

ITEC5920 - Applied Deep Learning  
Project Proposal  
Carleton University

Georges Ankenman - 100935237  
Davood Akhavannasab - 101133274

February 2, 2022

# Background

Nowadays, sensors can be seen in many devices such as smartphones, automobiles, and industrial instruments. Also, sensor information plays a critical role in a wide range of industries from health to safety. This information usually comes in a time-series format can be helpful in different fields such as disaster prevention, human health alerts, and air quality control by measurement interpretation. Some sensor information may contain noise (usually environmentally). We can find patterns by classifying the sensor information. The most common challenging issues of classification application are when we face imbalanced data and lack of training data. For example, if you consider a time series data set containing vibration data of a bridge, we will face a lack of earthquake information while it is a rare phenomenon. In several related works, CNN (Convolutional Neural Network), Multilayer Perceptrons (MLP), and LSTM (Long Short Term Memory network) have succeeded in time-series classification problems.

## Goals

- Classifying current sensor information by the following deep learning architectures:
  - CNN
  - MLP
  - LSTM
  - FCN (Fully Convolutional Neural Network)
- Produce reproducible Python code accessible in a public GitHub repository
- Prepare a 4-page report in IEEE format including Introduction, literature review, Problem statement, proposed solution, results, and conclusion

## Methodology

We are given a dataset that contains a matrix of 336 measurements with each measurement contains 3000 data samples. With this data, we must train a deep learning model that will allow for automatic classification of new data. Data is already standardised, so we do not need to reduce it further.

We will then proceed with training and comparing various models, notably Multilayer Perceptrons (MLP), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and a hybrid of models. We want to have multiple models to choose from as one model might be able to more accurately classify new sensor data over another given the available dataset. Once we have a trained model, we will want to validate the model by feeding it existing data to confirm the expected results, making adjustments accordingly. Once we have confirmed that the models are trained correctly, we will be feeding it new data and provide results. Metrics and success indicators will be determined at a later time, which will allow us to

## Pitfalls and Risks

One of the inherit pitfalls that we think we will come across is potential issues with the dataset and model. From wrongfully pruning data outliers in the data standardization phase of the project to issues with the trained model. We could find that a given model type trained on the the exact same dataset as the other models might have extra noise in the model which would lead to a false classification of the test data. Bad data inputs could be a problem we come across, mitigations techniques will need to be researched and implemented.

## References

- A. Shrestha and . J. Dang, "Deep Learning-Based Real-Time Auto Classification of Smartphone Measured Bridge Vibration Data," 2020.
- D. P. Francis, M. Laustsen and H. Babamoradi, "Classification of colorimetric sensor data using time series," 2021.
- <https://machinelearningmastery.com/when-to-use-mlp-cnn-and-rnn-neural-networks/>
- <https://machinelearningmastery.com/gentle-introduction-long-short-term-memory-networks-experts/>
- [https://keras.io/examples/timeseries/timeseries\\_classification\\_from\\_scratch/](https://keras.io/examples/timeseries/timeseries_classification_from_scratch/)

- <https://www.capitalone.com/tech/machine-learning/10-common-machine-learning-mistakes/>
- <https://towardsdatascience.com/machine-learning-pitfalls-e54ac3edc25>