

Figure 1: Neuron with corresponding biologically inspired labels.  
(Adopted figure from [1])

### 0.0.1 Neural Networks & Deep Learning

- *Summary of NN*
- *Structure of NN*
- *Training & Inference stages*
- *weight update methodologies*
- *Feed Forwards*
- *Feedback Neural Network*
- *Self-organizing Neural Network*
- *Weight parameters updated using back-propagation*

Deep learning is a subcategory of machine learning techniques where a hierarchy of layers perform some manner of information processing with the goal of computing high level abstractions of the data by utilising low level abstractions identified in the early

layers [2].

Neural networks fundamental purpose is to transform an input vector commonly referred to as  $X$  into an output vector  $\hat{Y}$ . The output vector  $\hat{Y}$  is some form of classification such as a binary classification or a probability distribution over multiple classes [3]. Between the input layer ( $X$ ) and the output layer ( $\hat{Y}$ ) there exists some number of interior layers that are referred to as hidden layers, the hidden and output layers are composed of neurons that pass signals derived from weights through the network, this model of computing was inspired by connectionism and our understanding of the human brain, see Fig. 1 for labels of the analogous biological components. Weights in a neural network effectively correspond to the synapses in the brain and the output of the neuron is modelled as the axon. All neurons in a Neural network have weights corresponding to their inputs, these weights are intended to mirror the value scaling effect of a synapse by performing a weighted sum operation [1].

Neural networks and deep neural networks are often referred to interchangeably, they are primarily distinguished by the number of layers, there is no hard rule indicating when a neural network is considered deep but generally a network with more than 3 hidden layers is considered a deep neural network, the rest of this dissertation will refer to DNNs for consistency. Each neuron in a DNN applies a non-linear activation function to the result of its weighted sum of inputs and weights, without which a DNN would just be a linear algebra operation [1], the cumulative effect of the activations in each layer results in elaborate causal chains of transformations that influence the aggregate activation of the network.

**Backpropagation** Although not the first to propose this approach [4] the 1986 paper Learning representations by back-propagating errors [5] popularised back-propagation for updating weights during training multi-layer networks.

DNNs can be categorised as feedforwards, feedback, and self-organising networks depending on their processing method [6].

There are many popular deep neural network architectures, this document will continue to outline the CNN & RNN architectures because these provide a high level overview of the type of models that will be used for the research posed in this dissertation.

### 0.0.2 Convolutional Neural Networks

*Convolutional Neural Network (CNN)*

- *A class of DNN*
- *CNN consist of: Convolutional Layers, Pooling layers & fully connected layers.*
- *Convolutional Layers contain sets of filters/kernels*

Much like traditional neural networks the CNN architecture was inspired by human and animal brains, the concept of processing the input with local receptive fields is conceptually similar some functionality of the cat's visual cortex [7]–[9]. The influential paper by Hubel & Weisel [7] ultimately had a significant influence on the design of CNNs via the Neocognitron, as proposed by Fukushima in [10] and again evaluated in [11] (**provide some comment on these papers**).

A critical aspect of image recognition is robustness to input shift and distortion, this robustness was indicated as one of the primary achievements of the Neocognitron in Fukushima's paper [10]. LeCun and Bengio provide comprehensive explanations of how traditional DNNs are so inefficient for these tasks

The local receptive fields enable neurons to extract low level features such as edges, corners, and end-points with respect to their orientation. CNNs are robust to input shift or distortion by using receptive fields to identify these low level features across the entire input space, performing local averaging and downsampling in the layers following convolution

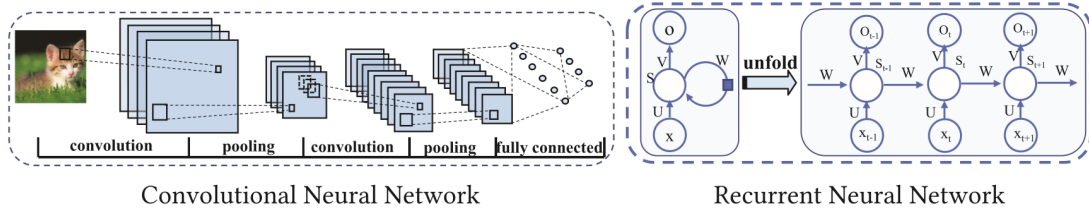


Figure 2: A typical example of a CNN (left) and RNN (right)  
**(Adopted figure from [6])**

layers means the absolute location of the features is less important than the position of the features relative to the position of other identified features [8]. Each layer produces higher degrees of abstraction from the input layer, in doing so these abstractions retain important information about the input, these abstractions are referred to as feature maps. The layers performing downsampling are known as pooling layers, they reduce the resolution or dimensions of the feature map which reduces overfitting and speeds up training by reducing the number of parameters in the network [9].

Convolutional Networks for Images, Speech, and Time-Series by LeCunn & Bengio

CNNs have been found to be effective in many different AI domains, popular applications include: computer vision, NLP, and speech processing.

### 0.0.3 Recurrent Neural Netowrks

*Recurrent Neural Network (RNN)*

RNNs are deep learning models that use loops in their layer connections to make predictions with sequential inputs and maintain state over those inputs, this architecture is designed specifically for time series predictions [12].