SugaRNN: Blood Glucose Predictions

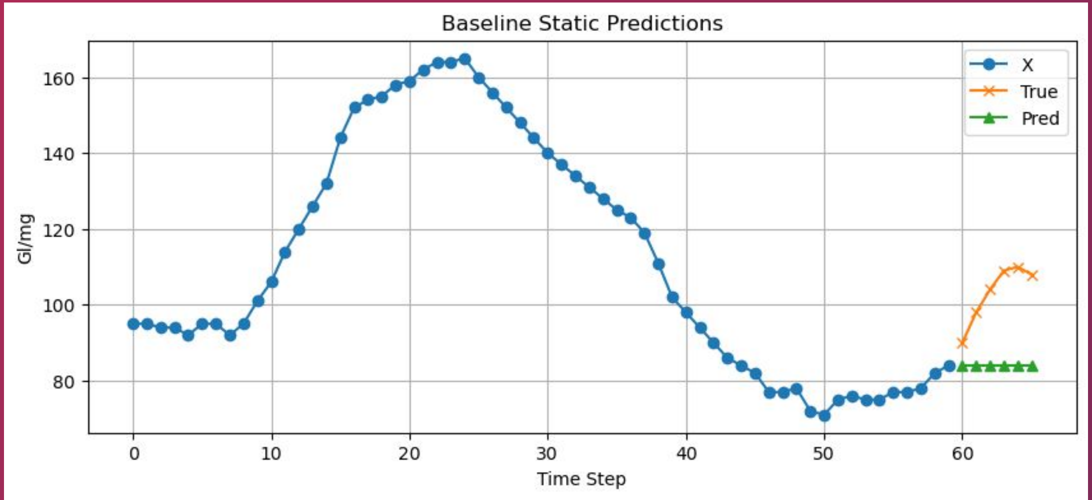
Andrew Plattel, Asa Adomatis, and Tucker Mackie

Introduction

- We want to use historic blood glucose (BG) values to predict a 30 minute window of future BG values
- Using the Hall dataset from the IrinaStatsLab GitHub repository, includes BG levels every 5 minutes for 57 different individuals, along with other descriptive variables
- We explored the use of RNNs, CNNs, and Time Series methods in predicting this window
- The desired result of our research is to hopefully progress the prediction and treatment of BG levels for diabetics

	id	time	gl	Age	BMI	A1C	FBG	ogtt.2hr	insulin	hs.CRP	...	freq_low	freq_moderate	freq_severe	glucotype	Height	Weight	Insulin_rate_dd	perc_cgm_prediabetic_range	perc_cgm_diabetic_range	SSPG
0	1636-69-001	2014-02-03 03:42:12	93.0	59.0	21.7	6.7	109.0	205.0	9.0	0.3	...	0.147059	0.369748	0.483193	2	176.3	68.0	0.1015	0.190404	0.026211	91.0
1	1636-69-001	2014-02-03 03:47:12	93.0	59.0	21.7	6.7	109.0	205.0	9.0	0.3	...	0.147059	0.369748	0.483193	2	176.3	68.0	0.1015	0.190404	0.026211	91.0
2	1636-69-001	2014-02-03 03:52:12	93.0	59.0	21.7	6.7	109.0	205.0	9.0	0.3	...	0.147059	0.369748	0.483193	2	176.3	68.0	0.1015	0.190404	0.026211	91.0
3	1636-69-001	2014-02-03 03:57:12	95.0	59.0	21.7	6.7	109.0	205.0	9.0	0.3	...	0.147059	0.369748	0.483193	2	176.3	68.0	0.1015	0.190404	0.026211	91.0
4	1636-69-001	2014-02-03 04:02:12	96.0	59.0	21.7	6.7	109.0	205.0	9.0	0.3	...	0.147059	0.369748	0.483193	2	176.3	68.0	0.1015	0.190404	0.026211	91.0

Baseline Model



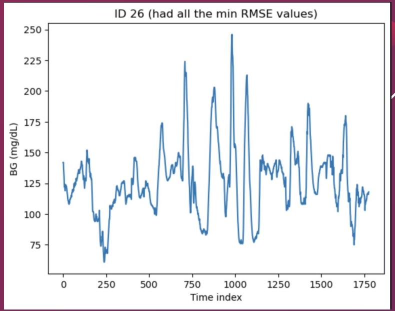
- Compare forecasting of other models to a model that doesn't predict any change
- **MAE:** 6.12 mg/dL
- **RMSE:** 10.12 mg/dL

RNN – Methods

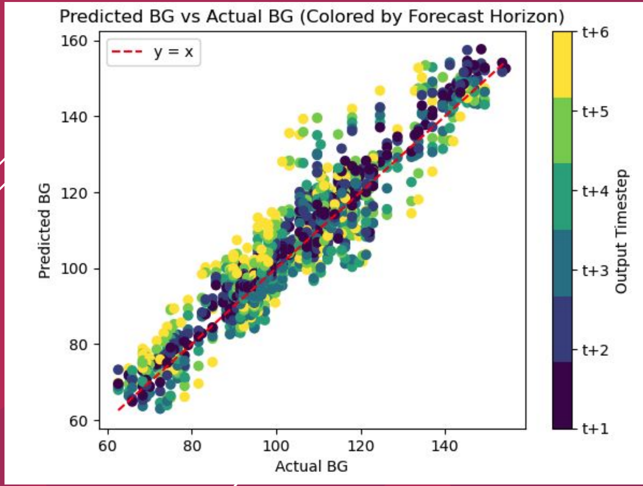
- Data transformation:
  - Convert `time` to datetime format and set it as the index
  - Create a dictionary that separates the data into 57 different dataframes, 1 for each ID
  - Create a normalized dictionary where we extract the `gl` variable, remove `NaN` values, and normalize using `MinMaxScaler` and `fit_transform`
- Model Design:
  - 2 hidden LSTM layers with 50 units, ReLU activation
  - Compiled the model with the ADAM optimizer and MSE for loss
  - Model inputs: number of sequences in and number of sequences out
    - Each sequence is about a 5 minute window
- Model Implementation:
  - Created a function that takes the ID_subset, sequences in and sequences out
  - Then splits the subset 80/20 for training/testing and reshapes the data to be usable with the RNN
  - Then makes predictions for each ID, calculates RMSE for each iteration along with the combined for that model, and outputs an accuracy plot of each prediction in a 30 minute window (6 predicted points)

RNN – Results

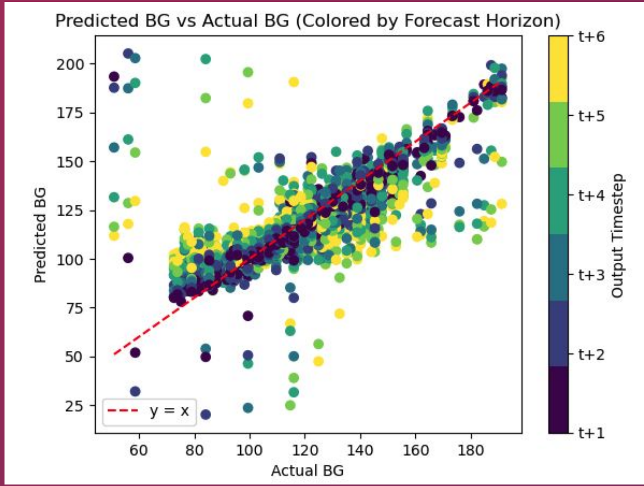
- Ran the models (total of 228) using 4 different sequences; 70, 60, 24, 12, 6
  - Respective RMSE values: 12.82 mg/dL, 12.70 mg/dL, 12.84 mg/dL, 12.26 mg/dL, 12.87 mg/dL
  - Respective MAE values: 8.45 mg/dL, 8.32 mg/dL, 8.44 mg/dL, 8.01 mg/dL, 8.53 mg/dL
  - 12 sequences (1 hour) of training performed the best in predicting a 30 minute window
  - Interestingly ID 26 had the minimum RMSE for each of the sequences
    - Likely due to smother variance in data values and less spikes/dips

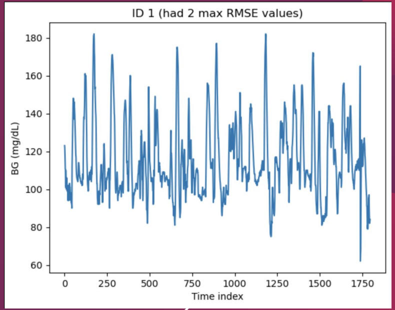


ID 26, trained on 12 sequences: RMSE of 6.43 mg/dL



ID 1, trained on 12 sequences: RMSE of 18.2 mg/dL

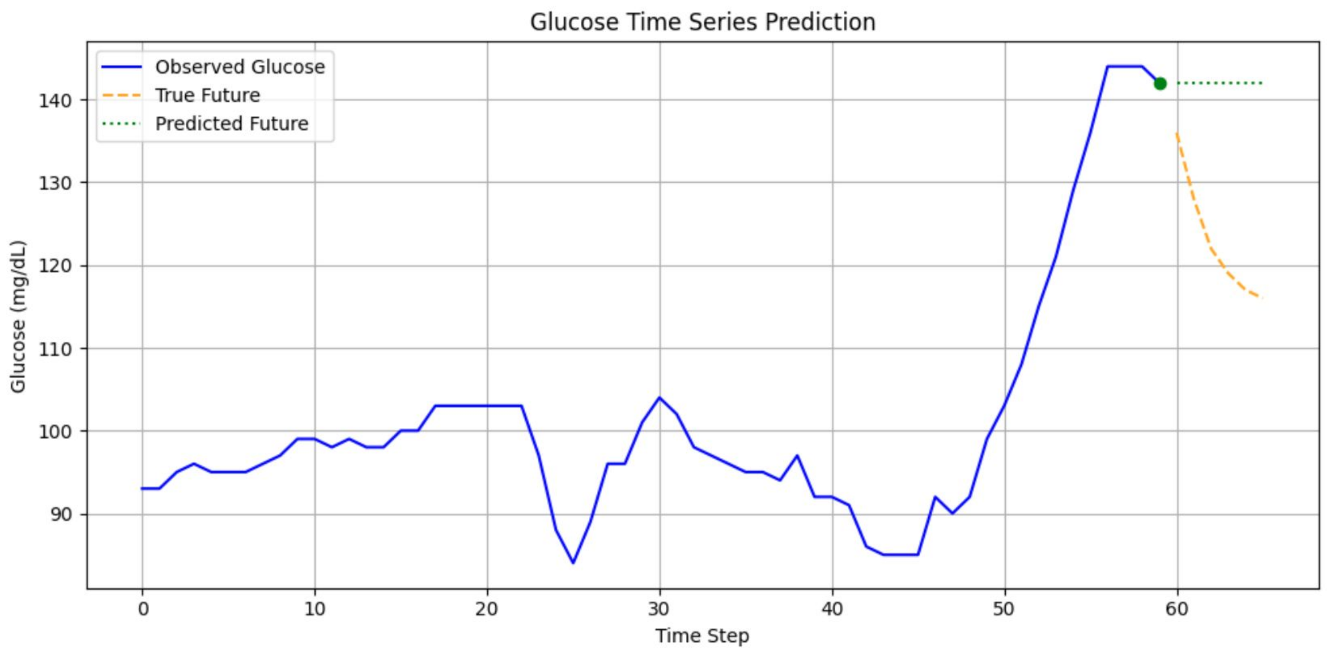




Time Series - Methods

- **Data Transformation:**
  - Create a sliding window design
  - Taking previous five hours of data points to predict the next 6 intervals - 30 minutes
  - Creates supervised input-output pairs where each input predicts the next future window
- **Model Design:**
  - Persentance Model Naive Forecasting was not built for learned behaviors or adding layers similar to CNN/RNN
  - Random Forest / ARIMA are not strong fits due to data constraints
- **Training:**
  - Train/Test split of 80/20 using chronological data
  - With no training models, the Persentance Model outputs the final results based off the previous 60 points

Time Series - Results

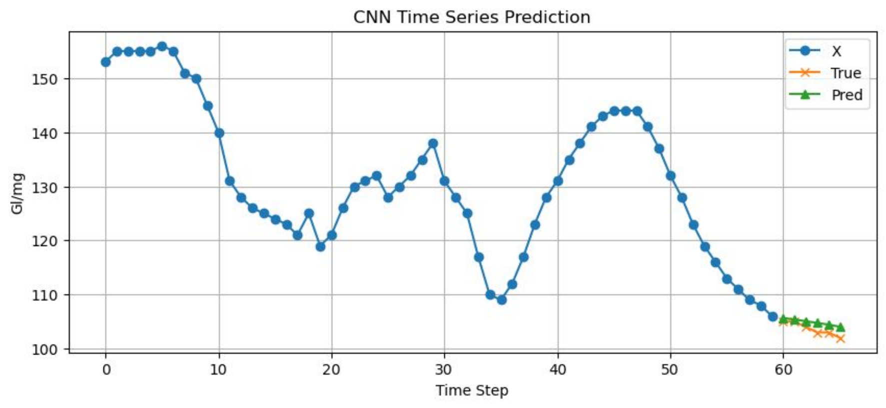


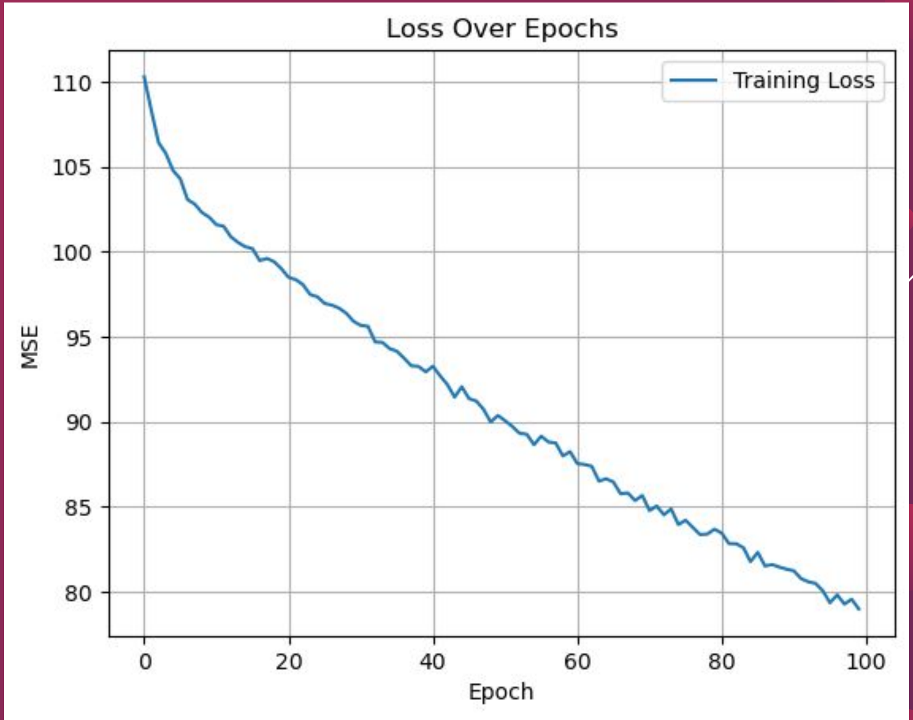
- **MAE:** 6.14 mg/dL
- **RMSE:** 10.59 mg/dL

CNN Model – Methods

- **Data Transformation:**
  - Created sequences representing 5 hours of previous glucose levels and sequences representing the next 30 minutes of blood glucose levels
  - Split time series blood glucose data and static, biometric data into separate inputs into the model
  - Added a column to the static data representing “volatility,” amount of times BG changed more than a threshold in the span of 5 minutes
  - Predict delta, change from last point in the sequence, rather than real values
- **Model Design:**
  - For time series input, 2 pairs of 1D convolution and 1D max pooling layers
  - For static input, two dense layers with 64 and 128 units
  - Once merged, two dense layers each with 256 units
  - All layers used ReLU activation functions
- **Training:**
  - Optimizer - ADAM
  - Loss - MSE

CNN Model – Results





- **MAE:** 6.36 mg/dL
- **RMSE:** 9.87 mg/dL
- RMSE better than no change

Conclusion

Results Summarized:

Model Type	No-Change	CNN-LSTM[1]	LSTM[2]	RNN	Time Series	CNN
RMSE (mg/dL)	10.17	9.81	6.45	**6.43** (12.26)	10.59	9.87
MAE (mg/dL)	6.12	5.65	N/A	**4.23** (8.01)	6.14	6.23

1) Mehrad Jaloli, MSc, and Marzia Cescon, PhD
2) Rabby, M.F., Tu, Y., Hossen, M.I. et al.

THANKS

Any questions?