

Neural networks, a subset of machine learning, are computational models inspired by the human brain's neural structure. They consist of layers of nodes (neurons) connected by weighted edges, where each node processes input data through an activation function and passes it to the next layer. These networks are primarily used for tasks like classification, regression, and pattern recognition. The most basic form, the feedforward neural network, processes data in one direction from input to output layers. More advanced architectures include convolutional neural networks (CNNs), which excel at image processing by learning spatial hierarchies, and recurrent neural networks (RNNs), designed for sequence data such as time series or natural language. Training a neural network involves adjusting the weights using optimization algorithms like gradient descent, where backpropagation calculates the error gradient to refine these weights, improving the network's predictions. Neural networks, especially deep learning models, require vast amounts of data and computational power, but once trained, they can perform highly complex tasks, from object recognition to language translation, with remarkable accuracy. Their ability to automatically learn features from raw data has made them a cornerstone of modern AI, transforming industries ranging from healthcare to finance by enabling systems to make decisions and predictions based on previously unattainable levels of insight.