Literature Review

Quality Prediction in Mineral Processing using Machine Learning

1) Dataset Review

- This Dataset comes from one of the most important parts of a mining process: a **froth flotation plant**!
- The main goal is to use this data to predict how much impurity is in the ore concentrate.
- As this impurity is measured every hour, if we can predict how much silica (impurity) is in the ore concentrate, we can help the engineers, giving them early information to take actions (empowering!).
- Hence, they will be able to take corrective actions in advance (reduce impurity, if it is the case) and also help the environment (reducing the amount of ore that goes to tailings as we reduce silica in the ore concentrate).

In [ ]:

```
import pandas as pd
import numpy as np
df = pd.read_csv("thesis/MiningProcess_Flotation_Plant_Database.csv")
```

In [2]:

```
df.head()
```

Out[2]:

	date	% Iron Feed	% Silica Feed	Starch Flow	Amina Flow	Ore Pulp Flow	Ore Pulp pH	Ore Pulp Density	Flotation Column 01 Air Flow	Flotation Column 02 Air Flow	...	Flotation Column 07 Air Flow	Flotation Column 01 Level	Flotation Column 02 Level	Flotation Column 03 Level	Flotation Column 04 Level
0	2017-03-10 01:00:00	55,2	16,98	3019,53	557,434	395,713	10,0664	1,74	249,214	253,235	...	250,884	457,396	432,962	424,954	443,558
1	2017-03-10 01:00:00	55,2	16,98	3024,41	563,965	397,383	10,0672	1,74	249,719	250,532	...	248,994	451,891	429,56	432,939	448,086
2	2017-03-10 01:00:00	55,2	16,98	3043,46	568,054	399,668	10,068	1,74	249,741	247,874	...	248,071	451,24	468,927	434,61	449,688
3	2017-03-10 01:00:00	55,2	16,98	3047,36	568,665	397,939	10,0689	1,74	249,917	254,487	...	251,147	452,441	458,165	442,865	446,21
4	2017-03-10 01:00:00	55,2	16,98	3033,69	558,167	400,254	10,0697	1,74	250,203	252,136	...	248,928	452,441	452,9	450,523	453,67

5 rows × 24 columns

In [5]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 737453 entries, 0 to 737452
Data columns (total 24 columns):
 #   Column                        Non-Null Count   Dtype 
---  ------                        --------------   ----- 
 0   date                          737453 non-null  object
 1   % Iron Feed                   737453 non-null  object
 2   % Silica Feed                 737453 non-null  object
 3   Starch Flow                   737453 non-null  object
 4   Amina Flow                    737453 non-null  object
 5   Ore Pulp Flow                 737453 non-null  object
 6   Ore Pulp pH                   737453 non-null  object
 7   Ore Pulp Density              737453 non-null  object
 8   Flotation Column 01 Air Flow  737453 non-null  object
 9   Flotation Column 02 Air Flow  737453 non-null  object
 10  Flotation Column 03 Air Flow  737453 non-null  object
 11  Flotation Column 04 Air Flow  737453 non-null  object
 12  Flotation Column 05 Air Flow  737453 non-null  object
 13  Flotation Column 06 Air Flow  737453 non-null  object
 14  Flotation Column 07 Air Flow  737453 non-null  object
 15  Flotation Column 01 Level     737453 non-null  object
 16  Flotation Column 02 Level     737453 non-null  object
 17  Flotation Column 03 Level     737453 non-null  object
 18  Flotation Column 04 Level     737453 non-null  object
 19  Flotation Column 05 Level     737453 non-null  object
 20  Flotation Column 06 Level     737453 non-null  object
 21  Flotation Column 07 Level     737453 non-null  object
 22  % Iron Concentrate            737453 non-null  object
 23  % Silica Concentrate          737453 non-null  object
dtypes: object(24)
memory usage: 135.0+ MB
```

Content

- The first column shows time and date range (from march of 2017 until september of 2017). Some columns were sampled every 20 second. Others were sampled on a hourly base.
- The **second and third columns** are quality measures of the **iron ore** pulp right before it is fed into the flotation plant.
- **Column 4** until **column 8** are the most important variables that impact in the ore quality in the end of the process.
- From **column 9** until **column 22**, we can see process data (level and air flow inside the flotation columns, which also impact in ore quality.
- The last two columns are the final iron ore pulp quality measurement from the lab.
- Target is to predict the last column, which is the % of silica in the iron ore concentrate.

Goal

- Is it possible to predict % Silica Concentrate every minute?
- How many steps (hours) ahead can we predict % Silica in Concentrate? This would help engineers to act in predictive and optimized way, mitigatin the % of iron that could have gone to tailings.
- Is it possible to predict % Silica in Concentrate whitout using % Iron Concentrate column (as they are highly correlated)?

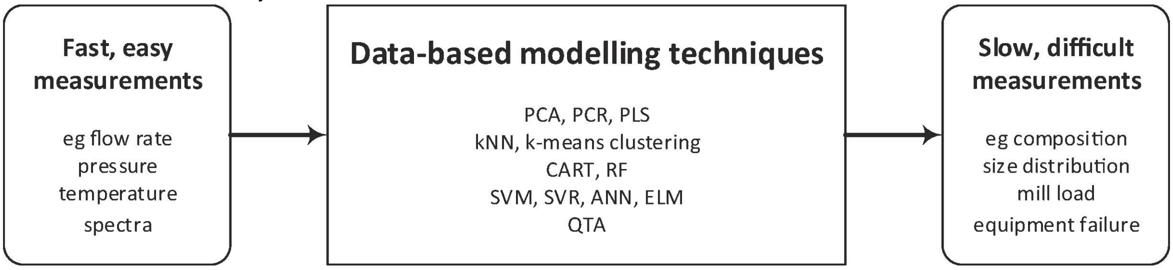
2) Machine learning applications in minerals processing: A review

- By J.T. McCoy, L. Auret
- [Research Paper (https://www.sciencedirect.com/science/article/pii/S0892687518305430)](https://www.sciencedirect.com/science/article/pii/S0892687518305430)

- ## In this paper :-
  - Machine learning applications in mineral processing from 2004 to 2018 are reviewed.
  - Data-based modelling; fault detection and diagnosis; and machine vision identified as main application categories.

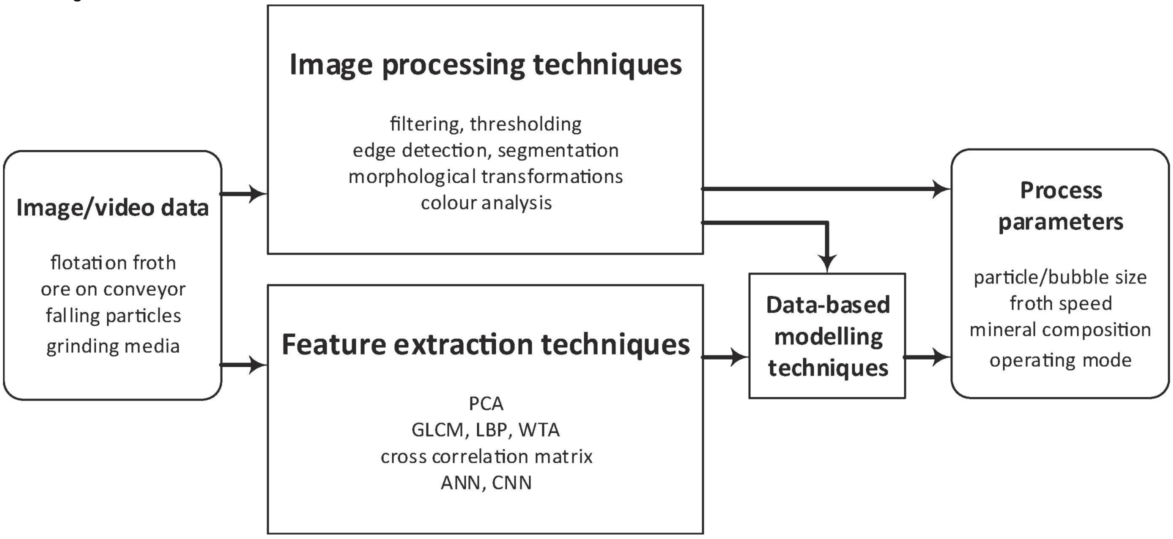
- ### Abstract of the paper
  - Machine learning and artificial intelligence are becoming increasingly important in many different areas of research and business. However, some people might be cautious because they've seen these technologies hyped up in the past without delivering as much as expected. This review is meant to give researchers and business people a clear understanding of how machine learning is being used in mineral processing. We've included a summary of all the techniques we've looked at, including where the data comes from, how successful the techniques have been, what areas of mineral processing they're used in, and what kinds of problems they're used to solve. We also suggest some ideas for future research, like collecting more data, comparing different techniques, getting industry involved, and thinking about the costs and benefits. This also talks about how education in mineral engineering might change in the future because of these technologies.

- ### Data-based modelling in minerals processing
  - The use of data-based modeling techniques in mineral processing is widespread, often employed as "soft sensors" to predict variables that are either infrequently measured or difficult to measure. These methods utilize various machine learning approaches such as neural networks, partial least squares (PLS), and qualitative trend analysis (QTA) to make predictions based on process measurements. For instance, neural networks have been applied to model hydrocyclones, milling circuits, flotation, and furnaces. Some innovative approaches include adaptive modeling for flotation monitoring, where multiple neural networks are employed based on data clustering, and Ensemble techniques for predicting dangerous seismic events in mining. Additionally, methods like QTA have been successfully used for froth condition monitoring in flotation. Moreover, advancements in software accessibility are expected to lead to more widespread adoption of advanced modeling techniques in mineral processing, potentially incorporating time-series information for better accuracy.



- ### Fault detection and diagnosis
  - Fault detection and diagnosis (FDD) is primarily concerned with the detection of abnormal/faulty conditions in complex processing systems, and the identification of the root causes of the faulty conditions. When FDD is performed on the basis of historical data (rather than first principles models of a system) then it is also referred to as data-based or statistical process monitoring.

- **Machine Vision**
  - Machine vision applies machine learning techniques to images and videos, rather than numerical data from sensors, posing challenges due to the high-dimensional nature of image data. Techniques involve image processing, such as filtering and segmentation, to extract relevant information like particle sizes in minerals processing. Another approach is feature extraction, which reduces image data to abstract features correlated with parameters of interest. These techniques are then used to relate image features or parameters to process conditions or key performance indicators using supervised learning methods. The process involves either image processing or feature extraction, followed by predicting process parameters, often supplemented by data-based modeling.



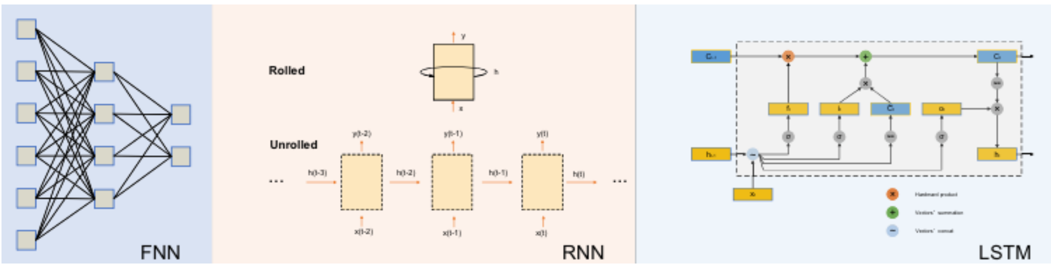
I will primarily be focusing on Data Based modelling approach for my purpose!

3) Purities prediction in a manufacturing froth flotation plant: the deep learning techniques

- By Yuanyuan Pu1, Alicja Szmigiel, Derek B
- [Research Paper (https://link.springer.com/article/10.1007/s00521-020-04773-2)](https://link.springer.com/article/10.1007/s00521-020-04773-2)

- **Abstract of the paper**
  - This study aims to leverage deep learning techniques to model a manufacturing froth flotation process and predict concentrate purities for iron and waste silica. By utilizing a long short-term memory (LSTM) architecture, we addressed the complexities of engineering data in terms of size and temporality. Monitoring and collecting 23 variables reflecting the flotation plant over six months, we then processed, split, and restructured the data for deep learning model implementation. Our model, built with stacked LSTM layers, was trained and tested using the prepared dataset. Results from testing demonstrate the effectiveness of our approach in accurately forecasting real-time concentrate purities. Compared to traditional machine models like random forest, our deep learning model shows superior performance in modeling the flotation process. This study lays groundwork for potential automation control in flotation processes, encouraging further exploration of deep learning applications in mineral processing engineering.

- This is a research paper that discusses the deep learning approach towards solving the problem that we target and that might come as one of the suitable options.
- This specific paper takes sequential prediction approach using Long short term memory alogrithm to train, this alogorithm helps in training while maintaining the previous context. Lets say we have to get prediction at time t, it will the t-1,t-2... into context to predict into future.
- This will help in help in getting real time purity prediction.
- In summary, this study introduces an LSTM model tailored for a manufacturing froth flotation plant, aimed at predicting processed material purities using measurable input variables. The process involves data preparation and restructuring, model construction and training, and subsequent analysis of results. Unlike traditional machine learning models that treat input data as static samples, our LSTM model leverages its unique architecture to capture temporal patterns, resulting in superior prediction accuracy compared to shallow models like random forests. Additionally, we conduct a thorough analysis of the factors influencing output purities, providing insights for adjusting input variables when there is a disparity between expected and predicted output purities. Our model also suggests methods for further improving performance. These findings offer valuable guidance to processing engineers, facilitating timely assessment of processed material purities and enabling productivity optimization and automated processing control. The success of LSTM model underscores the potential of deep learning techniques in the field of froth flotation and mineral processing.



Architecure used to build the model in the research paper.