Deep Learning Algorithms Comparison

Algorithms	Applications	Pros	Cons
Multilayer Perceptron's (MLPs) > the most basic deep learning algorithm (with single hidden layer). > a form of feed forward neural network. > uses supervised learning technique called backpropagation for training. > model (example)	> classification problems. > speech, image recognition & data compression.	> do not make any assumptions regarding the Probability density functions (PDF) unlike the other models which are based on Probability. > can distinguish data that is not linearly separable.	> when updating the weights in layers, network may be stuck in a local minimum which can hamper accuracy. > due to the hard-limit transfer function, the perceptron's can only give outputs in the form of 0 and 1.
Convolutional Neural Networks (CNN) > ConvNet is a popular deep learning algorithm which has hidden layers that perform convolutions. > has three main types of layers: convolutional (set of filters), pooling (down-sampling), and fully-connected (neurons) layers. > transform the original image layer by layer from the original pixel values to the final class scores. > use some regularization techniques like dropout where a particular node or connection is skipped. > model (example)	> main use-cases in image recognition and object detection tasks. > recommender systems. > natural language processing tasks. > financial time series forecasting.	> results are more accurate especially for image/object recognition use cases when compared to other ML algorithms.	> for training ConvNet, very high computation power is required. Thus, not very cost-effective.

Recurrent Neural Networks (RNNs) > a type of ANNs which uses sequential or time series data. > a class of neural networks that allow previous outputs to be used as inputs while having hidden states. > allow the processing of variable-length (or even infinite-length) sequences. > an output is produced which is coped and provided back to the network like a loop (internal memory). > model (example)	> mostly used in the fields of natural language processing and speech recognition. > ideal for tasks such as words auto completion, text recognition and video frames analysis.	> ability of to remember information throughout the training period plays a very pivotal role in time series prediction. > model size not increasing with size of input. > possibility of processing input of any length.	> computation is time-taking because of its recurrent nature. > difficulty of accessing information from a long time ago.
Long Short-Term Memory Networks (LSTMs) > a special kind of RNN highly capable of learning long-term dependencies. > network consists of different memory blocks called cells which remember things. > changes to memory blocks are done through mechanisms referred to as gates. > model (example)	> ideal for tasks such as sentences auto completion, caption generation and video frames analysis. > anomaly detection in network traffic data.	> can handle the information in memory for the long period of time as compare to RNN.	> high computation and resources are required to train the the model. > are prone to overfitting.

Restricted Boltzmann Machines (RBMs) > unsupervised neural network that belongs to energy based model. **>** a generative stochastic ANNs that can learn a probability distribution over its set of inputs. > are shallow, two-layer neural nets that constitute the building blocks of deep-belief networks. > model **(**example)	> feature extraction in pattern recognition and recommendation engines. > classification problems. > feature learning and topic modeling.	> can be pre-trained in a completely unsupervised way as the learning algorithm can make very efficient use of unlabeled large data.	> calculating energy gradient function while training is very difficult. > adjusting weights using the CD-k algorithm is not as easy as Backpropagation.
Radial Basis Function Networks (RBFNs) > a simple three-layer feedforward neural network with an input layer, a hidden layer consisting of a number of RBF non-linear activation units, and a linear output layer that acts as a summation unit to give the final output. > the hidden layer uses an unsupervised learning algorithm (k-means clustering). > mean squared error (MSE) is used to determine the error and the weights are tweaked accordingly to minimize MSE. > model (example)	> mostly used for speech recognition, time series analysis, image recognition and medical diagnosis.	> fast training phase when compared to MLP, as there is no backpropagation involved.	> classification takes time compared to MLP due to every node in the hidden layer has to compute the RBF function for the input sample vector.

Self Organizing Maps (SOMs) > enable data visualization by reducing the dimensions of data through SOMs network. > group similar data items together by creating a 1D or 2D map. > weights are initialized randomly for each node. At each step, one sample vector x is randomly taken from the input data set and the distances between x and all the other vectors are computed. > model (example)	> useful in the healthcare sector for creating 3D modeling. > Image analysis, fault diagnosis, process monitoring and control.	> easily interpret and understand the data using SOM. > dimensionality reduction is simpler to check for any similarities within our data.	> requires neuron weights to be sufficient to cluster the input data. > while training SOM, if we provide less or extensively more data may not get informative or very accurate output.
Generative Adversarial Networks (GANs) > unsupervised learning algorithm capable of automatically discovering and learning the patterns in the data. > GANs consist of two neural networks: a generator neural network that generates new examples and discriminator network responsible for evaluating the generated examples whether they belong to the actual training dataset. > model (example)	> widely used in the creative industry for 3D object generations. > useful in editing images (Deepfake), generating cartoon characters, illustrations for novels or articles.	> able to learn the internal representation (messy and complex distributions) in any data. > can recognize objects as well as can calculate the distance between them.	> when generates new data from original data, no evaluation metric to judge the accuracy of output. > high computation and time required for model training.

Autoencoders > unsupervised algorithm similar to Principal Component Analysis (PCA) to convert multi-dimensional data into low-dimensional data. > the encoder compresses the input into a latent space representation which can be reconstructed later to get the original input. > the decoder aims to reconstruct the code to its original form but may not be perfectly accurate as original and might have some loss. > model (example)	> used in the healthcare industry for medical imaging (breast cancer detection). > coloring of images, image compression, and denoising.	> usage of multiple encoder and decoder layers reduce the computational cost of representing some functions to a certain extent.	> not efficient as compared to GANs when it comes to reconstructing an image. > might lose important data from the original input after encoding.
Deep Belief Networks (DBN) > a networks model built by stacking a number of unsupervised Restricted Boltzmann Machines (RBM) layers. > uses a layer-by-layer approach for learning all the generative weights and all the top-down approaches. > networks are pre-trained by using the Greedy algorithm. > model (example)	> mostly to recognize, cluster, generate images, video sequences and motion-capture data.	> can work with even a small labeled dataset. > provide robustness in classification (view angle, size, position, color, etc).	> high computation and time required for model training.

Transformer Neural Networks > network model that adopts the mechanism of attention, differentially weighing the significance of each part of the input data. > an architecture for transforming one sequence into another one with the help of two parts (Encoder and Decoder) but it differs from existing sequence-to-sequence models because it does not imply any Recurrent Networks (GRU, LSTM, etc.) > model (example)	> used primarily in the field of natural language processing (NLP) and in computer vision (CV). > designed to handle sequential input data such as natural language, for tasks such as translation and text summarization.	> the model of choice for NLP problems, replacing RNN models such as LSTMs. > led to the development of pretrained systems such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer).	> attention can only deal with fixed-length text strings and has to be split into a certain number of segments or chunks before being fed into the system as input. > high computation and time required for model training.