# Prediction of Diabetes Glucose Level Using Gated Recurrent Unit and Long-Short Term Memory Networks

**Tina Behrouzi**

2021 March 14

# Abstract

To maintain blood glucose in target range, people with diabetes need to adjust insulin doses based on their blood glucose levels. Knowing the estimation of blood sugar ahead of time helps people to be prepared and also reduces their daily stress. In this project, I developed two novel LSTM and GRU networks for predicting 60 minutes of blood sugar into the future. Two approaches are considered for each network: the first approach is predicting all 60 minute data at the same time. The second approach is estimating 5 minute later blood sugar and deriving the next 60 minute data based on previous data and prediction values. The methods are compared based on root mean square error and training time and all codes are implemented in Python using Keras and Pandas libraries.

**Keywords:** Blood glucose level, Recurrent Neural Network, Long-Short Term Memory, Gated Recurrent Unit.

# Table of Contents

# Lists of Figures

# List of Tables

# Introduction

A recurrent Neural Network (RNN) is widely applied for time-series prediction. RNNs consider the time dependency of the data and process it time-step by time-step. A Simple RNN works only well for short term dependency. The two previous RNN networks that were employed for predicting blood glucose are Dilated RNN Chen et al., 2018 and LSTM Martinsson et al., 2018. A Long-Short Term Memory (LSTM) is one of the most popular RNN networks that can learn the data structure regarding a long relevant history of a input Hochreiter and Schmidhuber, 1997. The LSTM network also overcomes the vanishing gradient problem of simple RNN network when employing Backpropagation. The LSTM consists of memory blocks that contain forget, input, and output gates. These gates are responsible for learning the importance of internal state, current data, and previous output on the output of each layer, respectively.

Gated Recurrent Units (GRU) Chung et al., 2014 is a newer generation of RNNs and its structure is very close to LSTM. Compared with LSTM, GRU has less parameters and contains two gates, update gate and reset gate. The reset gate is similar to LSTM forget gate, and the update gate is a combination of forget and input gates. Both LSTM and GRU networks are optimized for glucose level prediction, and they are compared based on accuracy and computational time.

We wish to submit the original research, ”Graph Variational Auto-Encoder for Deriving EEG-based Graph Embedding,” for consideration by Pattern Recognition journal. In this paper, the new graph-based learning algorithm for identifying users based on their EEG brain waves is proposed. Given the increasing interest in EEG-based identification and graph learning methods, we felt compelled to design a novel graph variational auto-encoder method to obtain brain connectives' features. This work addresses two of the main problems in EEG biometric: the robustness of an identification method to the human condition and the vulnerability of graphical methods to classify subjects considering a few numbers of EEG channels (small graphs). We believe that it will be interesting for readers to see how we overcame these problems and got high accuracy with low computational cost on three databases. This paper proposes a more advanced graph-based feature extraction method comparing to the ”BrainPrint: EEG Biometric Identification Based On Analyzing Brain Connectivity Graphs” approach, published by Pattern Recognition on 26 April 2020. We confirm that this manuscript is the authors' original work, and it is neither submitted to nor published in another journal or conference paper. Moreover, We have no conflict of interest to disclose. Thank you for your consideration of this manuscript.

# Material and Methods

## 2.1 Methodology

### 2.1.1 Preprocessing

Both LSTM and GRU memory blocks have sigmoid and tanh activation, which cause them to be highly sensetive to input scale. Therefore, the input data $X$ is normalized between 0 and 1, $\frac{X-min(X)}{max(X)-min(X)}$. The average of training data minimum and maximum is applied for normalizing testing and validation data.

The sliding window method is applied to create samples. To predict the next 60 minutes data, the previous 2 hours history of data is assigned to each sample.

Two RNN networks, LSTM and GRU, are investigated for glucose level prediction in this project. For each one of these networks, 3 model is designed. The first model only considers sugar levels and disregards activity and heart rate data. This model predicts future glucose based on 5 minutes point. The second model considers all features and estimates the future value of blood sugar, heart rate, and activity level again based on 5 minutes point. The third model predicts the future 60 minutes glucose level considering all features.

### 2.1.2 LSTM

The LSTM network first layer returns the output sequence and has a 24 point input. The unit number is set to the number of prediction data points. The second layer only returns the final outputs and its unit number is the same as first layer. The default sigmoid activation function is used for the LSTM blocks. The final layer is Dence layer to derive final prediction values.

For the 1 feature - 5 minutes model, the units number and number of output neurons are 1. For 3 features - 5 minutes model, the LSTM and Dence layer units number are 3 (heart rate, activity, and glucose level output). The output unit number is 12 (1 hour) for the 60 minutes model.

### 2.1.3 GRU

The same structure as LSTM model is considered for GRU model. For all models, the first GRU layer returns the output sequence and the final GRU layer only returns the predicted value. However, for the 60 minutes model, the second GRU layer is added to the model which returns the time-sequence like the first layer. The final Dence layer returns the corresponding output. The unit number and number of the Dence layer output are set to required data point; which is 1, 3, and 12 for 1 feature - 5 minutes, 3 feature - 5 minutes, and 60 minutes models, respectively.

## 2.2 Dataset

The dataset consists of three recordings: blood glucose measurement in mg/dL, subject's heart rate, and activity measurment in kilometers. The glucose level samples were taken every 5 minutes; however, the activity and heart rate data recording time interval is not exactly 5 minutes and they are not collected at the same time. If we merge all three data based on blood glucose timestamp, there will be 14380 and 14447 missing data for activity and heart samples, respectively. To overcome this problem, samples' minutes plus rounded seconds are rounded every five minutes $round(\frac{min+round(sec)/60}{4.5}) * 5$. Additionally, more than one sample in five minutes interval is dropped. The number of missing heart and activity data reduces to 362 and 2040, respectively.

Starting time to collect heart and activity is one day and several hours later than glucose data. For prediction based on all 3 features, the blood sugar data until first heart rate recording is disregarded. Moreover, the linear interpolation is applied to cover for the rest of the missing heart and activity data. As a result, the merged data will only have 362 NaN values.

The first two weeks and the next one week of the data are considered for training and validation data, respectively. The rest of the data is assigned to testing data.

### 2.2.1 Implementation

The reported computation time is based on CPU @2.5 GHz Intel Core i7 processor. All codes are implemented in Python 3.6.

The Adam optimizer with the Mean Square Error (MSE) is applied for optimization Kingma and Ba, 2014. For all models, early stopping with a 15 patience value is considered. Learning rate is set to 0.001 (learning rates $e^{-4}, 0.5e^{-3}, e^{-3}, 0.5e^{-2}, e^{-2}$ were also tested but $e^{-3}$ had the best result for all models).

The Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are considered for evaluating and comparing methods.

# Results and Discussion

| Method | #features | MAE | RMSE | #epochs | CT/epoch(s) |
|---|---|---|---|---|---|
| LSTM | 3 | 4.125 | 7.854 | 200 | 3 |
| | 1 | 77.796 | 100.148 | 169 | 2 |
| GRU | 3 | 3.513 | 6.653 | 118 | 3 |
| | 1 | 77.245 | 99.519 | 62 | 2 |

Table 3-1: The LSTM and GRU error and computational time (CT) for five minutes glucose level future prediction

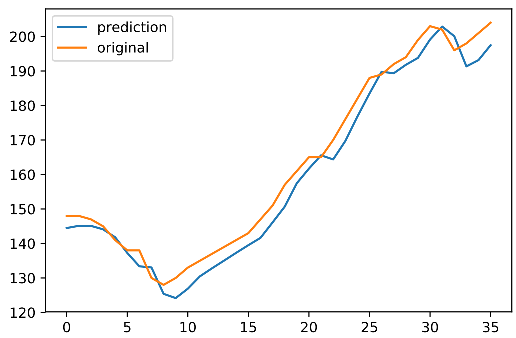


Figure 3-1: LSTM - 3 features

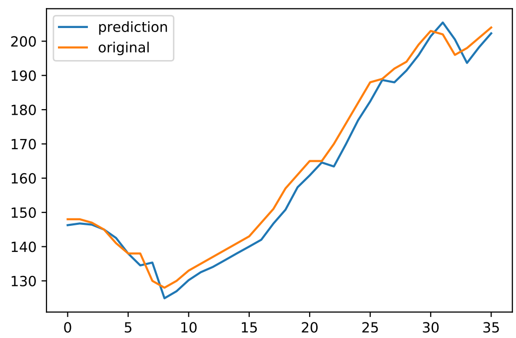


Figure 3-2: GRU - 3 features

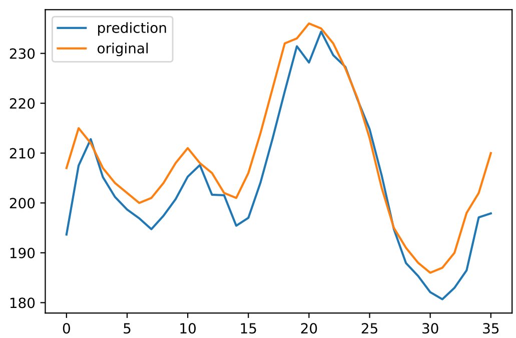


Figure 3-3: LSTM - 1 feature

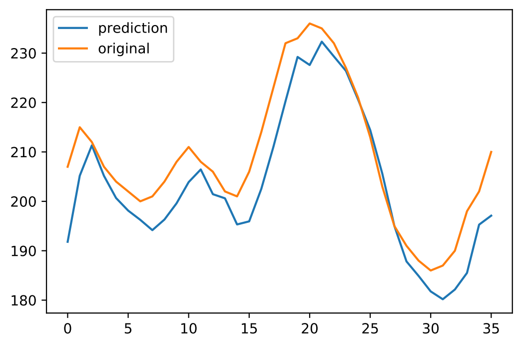


Figure 3-4: GRU - 1 feature

Figure 3-5: 3 hours comparison of the true value and the corresponding prediction value for each 5 min points

Table 3-1 and Fig. 3-5 indicate that the blood glucose prediction based on only sugar level is not as accurate as 3 features model that also considers heart rate and activity data. Moreover, the GRU model computational cost and error rate is less than LSTM model. The ability to predict the next 60 minutes based on previous prediction for GRU model is more accurate than

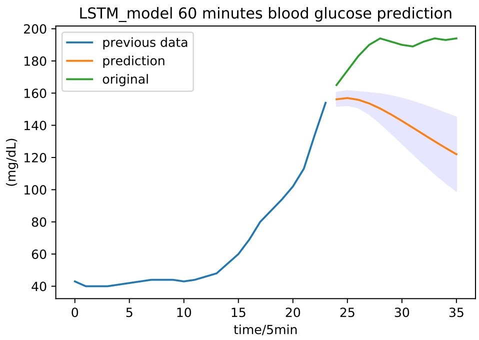


Figure 3-6: LSTM - 3 features

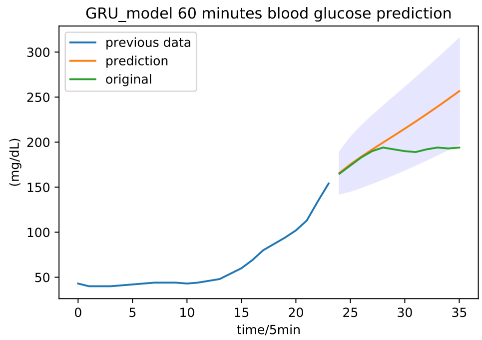


Figure 3-7: GRU - 3 features

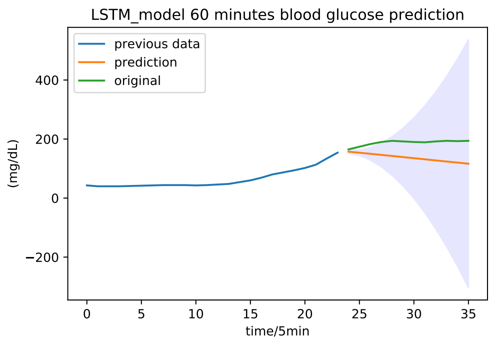


Figure 3-8: LSTM - 1 feature

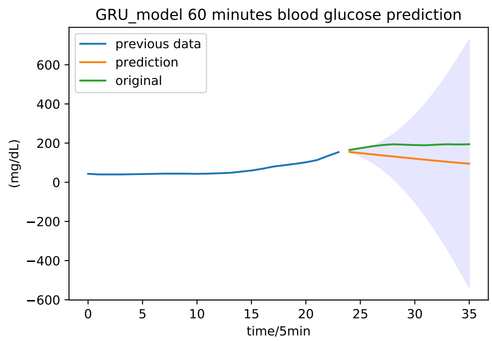


Figure 3-9: GRU - 1 feature

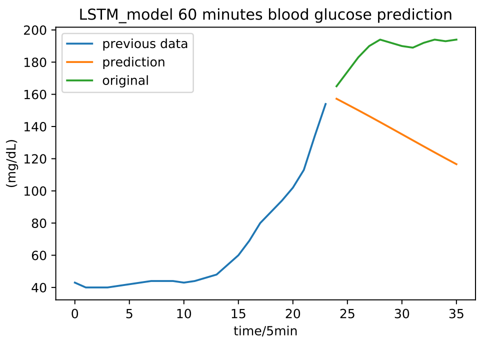


Figure 3-10: LSTM - 1 feature

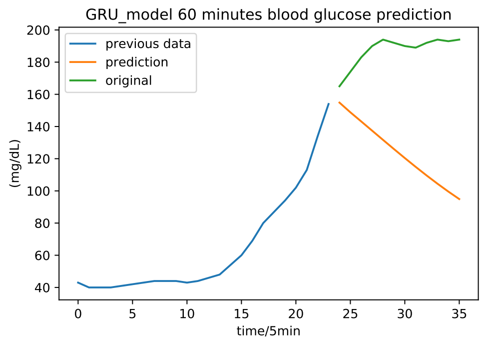


Figure 3-11: GRU - 1 feature

Figure 3-12: Using five minute prediction for 60 minutes prediction. The blue shade illustrates the confidence interval other models based on Fig. 3-12. The 1 feature model fails to predict the next 60 minutes pattern. Therefore, all 3 features are required for predicting future glucose level, and GRU model is more accurate and has less cost compared to LSTM model for this data.

Because blood glucose prediction for 1 feature model resulted in higher MAE and RMSE error, the 1 feature based future sugar level estimation is not derived for 60 minutes based prediction. Considering Fig 3-15 and Table 3-2, the 60 minutes prediction model is more

| Method | MAE | RMSE | #epochs | CT/epoch(s) |
|---|---|---|---|---|
| LSTM | 16.974 | 26.943 | 110 | 2 |
| GRU | 16.206 | 26.423 | 66 | 2 |

Table 3-2: The LSTM and GRU error and computational time for five minutes glucose level future prediction

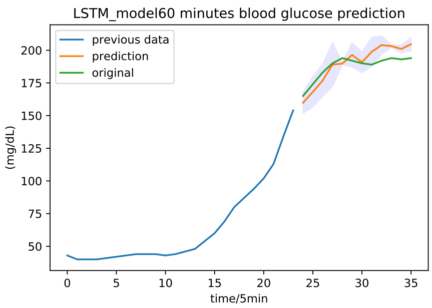


Figure 3-13: LSTM - 3 features

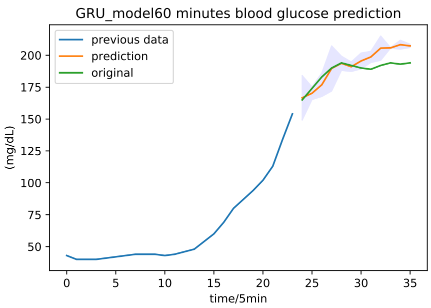


Figure 3-14: GRU - 3 features

Figure 3-15: Using five minute prediction for 60 minutes prediction

accurate than the other two models that predict sugar level based on 5 minutes steps. The 60 model was about to estimate glucose pattern more accurately. Additionally, for all models, the GRU network resulted in less error with less computational cost compared to LSTM model.

# Conclusion

Six novel models for predicting blood glucose level are proposed. Three of them are based on LSTM network and other three are GRU networks. For each network, the impact of considering heart rate and activity data and also difference between 5 and 60 minutes point prediction are evaluated. We derived that for accurate prediction, all 3 features are required. Moreover, The GRU model resulted in less computational cost with less error compared to LSTM model.

In the future, the power of these models for more number of subjects will be evaluated.

# References

Chen, J., Li, K., Herrero, P., Zhu, T., & Georgiou, P. (2018). Dilated recurrent neural network for short-time prediction of glucose concentration. *KHD@ IJCAI*, 69–73.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Martinsson, J., Schliep, A., Eliasson, B., Meijner, C., Persson, S., & Mogren, O. (2018). Automatic blood glucose prediction with confidence using recurrent neural networks. *KHD@ IJCAI*.

# Appendix

## A.1 Data

| point_value(mg/dL) | point_timestamp | new_date |
|---|---|---|
| 144 | 2017-05-16 22:18:50 | 2017-05-16 22:20 |
| 144 | 2017-05-16 22:23:40 | 2017-05-16 22:25 |
| 132 | 2017-05-16 22:28:40 | 2017-05-16 22:30 |
| 124 | 2017-05-16 22:33:40 | 2017-05-16 22:35 |
| 124 | 2017-05-16 22:38:39 | 2017-05-16 22:40 |

| point_value(kilometers) | point_value |
|---|---|
| 0.0 | 110.0 |
| 0.0 | 85.0 |
| 0.0 | 85.0 |
| 0.0 | 90.0 |
| 0.0 | 82.0 |

Figure A-1: Merged data first 5 rows

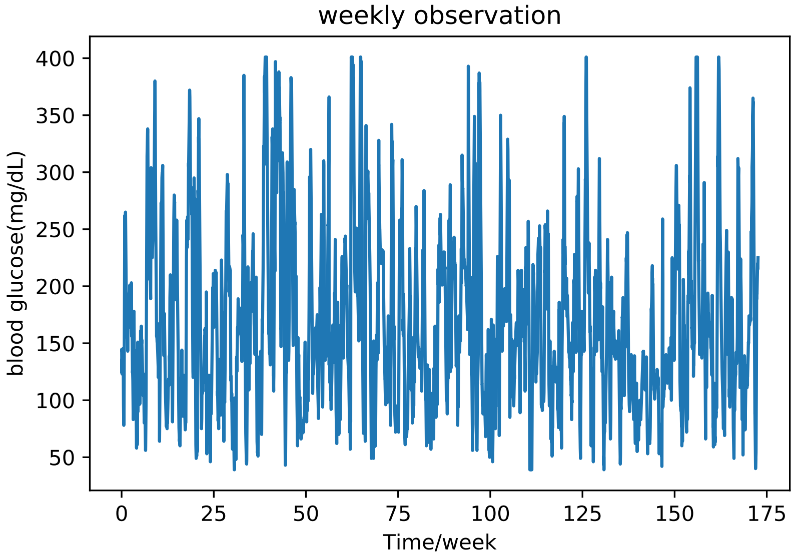


Figure A-2: Glucose level data

Fig. A-3 and A-2 show that the subject's glucose level does not have a daily or weekly pattern.

## A.2 Loss

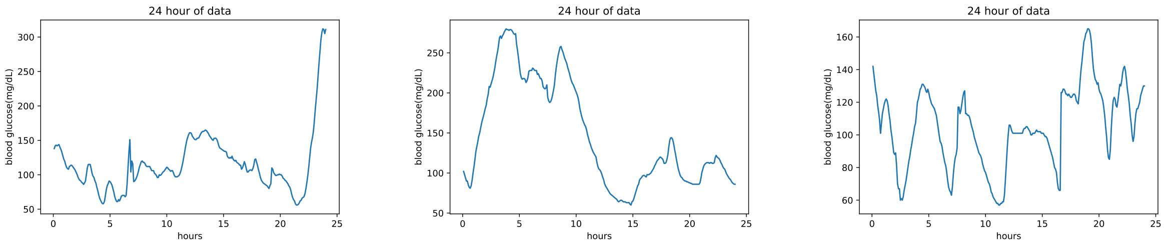


Figure A-3: Blood sugar level in different days

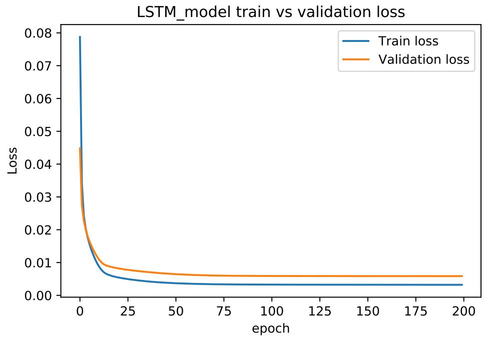


Figure A-4: LSTM - 3 features

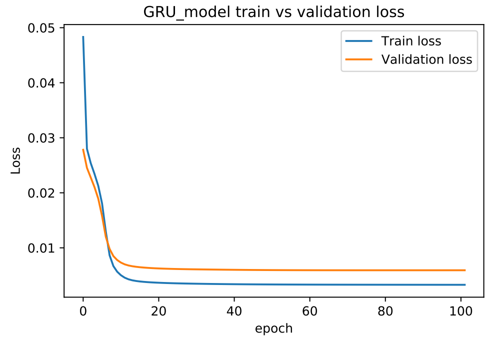


Figure A-5: GRU - 3 features

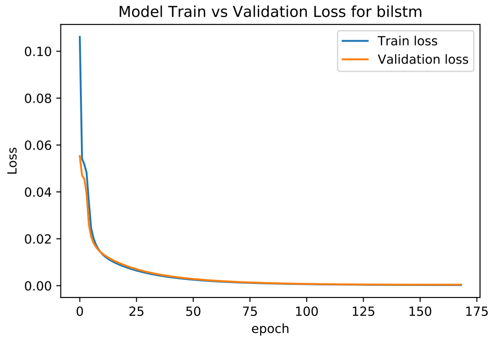


Figure A-6: LSTM - 1 feature

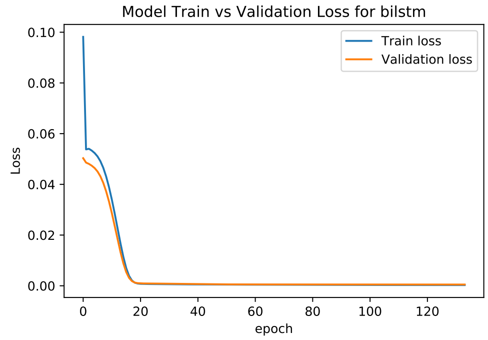


Figure A-7: GRU - 1 feature

Figure A-8: Training and validation loss for 5 minutes ahead blood sugar prediction

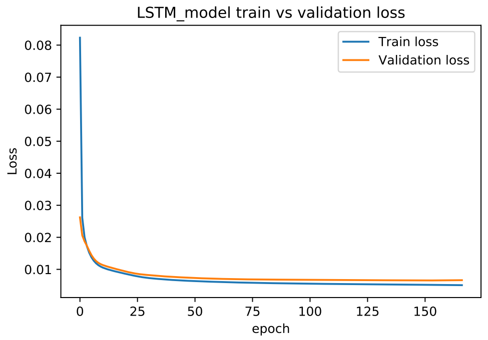


Figure A-9: LSTM

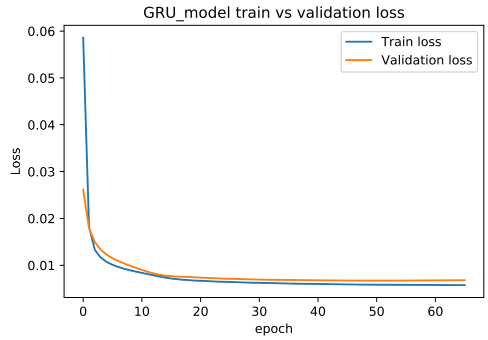


Figure A-10: GRU

Figure A-11: Training and validation loss for 60 minutes ahead blood sugar prediction

## A.3 Files

There are 3 files in code folder. ”GRU_LSTM_5min_1feature.py” corresponds to 5 minutes future data prediction using only blood suger data. ”GRU_LSTM_5min_3features.py” is python code for 5 minute glucose level prediction using all 3 features. ”GRU_LSTM_60min.py” corresponds to 60 minutes sugar level prediction using all 3 features. The codes in each file consists of comments for clarification.

### A.3.1 Required Libraries:

- pandas
- tensorflow
- sklearn
- matplotlib
- numpy
- copy (optional - for plotting prediction confidence interval)