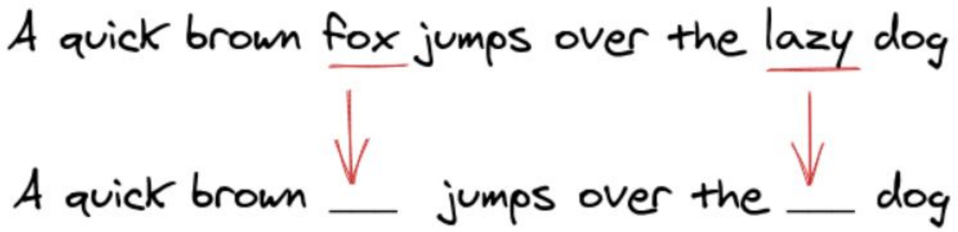
Predicting Human Motion Data With BERT-like Model

Ray Zaremski

Vishnu Sharma

BERT

- Bidirectional Encoder Representations from Transformers (BERT) is a machine learning technique used for natural language processing
- This technique is commonly used to predict missing data in a document or predict future data



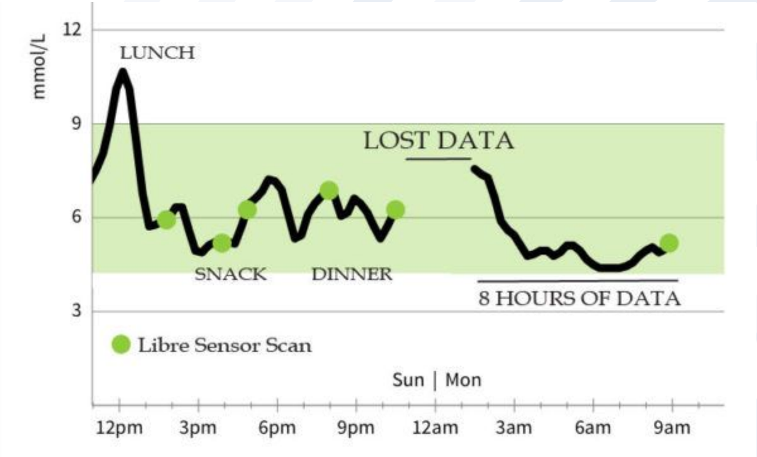
https://amitness.com/2020/02/albert-visual-summary/

Can BERT be extended?

- The problem that BERT aims to solve can be extended to non-NLP problems
  - Can we extend the work done with BERT to populate missing sensor data or predict future sensor data?
- This would also show that the idea of BERT can extend to not just strictly NLP tasks

Why Would We Need to Predict Sensor Data?

- Accelerometers can fail, leaving a combination of valid and missing data that has been recorded
- Sensor datasets may be too sparse for a user's needs
- Data could be corrupted
- Electromagnetic interference can cause segments of unreliable data



https://www.makingyouthink.ca/2019/freestyle-libre-tips/

Future Sensor Data

- The same concept can be extended to future sensor data for similar reasons
- It would generate additional data for a user to run experiments with
- It can be used to diagnose issues with sensor data by comparing the recorded data with the prediction

Evaluation Platform

- Python 3 with Tensorflow and Keras
- Charts were generated using Matplotlib
- There was no significant performance increase for GPU vs. CPU, so no GPU was used

TensorFlow

Dataset

- Dataset was taken from UCI and involves 15 participants using chest-mounted accelerometer to perform 7 different activities
- We only account for one of these activities (going up and down stairs)
- We are training on 80% of the data and testing on 20%

Long Short-Term Memory (LSTM)

- Artificial recurrent neural network (RNN) architecture used in deep learning
- Works well with classifying and predicting time series data
- Solves vanishing gradient problem

LSTM vs. ARIMA

- Both excel in time series forecasting
- ARIMA has better results for short-term forecasting, while LSTMs are preferred in a long-term context
- LSTMs offer overall better performance
- ARIMA has higher complexity

MinMaxScaler

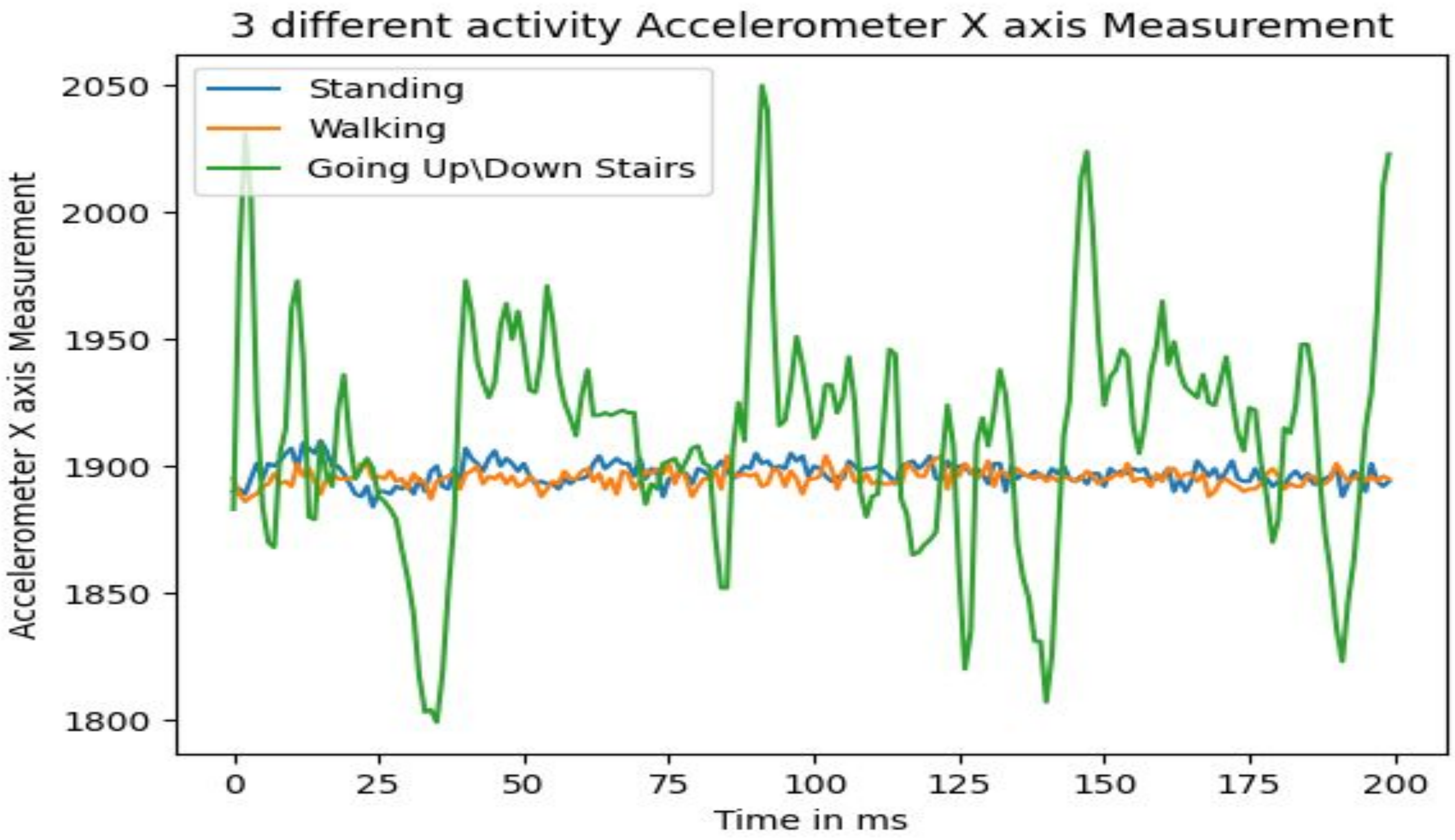
- Input variables may have different units (e.g. feet, kilometers, hours)
  - Could mean that variables have different scales
- Differences in the scales across input variables may increase the difficulty of the problem being modeled
- We used the default MinMaxScaler configuration and scaled values between 0 and 1

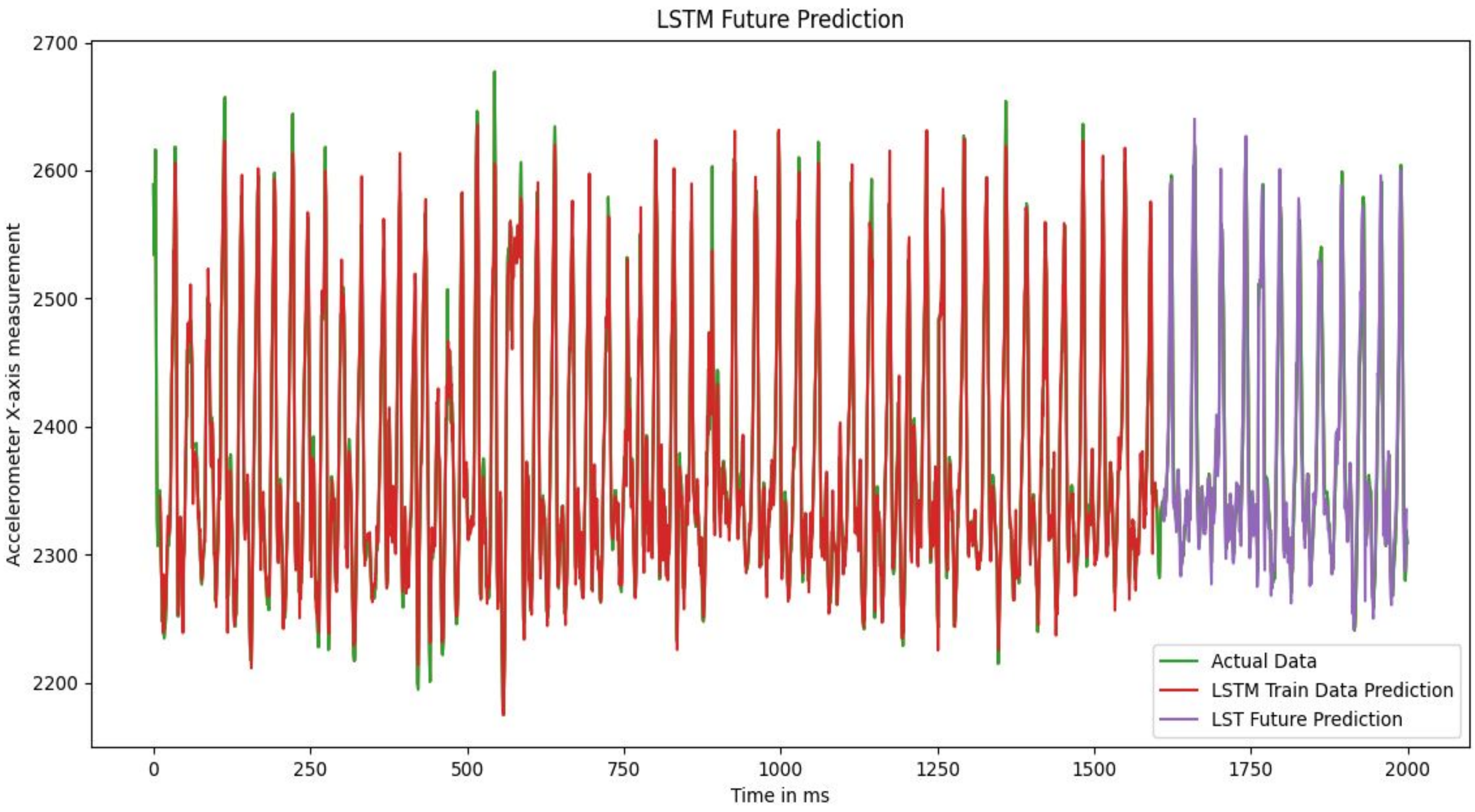
Time Series Generator

- Keras functionality that takes a sequence of data points at equal intervals and time series parameters
- Produces training and validation batches

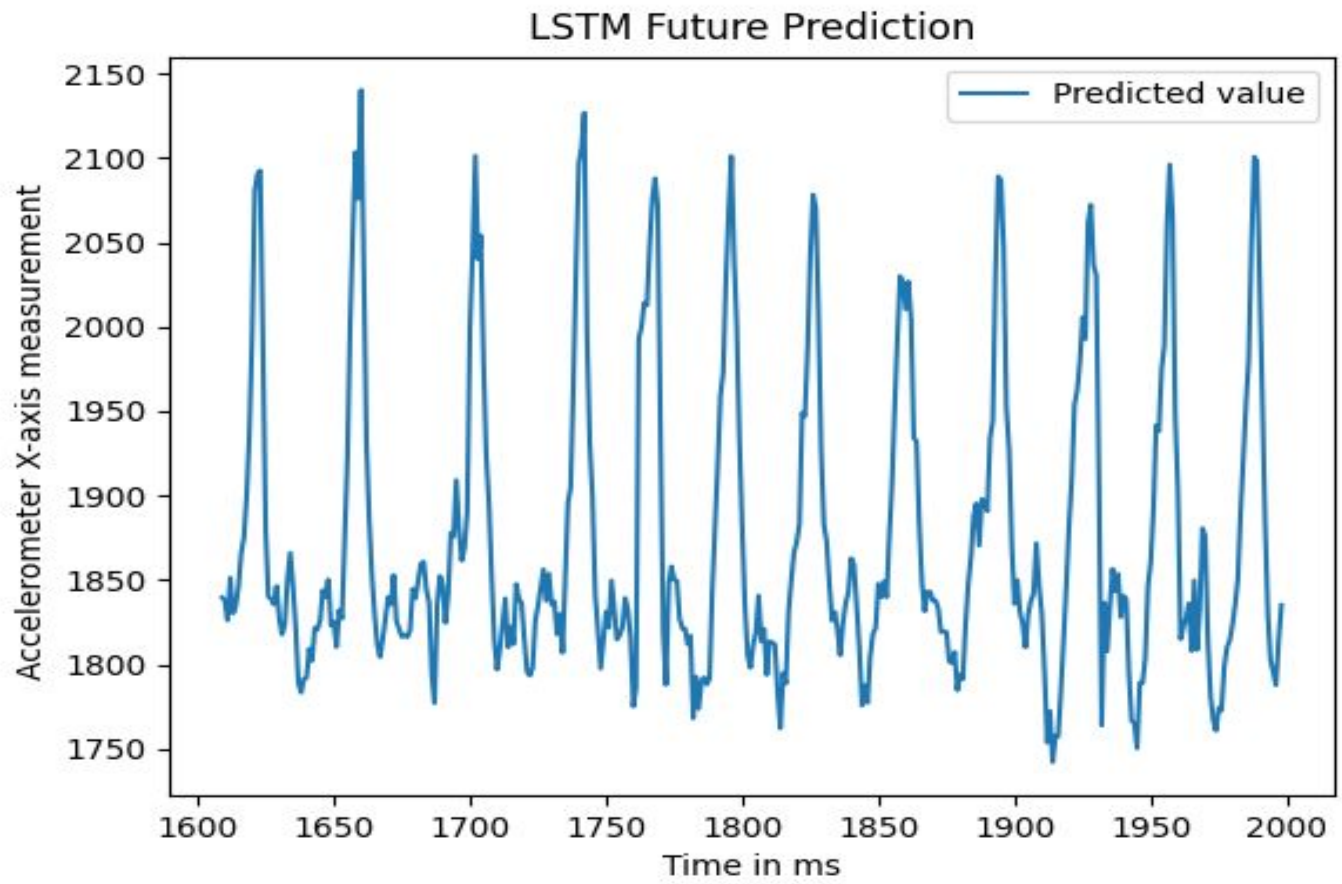
Rectified Linear Units (ReLU)

- The ReLU is a linear activation function that will output the input directly, if it is positive, otherwise it will output zero
- It has become the default activation function for many types of neural networks
    - Using it within a model makes the training easier and will often improve performance

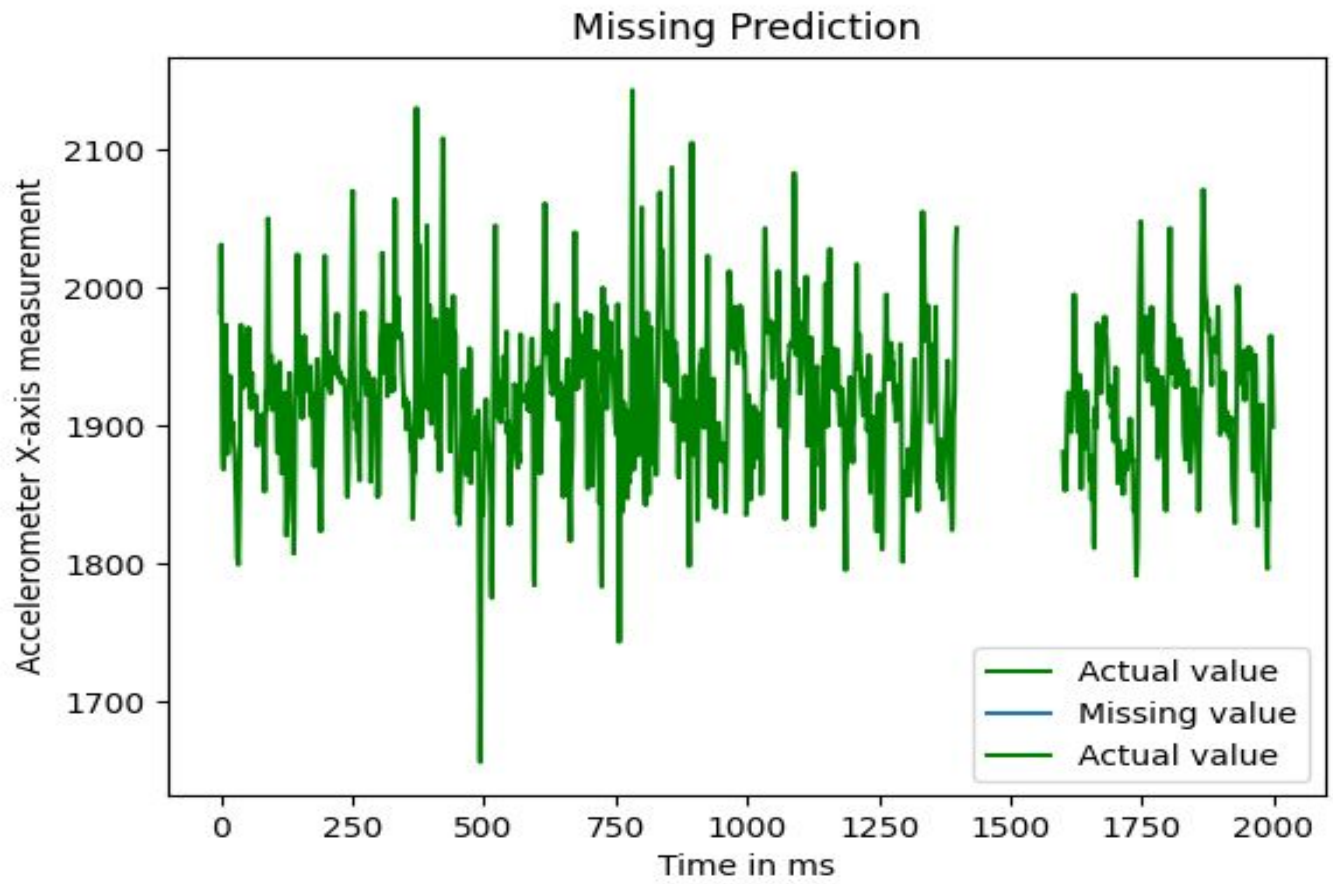

3 different activity Accelerometer X axis Measurement
Standing
Walking
Going Up\Down Stairs
Accelerometer X axis Measurement
2050
2000
1950
1900
1850
1800
0
25
50
75
100
125
150
175
200
Time in ms



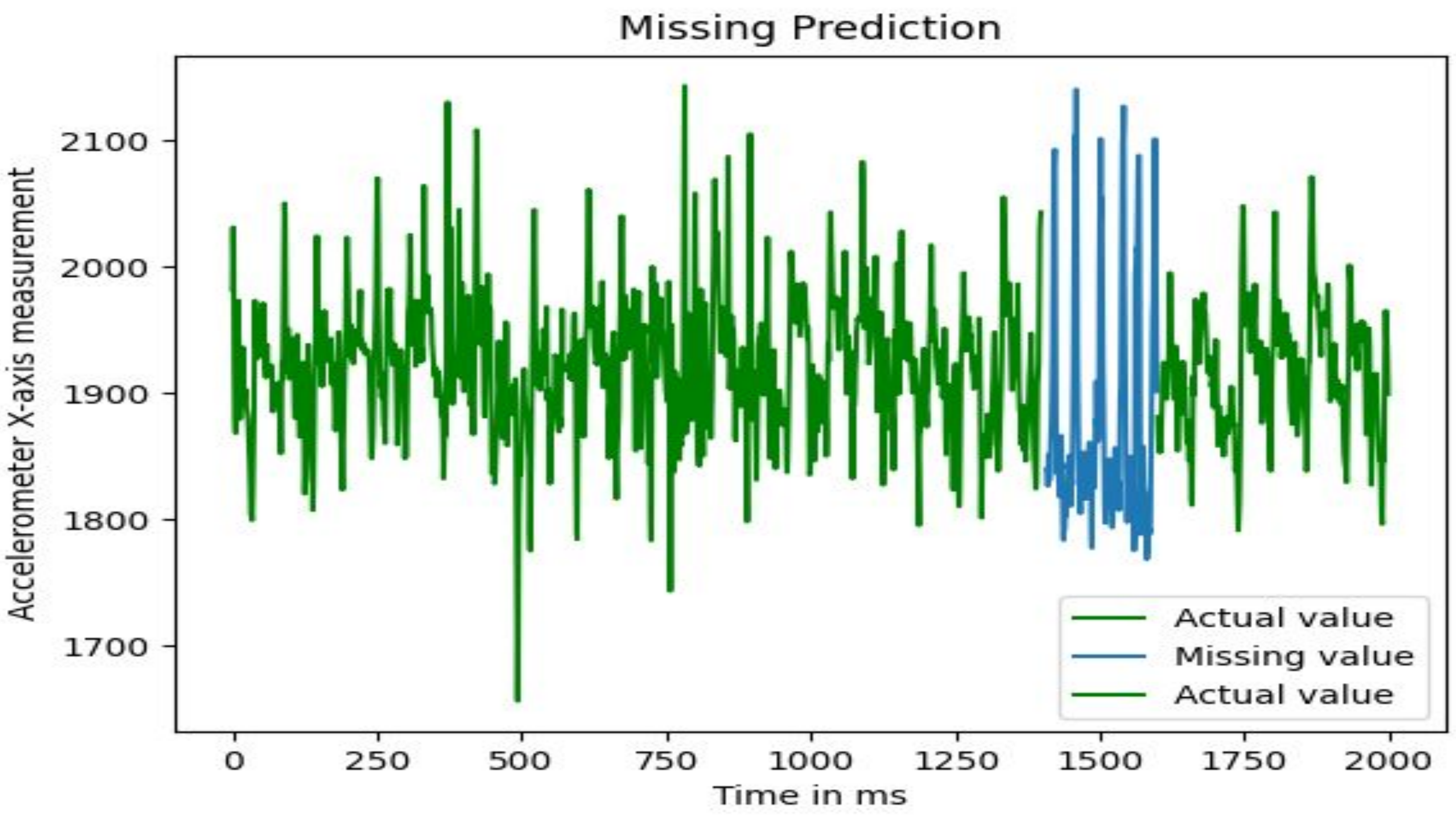
LSTM Future Prediction
2700
2600
2500
2400
2300
2200
Accelerometer X-axis measurement
0
250
500
750
1000
1250
1500
1750
2000
Time in ms
Actual Data
LSTM Train Data Prediction
LST Future Prediction



LSTM Future Prediction
Predicted value
Accelerometer X-axis measurement
Time in ms
2150
2100
2050
2000
1950
1900
1850
1800
1750
1600
1650
1700
1750
1800
1850
1900
1950
2000



Missing Prediction
2100
2000
1900
1800
1700
Accelerometer X-axis measurement
0
250
500
750
1000
1250
1500
1750
2000
Time in ms
Actual value
Missing value
Actual value


Missing Prediction
Accelerometer X-axis measurement
2100
2000
1900
1800
1700
0
250
500
750
1000
1250
1500
1750
2000
Time in ms
Actual value
Missing value
Actual value

Results

Train score: 19.60 RMSE

Test score: 20.22 RMSE

Further Work

- This work was done using accelerometer sensor data, but this could be extended to magnetometer and gyroscope data
- Subsequently, IMU sensor data could be used
- Although an LSTM model was used here, other models could lead to comparable results

Questions?