***Activation function:***

The activation function determines the output value of a neuron in a neural network as a function of the sum of the weighted inputs it receives .

***Artificial Intelligence (AI):***

The suite of computational tools and efforts that seek to mimic human intelligence and information processing

***Artificial Neural Networks (ANN):***

Artificial neural networks are computational systems designed to imitate how a human brain processes information. In these models, ‘neurons’ successively transform information to predict some outcome

***Backpropagation:***

Backpropagation is the minimisation of some loss function to optimise a model’s parameterisation—in the case of an ANN this involves tuning the weights. For DNN the process works backward from the final fully connected layer

***Convolutional Layers:***

The core component of a CNN architecture, the convolutional layers apply the same set of learned filters over the entire image to extract the features that will be used for classification and prediction

***Convolutional Neural Networks (CNN):***

A form of specialised deep neural network typically used for image classification. CNNs are made of three types of layers: convolutional, pooling and fully connected layers

***Data augmentation:***

A technique used to synthesise new training data by modifying existing data (for example, image rotation, image cropping, and so on). This technique can be used to mitigate against overfitting when training a DL model with small training sets

***Deep learning (DL):***

Deep learning is the component of machine learning that uses neural network algorithms with multiple hidden layers (the multiple layers make them ‘deep’)

***Error function:***

Also known as a loss function, the error function quantifies the deviation of a network’s prediction from the ground true value. This error is minimised during model training

***Feature extraction:***

The conversion of (raw) input data into (simpler) representations to make classifications or predictions without losing key characteristics of the original data. Feature extraction often results in better DL model performance than using the raw data

***Feature selection:***

The process of input variable selection (synonymous with variable selection in classic frequentist regression models). The reduction of input parameters can aid in both computational costs and spurious associations

***Feed-forward Neural Networks (FNN):***

An architecture of artificial neural network in which information is fed from the input layer neurons to the hidden layer neurons before being transferred to output layer neurons. Information flow is in one direction only (input layer → hidden layers → output layer)

***Fully connected layers:***

In a fully connected layer of an ANN, all neurons are connected with all other neurons in the preceding and succeeding layers

***Hidden layer(s):***

The layer(s) between input and output layers containing neurons. These neurons receive inputs which are weighted and then produce outputs based on an activation function

***Input Layer:***

The layer in an ANN which receives the initial raw data, processes it and passes it onto the hidden layers. The input layer is the first step in an artificial neural network

***Layers:***

Neural networks are made up of a series of layers which receive the input information (input layer), process it through a series of (hidden) layers before making predictions (output layer). There are typically three categories of layers in neural networks; input layers, hidden layers and output layers, although their organisation will vary depending on the ANN architecture

***Long short-term memory (LSTM):***

An architecture designed to overcome some technical problems with training RNN models. In place of neurons, these models use memory blocks that are connected between layers. Each block has a memory of recent sequences and a gate that controls its state and the information it outputs; this architecture allows the information flow from a block to be conditional on its state

***Machine Learning (ML):***

Machine learning is a subset of AI that develops algorithms designed to iteratively learn (for example, identify patterns) from data

***Neurons/Nodes:***

Each layer in a neural network is comprised of a series of neurons, each of which is a mathematical operation. These neurons apply the operation to incoming data, multiply it by a weight, and pass the resulting value through an activation function to other neurons in the network.

***Output Layer:***

The final layer in an artificial neural network where the information that was processed by the hidden layers is reformulated to create the desired predictions. The neurons in this layer also have their own weights that are applied to aid in the derivation of the prediction

***Pooling layers:***

The pooling layer in a CNN aggregates information by merging the results of multiple CNN filters

***Recurrent Neural Network (RNN):***

An ANN that can represent auto-correlation between data points by incorporating dependencies between observations. This architecture makes RNNs particularly useful for predicting time-series data

***Saliency maps:***

Heatmaps developed to highlight the most important portions of an image in a DL model (usually a CNN) making a prediction—they are a tool designed to improve DL model interpretability

***Supervised training:***

The practice of providing the ANN with data that are ‘labelled’ in some way (for example, wildlife imagery with the species in the image tagged). This process enables training a model for a particular predictive task and then assessing its (predictive) performance

***Testing data:***

Data used to test a model’s performance for a given task; often a subset of all data available and not used in model training

***Training data:***

Data used to train a model for a particular task. These data are typically held separate to the testing data to prevent overfitting

***Transfer learning:***

The practice of using knowledge gained from solving one problem in a separate but related problem. In the context of DL models, this is applying a model trained in one context in a new setting

***Unsupervised training:***

The practice of training a model with unlabelled input data; clustering algorithms are a well-known example of this approach

***Weights:***

The weights in an ANN model control the information flowing from a node (in some ways analogous to the slope in a regression model). The weights are combined for all nodes in a layer in an activation function which determines how information is passed out of a layer