# 21F-CISC684-010: Introduction to Machine Learning
# Assignment: Project Summary

October 31, 2021

## Group Member


- Shayla Sharmin (ID: 702610714)
- Prerana Khatiwada (702616768)


## Selected Paper

M. A. Russo, A. Filonenko, and K.-H. Jo, **"Sports classification in sequential frames using CNN and RNN"**, in Proceedings of the International Conference on Information and Communication Technology Robotics (ICT-ROBOT), Busan, South Korea, 6-8 September 2018, pp. 1–3. DOI: 10.1109/ICT-ROBOT.2018.8549884.

## Summary


The paper we'll use is a sports categorization that falls within the category of scene context analysis. This work can be used for post-game analysis, automatic scoreboard updates, and even automated commentary. Researchers have worked on scene context analysis for years, and as part of this, Russo et al., developed a system that can distinguish football, cricket, tennis, basketball, and ice hockey into five distinct sports based solely on visual information, using CNN (convolutional neural networks) to learn features or visual cues and RNN (recurrent neural networks) to identify the temporal relationship between frames.

A sequence of RGB color frames is employed as input to the convolution network in this paper, with the weights shared among the layers. The first layer in this project is a 32-feature map, and the next three layers are three types of dilated convolution layers. The network is taught the relationship between human action sequences and their surrounding environment context to solve the categorization challenge. They demonstrated how dilated convolution allowed them to observe a larger area and how adding the context module helped to improve accuracy in this experiment. The activation function, which is placed at the output of the CNN layers, is a rectified linear unit. After each activation, batch normalization was performed with RMSProp as the Optimizer. They chose a learning rate of 0.0001, a loss function of category cross entropy, and 60 epochs. They employed gated recurrent units for the recurrent component (GRU).

They created their own dataset, which included a total of 300 video frames drawn from 50 different YouTube videos. They achieved 99.58 percent accuracy on the training dataset and 96.66 percent accuracy on the test dataset. We will incorporate their work in our project and attempt to gather their dataset in order to test the results. In addition, we intend to use datasets from UCF or KTH. We intend to apply a similar approach to a new dataset with a completely different sports database. We'll aim to improve the model by hyper-tuning it for the new dataset in the hopes of improving accuracy.