trong đó Generator học cách sinh dữ liệu giả để lừa mô hình Discriminator, còn Discriminator lại học cách phân biệt giữa dữ liệu giả và dữ liệu thật. Thông qua quá trình huấn luyện thì cả hai mô hình này đều cùng cải thiện được khả năng của mình.

Một số ứng dụng phổ biến của GAN là: Tạo khuôn mặt người, thay đổi độ tuổi khuôn mặt, sinh ảnh vật thể, tạo nhân vật hoạt hình,…