Background:

Human action Recognition were divided into 3 different levels: human detection (low- level vision), human-tracking (intermediate-level vision) and behaviour understanding methods (high-level vision). The high-level vision has always been a challenging time series classification task. Traditionally, methods from the field of signal processing were used to analyze and distill the collected sensor data. Such methods were for feature engineering, creating domain-specific, sensor-specific, or signal processing-specific features and views of the original data. Statistical and machine learning models were then trained on the processed version of the data. But there exists limitations for this method. One of the limitations is the single processing and domain expertise required to analyze the raw data and engineer the features required to fit a model. Therefore, this is extremely expensive and not scalable. Recently, deep neural network has started delivering the promised of feature learning and are achieving satisfied results in the human activity recognition. There are two main approaches to neural networks that are appropriate for this purpose. CNN and RNN. (Brownlee, 2019)

<https://machinelearningmastery.com/deep-learning-models-for-human-activity-recognition/>

With the recent advent in deep learning, re-current neural network(RNN) have become the dominant tool for sequence tasks such as sequential learning, tracking, object recognition and more. Although RNNs are less powerful than convolutional neural network (CNN) in terms of feature compatibility and image detection, some of them are designed and built to have the ability to process temporal information-data that comes in sequences while CNN is incapable of effectively interpreting temporal information. As a result, these 2 networks are used for completely distinct purposes.

All RNNs have feedback loops in the recurrent layer which allows it to store information in memory over time. This is one of the key architecture in RNN to help it handle sequence data. However, it could be quite difficult to train standard RNN models to solve problems that require learning long-period temporal dependencies. This is due to the fact that the gradient of the loss function decays exponentially as time passes. That means as the model training time gets longer, the time taken to get the result would increase in a much faster rate.

Long-Short-Term-Memory (LSTM) network is a type of RNN that uses special units in addition to standard units in the layer. The units in LSTM contains a “memory cell” that can maintain information for longer period of time than usual. Moreover, a set of gated is used to control when information enters the memory, output or it is forgotten. This special architecture makes the model get used to longer-term dependencies. There exists another similar algorithm with simplified structure named Gated Recurrent Unit (GRU) that is not used in this project. It also uses set of gates to control the flow of information similar to LSTM. But the difference is that GRU does not use separate memory cells so that it has fewer gates than LSTM.

(Lendave, 2021)

<https://analyticsindiamag.com/lstm-vs-gru-in-recurrent-neural-network-a-comparative-study/>

LSTM is considered a type of RNN, it is in fact an extension of RNNs and that is introduced to handle situations where RNNs fail. Although RNN has been used in many deep learning applications, and out of its various applications, the most popular ones are in the fields of speech processing, non-Markovian control, and music composition. Nevertheless, there are drawbacks to RNNs. First, it fails to store information for a longer period of time. At times, a reference to certain information stored quite a long time ago is required to predict the current output. But RNNs are absolutely incapable of handling such “long-term dependencies”. But this is one of key feature required to detect human behavior is video. Second, there is no finer control over which part of the context needs to be carried forward and how much of the past needs to be ‘forgotten’. Other issues with RNNs are exploding and vanishing gradients (explained later) which occur during the training process of a network through backtracking. Thus LSTM was brough into picture. (2021)

<https://www.geeksforgeeks.org/understanding-of-lstm-networks/>

J. Brownlee, “Deep learning models for human activity recognition,” *Machine Learning Mastery*, 05-Aug-2019. [Online]. Available: https://machinelearningmastery.com/deep-learning-models-for-human-activity-recognition/. [Accessed: 07-Apr-2022].

V. Lendave, “LSTM vs gru in recurrent neural network: A comparative study,” *Analytics India Magazine*, 14-Dec-2021. [Online]. Available: https://analyticsindiamag.com/lstm-vs-gru-in-recurrent-neural-network-a-comparative-study/. [Accessed: 07-Apr-2022].

“Understanding of LSTM networks,” *GeeksforGeeks*, 25-Jun-2021. [Online]. Available: https://www.geeksforgeeks.org/understanding-of-lstm-networks/. [Accessed: 07-Apr-2022].