**The Impact of Activation Function in Neural Network**

**Introduction:**

The activation function of an unit in an artificial neural network specifies the output of that node given an input or set of inputs. A standard integrated circuit can be thought of as a digital network of kernel function that can be switched on or off depending on the information.

An artificial neural network's activation function is a crucial component since it determines whether or not a neuron should be engaged. The activation function of a node in an artificial neural network determines the output of that node given an input or group of inputs. Activation functions are an important aspect of a neural network's design. The hidden layer's activation function determines how well the network model learns the training dataset. The type of predictions the model can produce is determined by the activation function used in the output layer.

As a result, the activation function is a decision function, and non-linear decision functions are required, which is why all activation functions have some non-linearity.

**Literature Review:**

In paper [1], the accompanying paper discusses many forms of activation functions, with a focus on the 'Relu' activation function. The 'Relu' activation function allows the network to verify which weight of the input delivers the best accuracy for disease prediction using back propagation. The loss function and optimizer approaches are employed and compared between the RNN and Convolutional Neural Network to provide superior deep learning optimization. Softmax has the ability to handle many classes. Only one class in the other activation function normalizes the output and divides it by the total, resulting in a probability of input to a specific class. Adaptive Moment Estimation (Adam) also adjusts the Learning Rate (LR) and updates the network's weights on a regular basis based on the training data. To scale the learning rate, it uses the epidemiological moving average of the gradients. Adam's disadvantage is that it uses an exponential decaying average of previous data.

**Reference:**

1. <https://www.researchgate.net/publication/326708262_Activation_functions_and_their_characteristics_in_deep_neural_networks>