# **What is a convolutional neural network (CNN)?**

A convolutional neural network (CNN) is a category of [machine learning](https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML) model. Specifically, it is a type of [deep learning](https://www.techtarget.com/searchenterpriseai/definition/deep-learning-deep-neural-network) algorithm that is well suited to analyzing visual data. CNNs are commonly used to process image and video tasks. And, because CNNs are so effective at identifying objects, they are frequently used for [computer vision](https://www.techtarget.com/searchenterpriseai/definition/machine-vision-computer-vision) tasks, such as [image recognition](https://www.techtarget.com/searchenterpriseai/definition/image-recognition) and [object recognition](https://www.techtarget.com/whatis/definition/object-recognition), with common use cases including self-driving cars, facial recognition and medical image analysis.

Older forms of [neural networks](https://www.techtarget.com/searchenterpriseai/definition/neural-network) often needed to process visual data in a gradual, piece-by-piece manner -- using segmented or lower-resolution input images. A CNN's comprehensive approach to image recognition enables it to outperform a traditional neural network on a range of image-related tasks and, to a lesser extent, speech and audio processing.

CNN architecture is inspired by the connectivity patterns of the human brain -- in particular, the visual cortex, which plays an essential role in perceiving and processing visual stimuli. The [artificial neurons](https://www.techtarget.com/searchcio/definition/artificial-neuron) in a CNN are arranged to efficiently interpret visual information, enabling these models to process entire images.

CNNs also use principles from linear algebra, particularly convolution operations, to extract features and identify patterns within images. Although CNNs are predominantly used to process images, they can also be adapted to work with audio and other signal data.

**CNN hoạt động như thế nào?**

CNN được xây dựng dựa trên các lớp (layers) chuyên biệt, mỗi lớp thực hiện một chức năng cụ thể để xử lý dữ liệu. Dưới đây là các thành phần chính:

**Lớp Tích chập (Convolutional Layer):**

Đây là trái tim của CNN. Lớp này sử dụng các bộ lọc (filters) để "quét" qua dữ liệu đầu vào (như hình ảnh) và trích xuất các đặc điểm quan trọng, chẳng hạn như cạnh, góc, hoặc hoa văn.

Ví dụ: Trong một bức ảnh, lớp tích chập có thể phát hiện các đường viền hoặc vùng sáng tối.

**Lớp Kích hoạt (Activation Layer):**

Thường sử dụng hàm ReLU (Rectified Linear Unit), lớp này giúp mô hình học các mối quan hệ phi tuyến tính bằng cách loại bỏ các giá trị âm, giữ lại các giá trị dương.

**Lớp Gộp (Pooling Layer):**

Lớp này giảm kích thước dữ liệu (downsampling) để giảm tải tính toán và tránh hiện tượng quá khớp (overfitting). Một phương pháp phổ biến là Max Pooling, lấy giá trị lớn nhất trong một vùng nhất định.

**Lớp Kết nối Đầy đủ (Fully Connected Layer):**

Ở giai đoạn cuối, các đặc điểm đã trích xuất được "làm phẳng" (flattened) và đưa vào một mạng nơ-ron truyền thống để phân loại hoặc đưa ra dự đoán. Ví dụ: Xác định ảnh chứa mèo hay chó.

Ứng dụng của CNN

CNN được sử dụng rộng rãi trong nhiều lĩnh vực, bao gồm:

Nhận diện hình ảnh: Phân loại ảnh (ví dụ: phân biệt mèo và chó), nhận diện khuôn mặt trên mạng xã hội như Facebook.

Xe tự hành: Phát hiện vật thể trên đường, như biển báo giao thông hoặc người đi bộ.

Y học: Phân tích hình ảnh y tế để phát hiện bệnh, như ung thư qua ảnh X-quang.

Xử lý ngôn ngữ tự nhiên: Dù ít phổ biến hơn, CNN cũng được dùng để phân tích văn bản hoặc chuỗi dữ liệu.

**Lịch sử và tầm quan trọng**

CNN được giới thiệu lần đầu vào năm 1980 bởi một nhà khoa học máy tính Nhật Bản, nhưng nó trở nên nổi tiếng nhờ Yann LeCun với mô hình LeNet-5 (1998), dùng để nhận diện chữ viết tay.

Bước ngoặt lớn xảy ra vào năm 2012, khi mô hình AlexNet (dựa trên CNN) chiến thắng cuộc thi ImageNet, đánh dấu sự bùng nổ của học sâu trong AI.

Tại sao CNN đặc biệt?

Hiệu quả với dữ liệu hình ảnh: Không như mạng nơ-ron thông thường, CNN tận dụng cấu trúc không gian của hình ảnh (pixel gần nhau thường liên quan), giảm số lượng tham số cần tính toán.

Tự động hóa: CNN tự học các đặc điểm từ dữ liệu mà không cần con người thiết kế thủ công.