Predicting the temporal evolution of patient parameters in a diabetic population using Recurrent Neural Networks

Supervisors

Prof Emanuele Trucco, University of Dundee

Dr Alex Doney , University of Dundee

Dr Prathiba, MDRF, Chennai, India

PhD Student

Jyothsna Divyananda

Funded by NIHR

Introduction

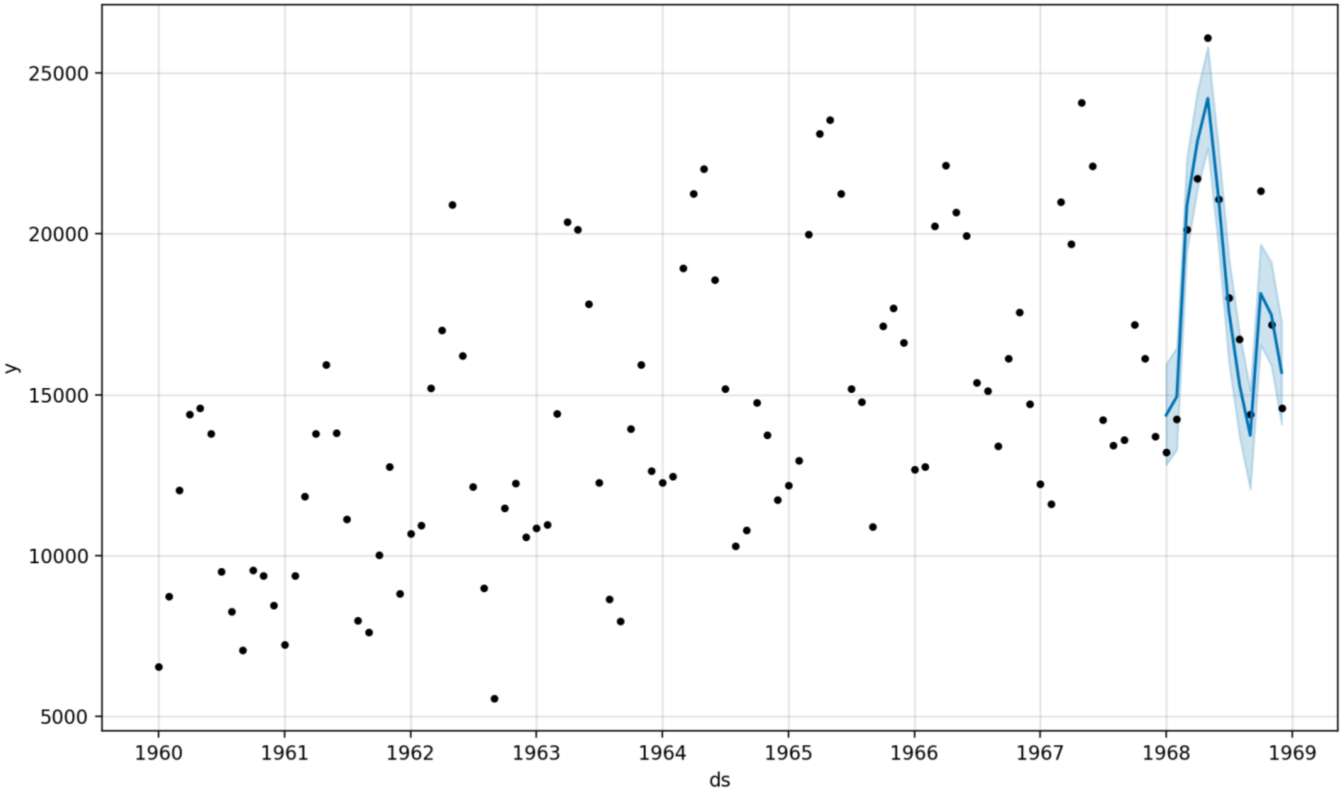


Fig 1: Example of Time series (1)

- Time series data is a sequence of observations collected from a process with equally spaced periods of time.
- Examples: count of sunspots, Ocean tides, weather recordings, medical records, Daily sales records, stock market etc.
- Applications of Time Series Analysis: Budget analysis, economic forecasting, sales forecast, workload projections etc

Recurrent Neural Network

- Recurrent Neural Networks(RNN) are used for time series forecasting
- We are employing stacked Long Short-Term Memory (LSTM) [1]
- LSTM's are used for NLP, TSF, speech recognition, handwriting recognition etc.

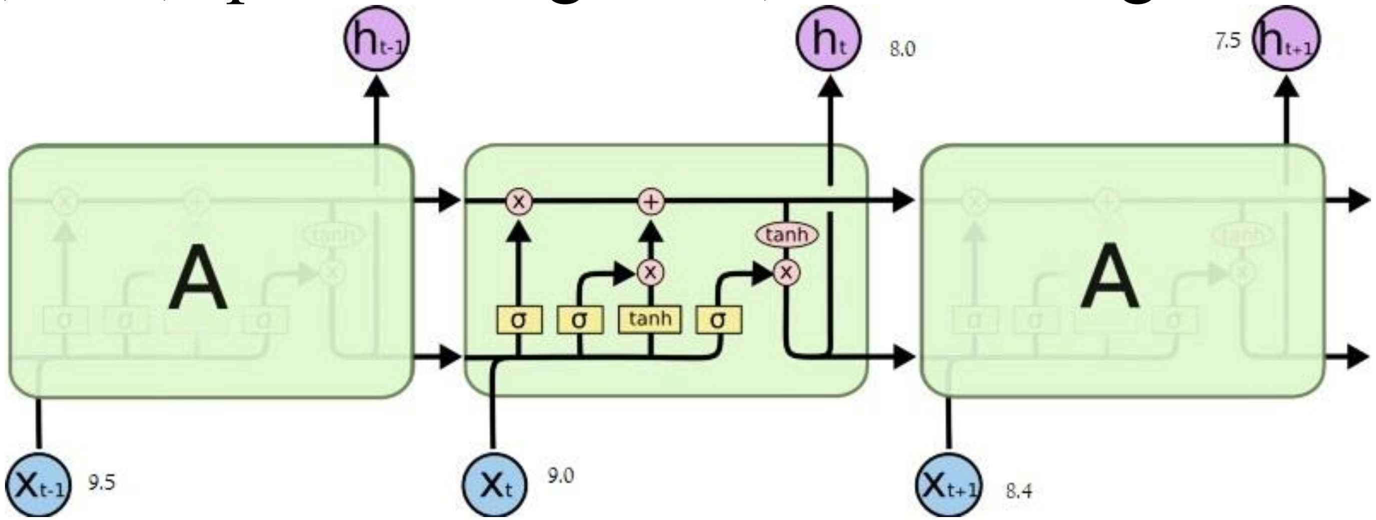


Fig: Basic structure of LSTM

1.Hochreiter S, Schmidhuber J. Long short-term memory. Neural computation. 1997 Nov 15;9(8):1735-80.

Aims and Challenges

- Aims:
  - Predict progression of a single physiological parameter (HbA1c) by univariate and multivariate forecasting
  - Predict progression of sets of physiological parameters (e.g., starting with hdl, non hdl, HbA1c, body mass index)
  - Predict physiological parameters of diabetic participants given specific drugs

- Challenges
  - Time series data maybe / are short
  - Requirement of large amount of data to train the models
  - Missing data

Data

- GoDARTS longitudinal Bioresource [1]
- Hdl, non hdl, BMI and HBA1c
- Population on diet

1. Harry L Hébert, Bridget Shepherd, Keith Milburn, Abirami Veluchamy, Weihua Meng, Fiona Carr, Louise A Donnelly, Roger Tavendale, Graham Leese, Helen M Colhoun, Ellie Dow, Andrew D Morris, Alexander S Doney, Chim C Lang, Ewan R Pearson, Blair H Smith, Colin N A Palmer, Cohort Profile: Genetics of Diabetes Audit and Research in Tayside Scotland (GoDARTS), *International Journal of Epidemiology*, Volume 47, Issue 2, April 2018, Pages 380–381j

Data Pre-processing

- Unevenly spaced data
- Spline interpolation
- Requirement of equal number of time steps

1. HbA1c, Lipids, BMI – input variables
2. Subsets of Patients based on diet block generated
3. Subset generated based on dates in the drug block csv
4. Subset of patients with more than 6 measurements from each variable selected and subsets generated
5. Spline interpolation of data
6. Subset of patients with evenly spaced measurements acquired.

Experiment 1
Univariate Forecasting

Data

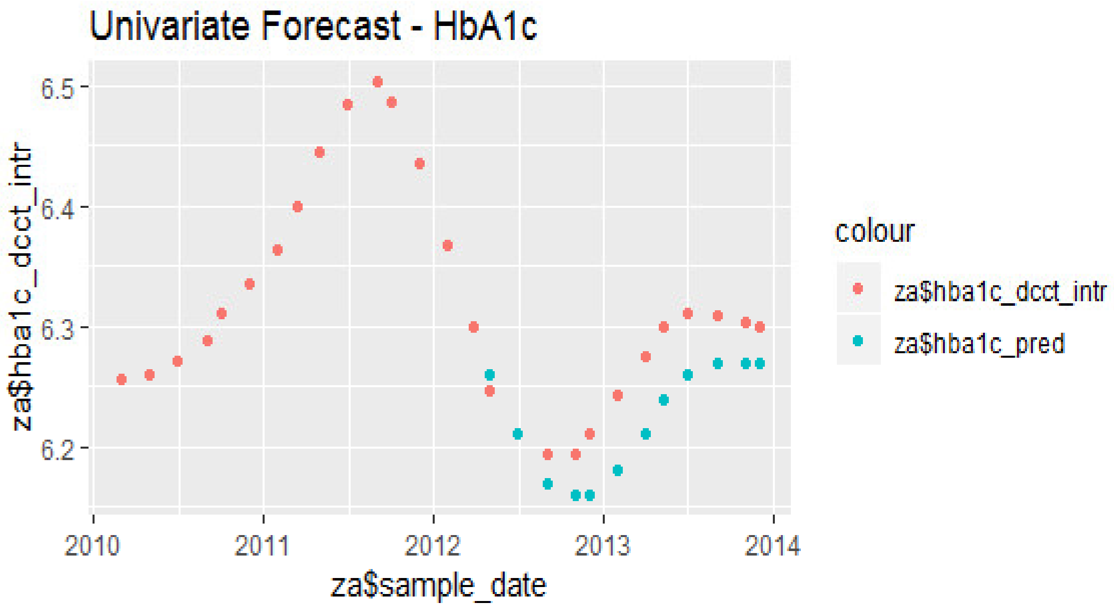
- Experiment 1
  - Patients on diet are chosen.
  - Number of patients: 1200
  - We select 5 year period for each patient with 30 measurements placed 2 months apart.
  - Feature selected: HbA1c

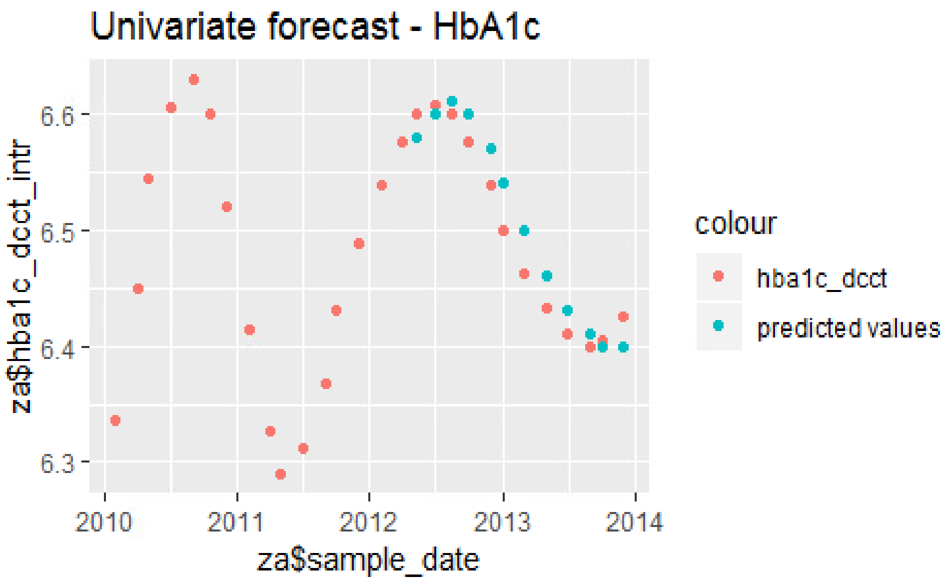
- Data is randomly divided into training 90% and testing 10%

LSTM Training

- Deep learning framework: Keras (TF backend)
- Number of training epochs: 150
- Optimizer: Adam (learning rate=0.0001)
- Loss: Mean Squared Error
- Metrics: Mean absolute error, Root mean square error
- Training strategies:
  - Model with best validation performance is saved

Results

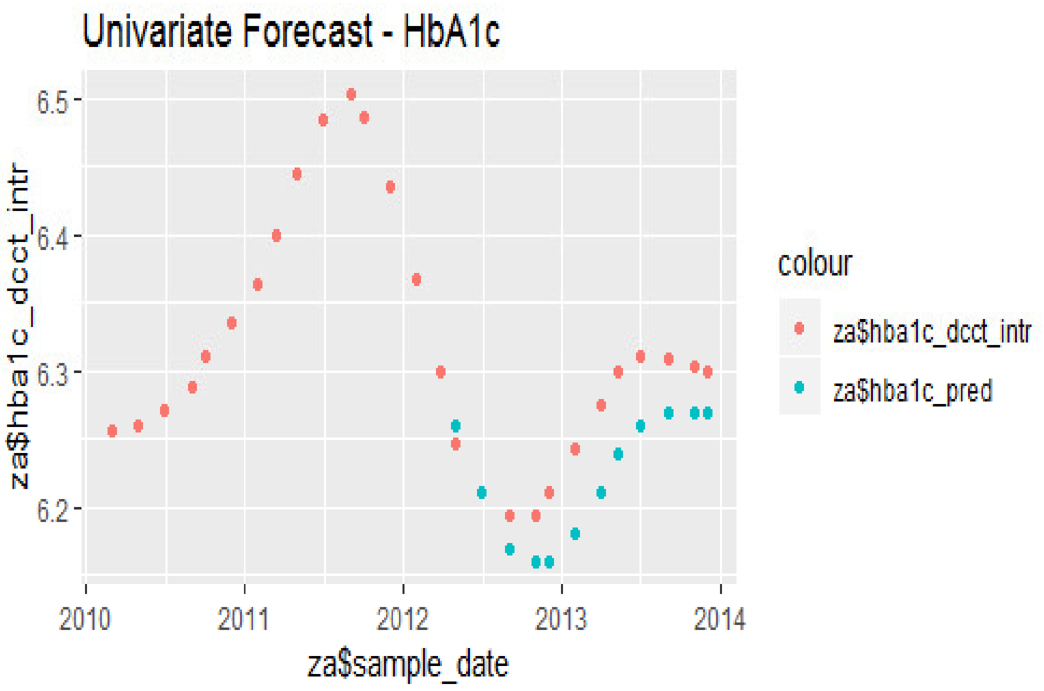




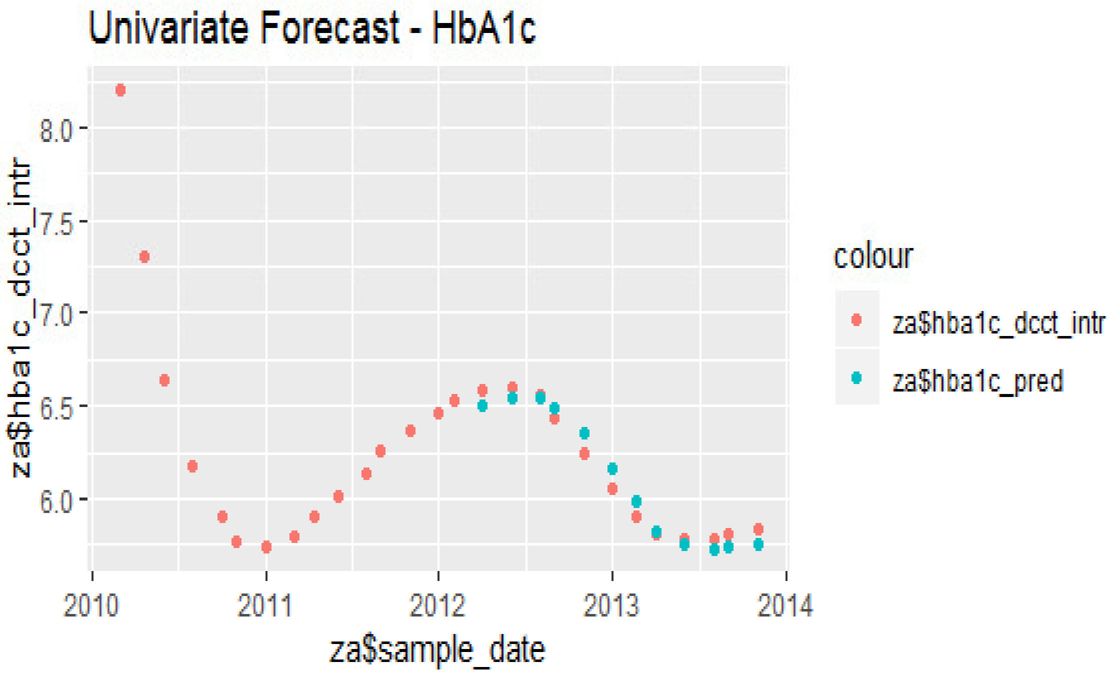
MAE:0.043, RMAE:0.039

MAE:0.043, RMAE:0.039

Results

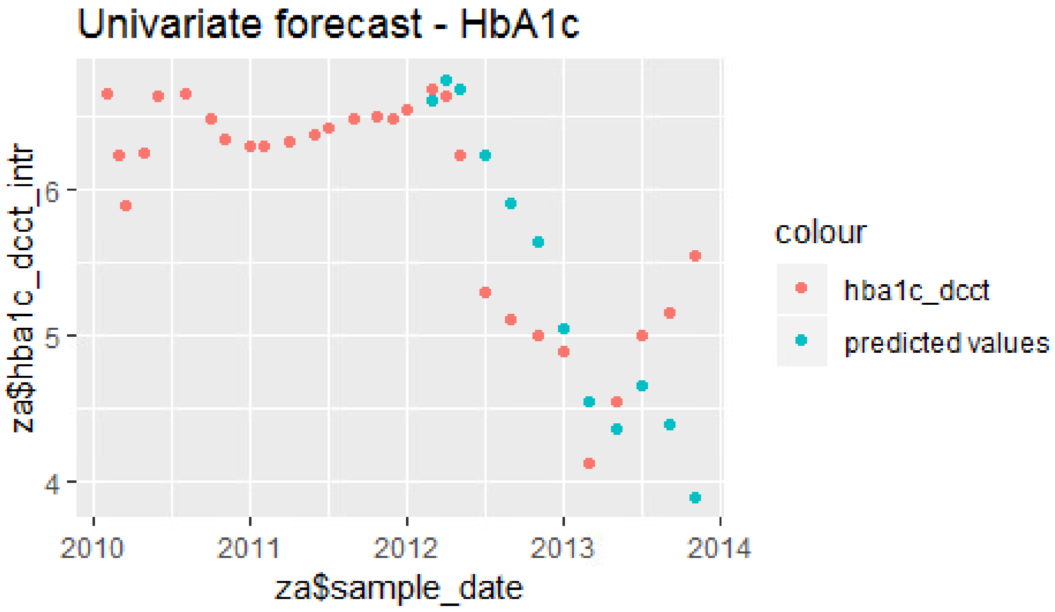


MAE:0.043, RMSE:0.039

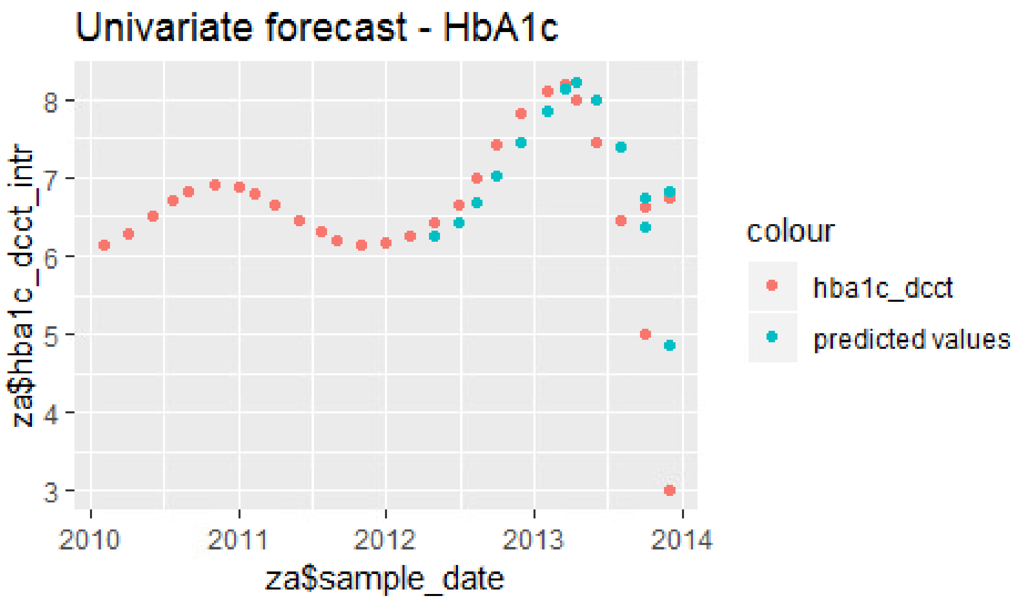


MAE:0.064, RMSE:0.072

Results

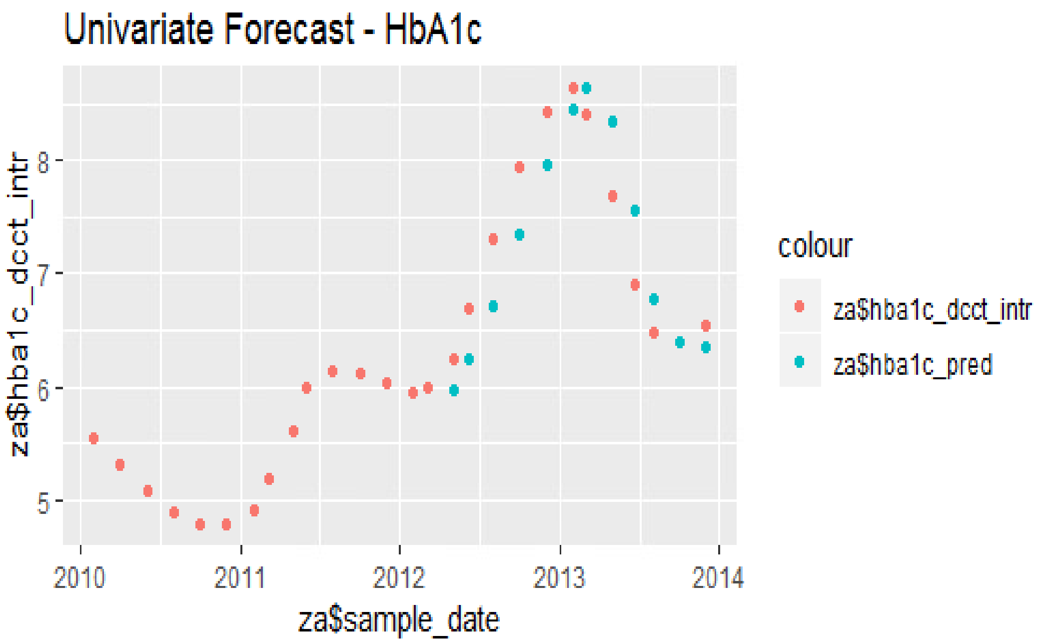


MAE: 0.41, RMSE:0.551

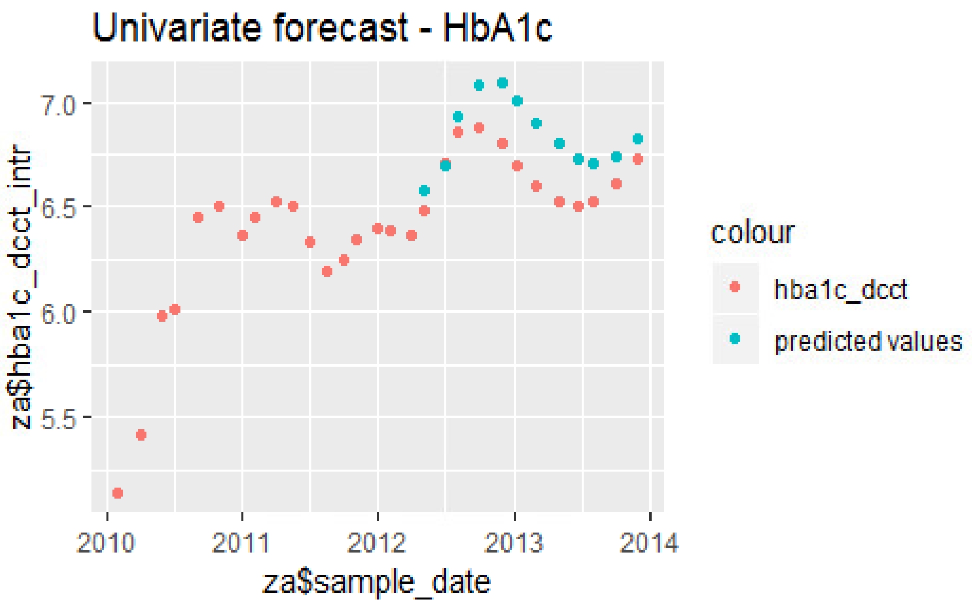


MAE: 0.555, RMSE:0.763

Results



MAE: 0.383, RMSE:0.435



MAE: 0.182, RMSE:0.207

Related work

- Data Driven Patient-Specialized Neural Networks for Blood Glucose Prediction, 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)
  - OhioT1DM, blood glucose level values sampled every 5 minutes, for about two months of observation
  - Dataset consists 6 patients (two men and four women)
- Kang S. Personalized prediction of drug efficacy for diabetes treatment via patient-level sequential modeling with neural networks. Artif Intell Med. 2018 Apr;85:1-6. doi: 10.1016/j.artmed.2018.02.004. Epub 2018 Feb 23. PMID: 29482961.
  - EMR from Seoul National University Hospital, 3169 T2DM patients between years 2003 – 2013

- https://ieeexplore.ieee.org/document/9105950
https://pubmed.ncbi.nlm.nih.gov/29482961/

Experiment 2

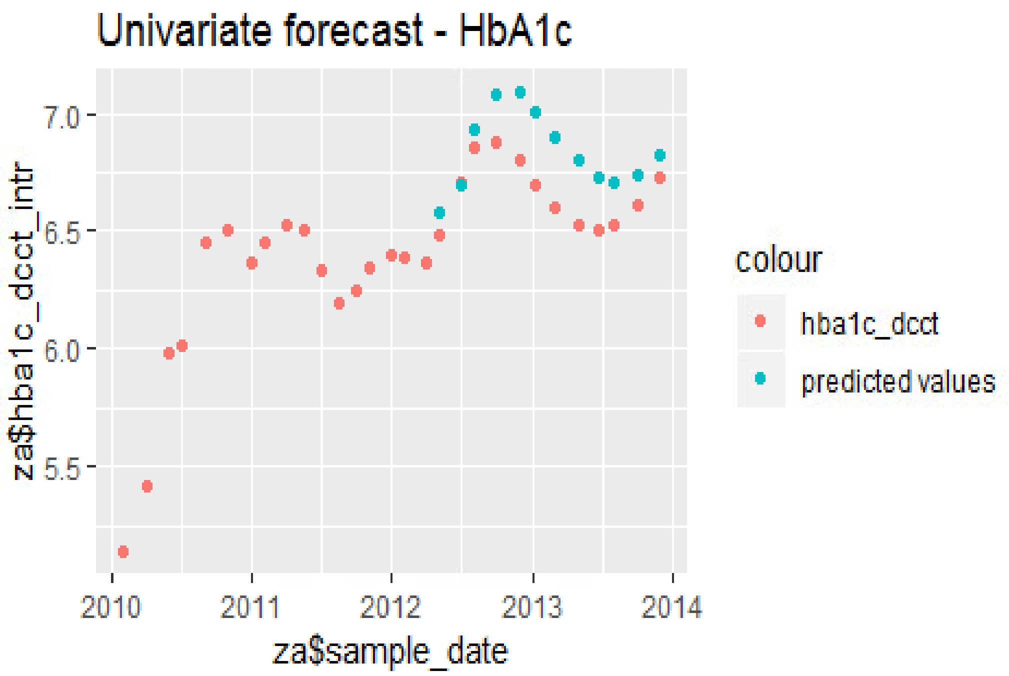
Univariate Forecasting with varied years of training

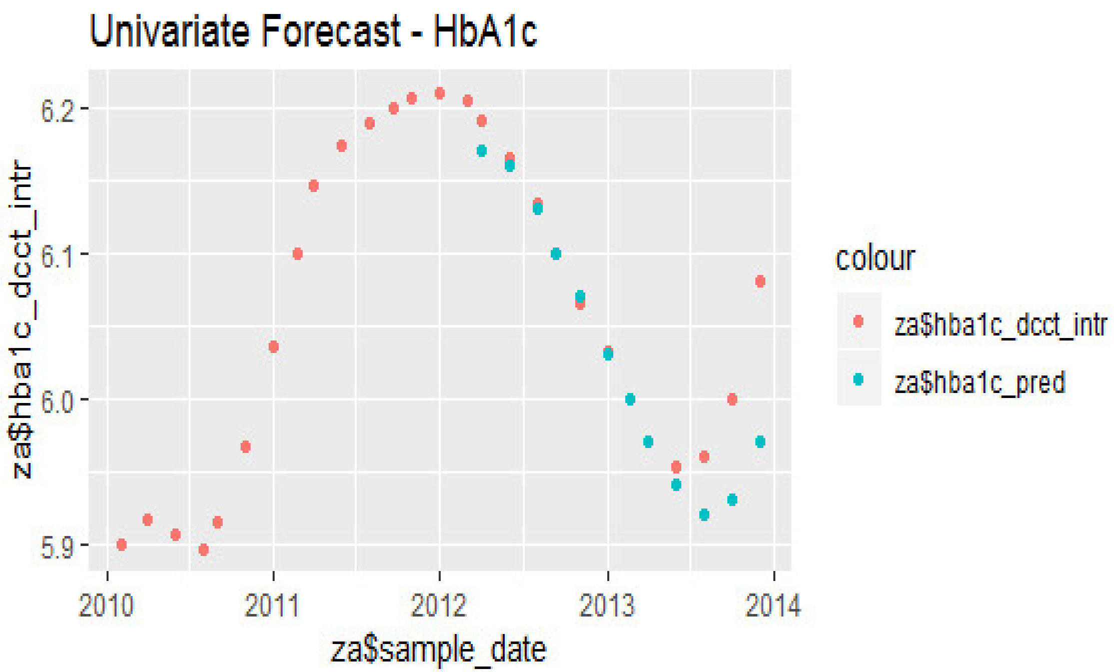
Data

- Experiment 2
  - We select 5 year period for each patient with 30 measurements placed 2 months apart.
  - Number of patients: 1200
  - The years used for training are decreased as in 3 and half years, 3 years, 2 and half etc
  - Feature selected: HbA1c

- Data is randomly divided into training 90% and testing 10%

Results





MAE: 0.182, RMSE:0.207
With two and half years of training

MAE:0.022, RMAE:0.04
With two years of training

Results

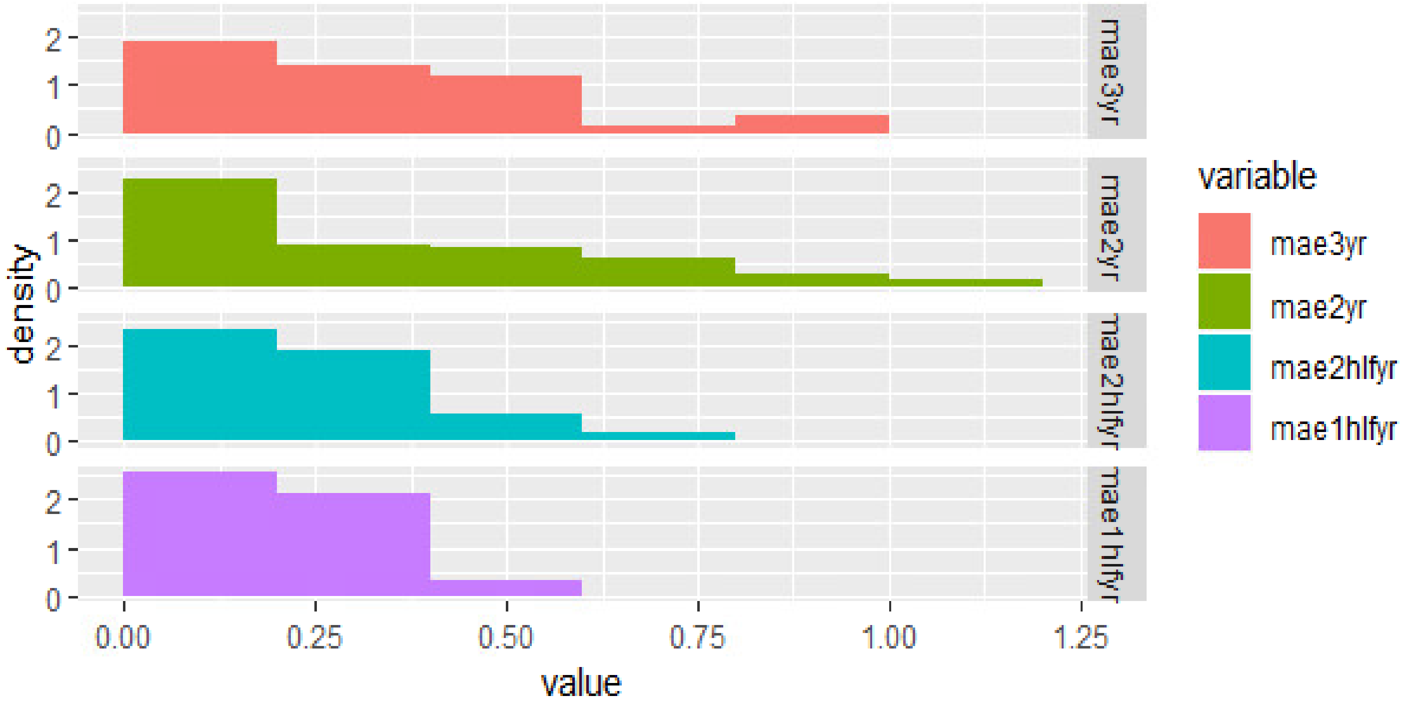


Fig: Histogram of MAE's of all patients over decreasing years of training, from the top 3 years, 2 years, 2 and half years and one and half years**.** The error is computed 12 months of prediction time. (These are the histograms of MAE of all the patients in the test set)

Results

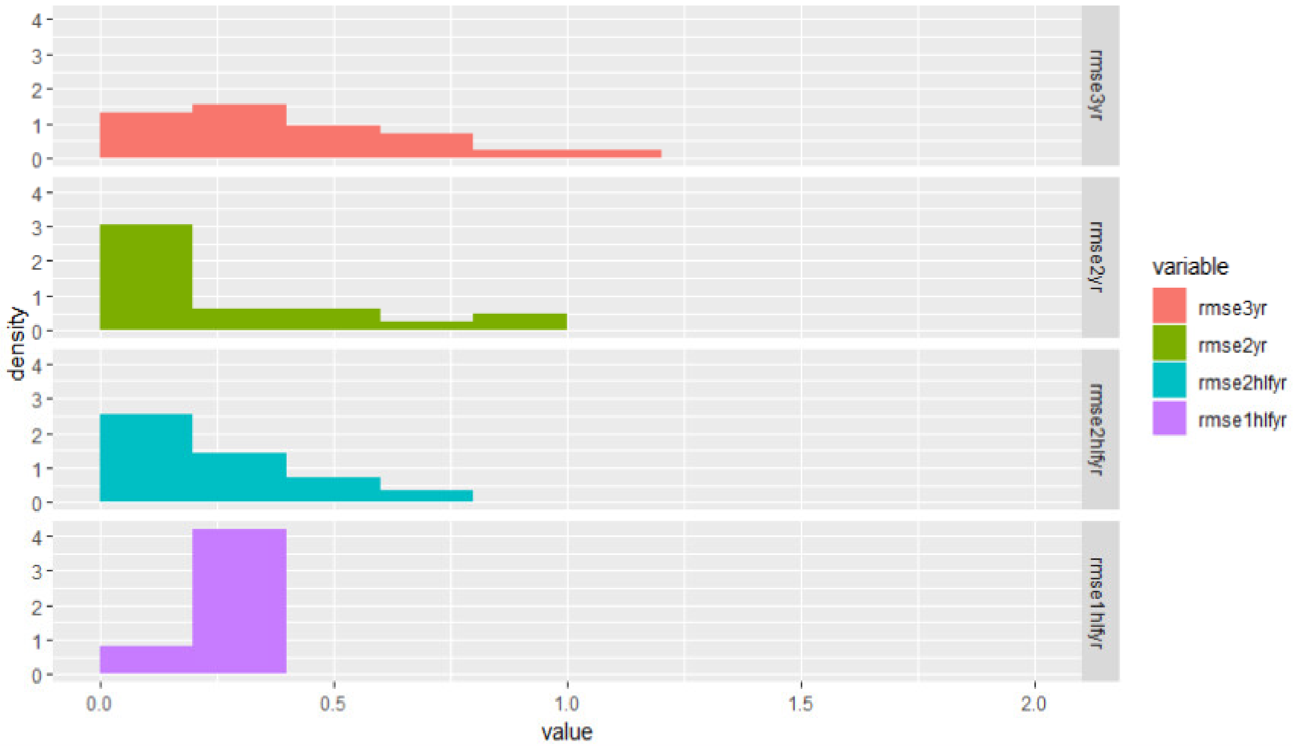


Fig: Histogram of RMSE's of all patients over decreasing years of training, from the top 3 years, 2 years, 2 and half years and one and half years**.** The error is computed 12 months of prediction time. (These are the histograms of RMSE of all the patients in the test set)

Results

No of years of training	RMSE(Std Dev)	MAE(Std dev)
One and half years	0.63 (1.22)	0.58(1.2)
Two years	0.580(0.875)	0.560(0.847)
Two and half years	0.277(0.272)	0.250(0.252)
Three years	0.254(0.26)	0.26(0.219)

Table 1: Global mean and standard deviation(mean and standard deviation over all individual patient's)

Experiment 2
Multivariate Forecasting

Data

- Experiment 3
  - Synthetic data generated based on real data
  - We select 5 year period for each patient with 30 measurements placed 2 months apart – raw data
  - Feature selected: HbA1c, hdl, BMI, non hdl

- Data is randomly divided into training 70%, testing 15% and 15% validation

Related work

- Liu, Zitao, and Milos Hauskrecht. “A Personalized Predictive Framework for Multivariate Clinical Time Series via Adaptive Model Selection.” Proceedings of the ... ACM International Conference on Information & Knowledge Management. ACM International Conference on Information and Knowledge Management vol. 2017 (2017): 1169-1177. doi:10.1145/3132847.3132859
  - EHRs of post-surgical cardiac patients
  - 500 patients

Current and Future work

- Article – “Predicting parameter progression in a dieting diabetic population using Recurrent Neural Networks” – In progress
- Multivariate multiple step forecasting. – In progress
- Data preparation of multivariate forecasting – In progress
- Multivariate input single variable output forecasting

Acknowledgements

- Participants of GoDarts
- Prof. E Pearson & Dr. Louise, Ninewells hospital, Dundee
- Anand, Sushrima, Jacob, Syed
- HIC team – Dr. Joseph Ward

Funding & disclaimer

"The research was commissioned by the national institute for health research using official development assistance (ODA) funding [INSPIRED 16/136/102]."

"The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the department of health and social care. "

THANK YOU
FOR
YOUR
ATTENTION
ANY QUESTIONS?