## DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING (ARTIFICIAL INTELLIGENCE & MACHINE LEARNING)

| Parameters | RNN | LSTM | GRU |
|---|---|---|---|
| Structure | Simple | More complex | Simpler than LSTM |
| Training | Can be difficult | Can be more difficult | Easier than LSTM |
| Performance | Good for simple tasks | Good for complex tasks | Can be intermediate between simple and complex tasks |
| Hidden state | Single | Multiple (memory cell) | Single |
| Ability to retain long-term dependencies | Limited | Strong | Intermediate between RNNs and LSTMs |
| Gates | None | Input, output, forget | Update, reset |
| Gates Function | None | The input gate and target gate are coupled by an update gate which takes responsibility of reset gate | The reset gate is applied directly to the previous hidden state |
| Type of Gate | None | LSTM has three gates: Input Gate, Forget Gate and Output Gate | GRU has two gates: reset and update gates |

| Parameters | RNN | LSTM | GRU |
|---|---|---|---|
| Design | RNN are designed to remember or "store" information from previous inputs. | Special Type of RNN. Both LSTMs and GRUs are designed to address the problem of "vanishing gradients" in RNNs. | Special Type of RNN. Both LSTMs and GRUs are designed to address the problem of "vanishing gradients" in RNNs. VGs occurs when the gradients of weights become very small and has difficulty learning. |
| Memory | RNNs remember information from previous inputs but may struggle with long-term dependencies. | LSTM do not store information as it has output gate. | GRUs do not store information like the LSTMs do and this is due to the missing output gate. |
| Type of Input Data | DL used for processing sequential data, such as text, audio, or time series data. | DL answer for problems involving sequences and time series. | DL answer for problems involving sequences and time series. |
| Learning Type | They can be used for both supervised and unsupervised learning. | They can be used for both supervised and unsupervised learning. | They can be used for both supervised and unsupervised learning. |
| Activation Function | Use sigmoid or tanh function for hidden layers. The tanh has better performance. Only the identity AF is considered linear. All other AFs are non-linear. | in LSTM sigmoid is used as the gating function and the tanh is used as the output AFs, which o/p values bet. 0 and 1. We can use ReLU or LeakyReLU. | tanh or sigmoid. If you pass None, no activation is applied (ie. "linear" activation. |
| Application | useful for tasks such as language translation, speech recognition, and time series forecasting. | useful for tasks such as language translation, speech recognition, and time series forecasting. | useful for tasks such as language translation, speech recognition, and time series forecasting. |

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**
**(ARTIFICIAL INTELLIGENCE & MACHINE LEARNING)**