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

Presented By : Jyothsna Divyananda

Supervisors

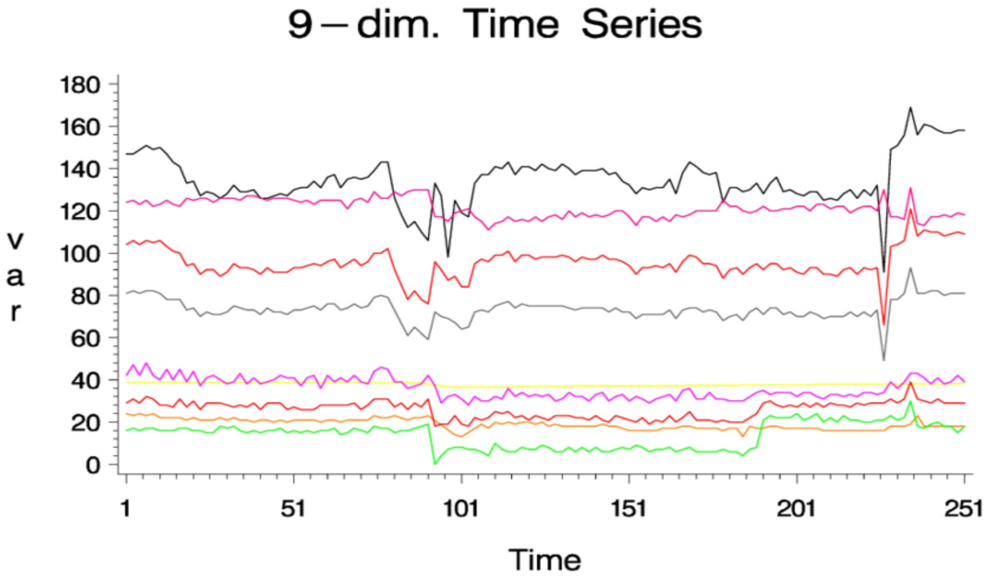
Prof Emanuele Trucco, Dr Alex Doney, University of Dundee

Dr Prathiba, MDRF, Chennai, India

Multivariate Time Series(MTS)

- Multivariate time series has more than one-time dependent variable.
- The variables depend not only on their past values but also have some dependency on other variables.

Fig 1: Example of multivatiate time series (MTS), (a) example plot of MTS, (b) example table for MTS



a

Date	Humidity	Temperature	Wind_dir	Wind_speed	Sample_Measurement
2013-09-12 00:00:00	74.0	25.044	354.0	0.0	11.8
2013-09-12 01:00:00	74.0	25.044	354.0	0.0	13.2
2013-09-12 02:00:00	75.0	24.490	303.0	3.0	8.2
2013-09-12 03:00:00	88.0	22.000	303.0	3.0	13.8
2013-09-12 04:00:00	78.0	23.000	272.0	4.0	14.7

b

https://www.researchgate.net/publication/28351883_Statistical_Methods_in_Intensive_Care_Online_Monitoring/figures?lo=1&utm_source=google&utm_medium=organic
https://datascience.stackexchange.com/questions/37212/multivariate-time-series-forecasting-using-lstm

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

Experiment 1
Multivariate Forecasting
(Synthetic Data)

Data

- Experiment 1
  - Dataset 1 -SML2010 Data Set
  - 40 days' worth data
  - The data has been sampled every minute
  - Target variable: Temperature
  - Features: 18

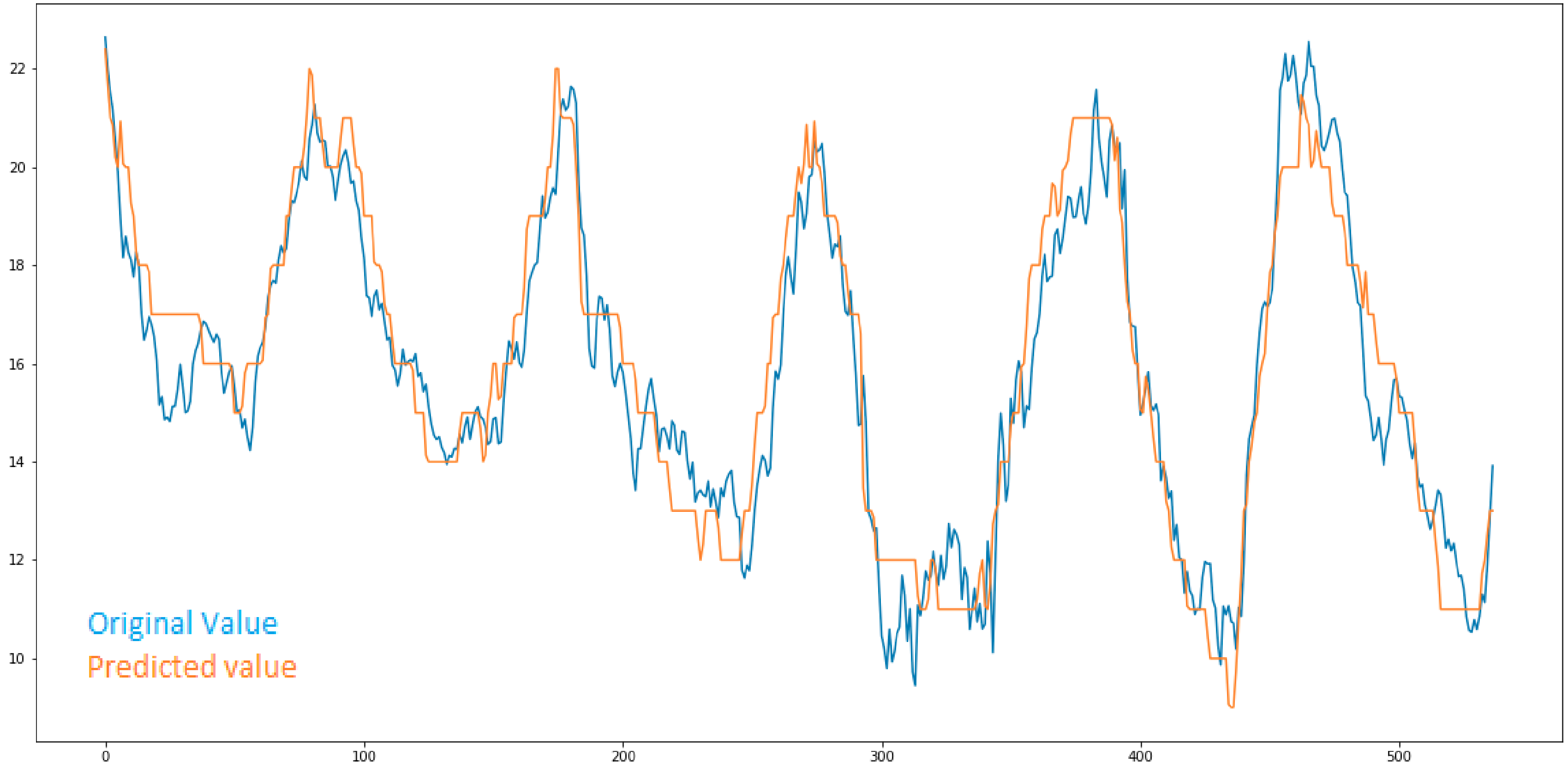
  - Dataset 2
  - Synthetic data generated based on real data
  - Target variable: Hdl
  - Features selected: HbA1c, Hdl, Chol

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

GRU Training (SML2010 data)

- Deep learning framework: Keras (TF backend)
- Number of training epochs: 400
- Optimizer: Adam (learning rate=0.001)
- Loss: Mean Squared Error
- Metrics: Mean absolute error, Root mean square error
- Training strategies:
  - Model with best validation performance is saved

Results

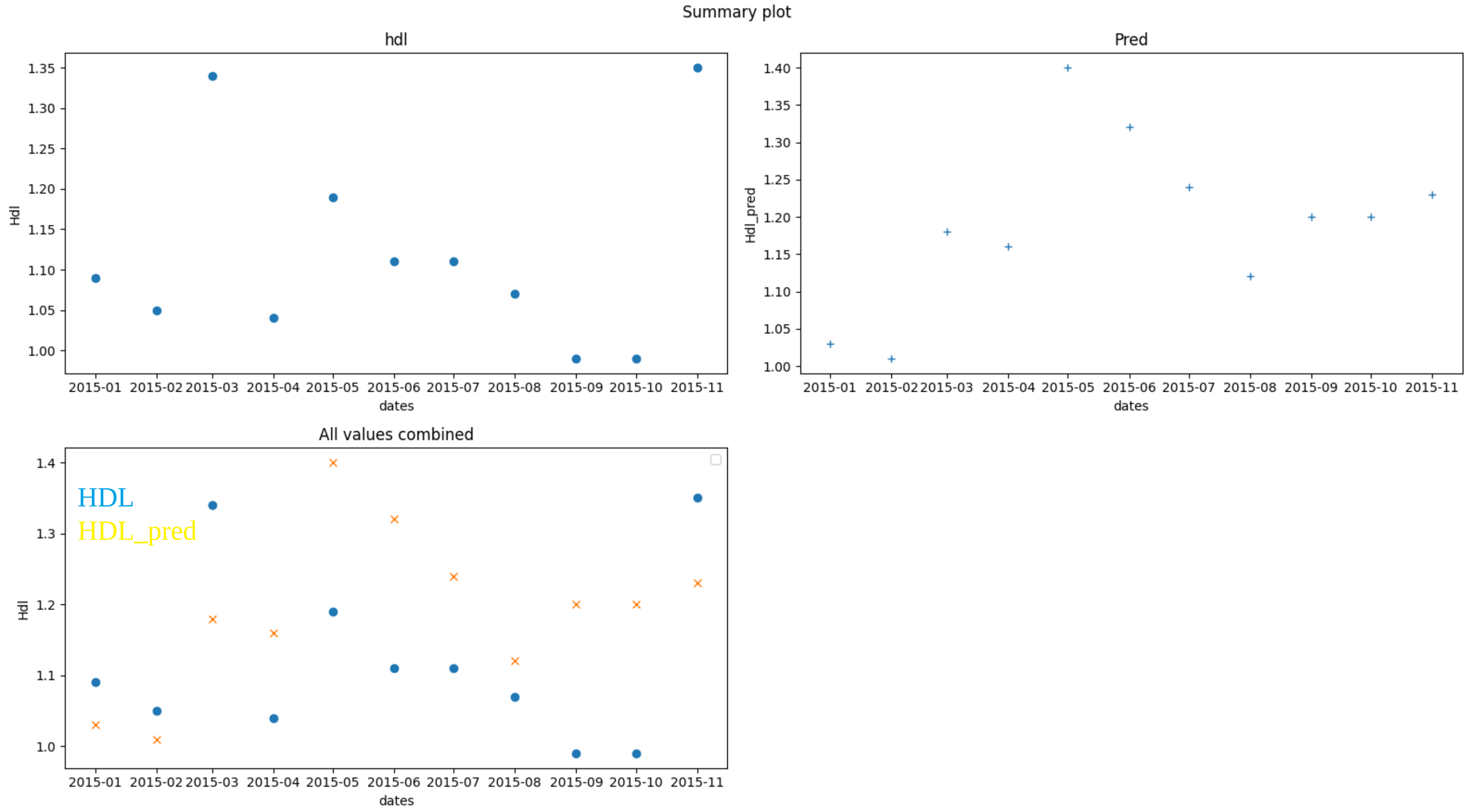


MSE: 1.12
MAE: 0.85

GRU Training (Synthetic data)

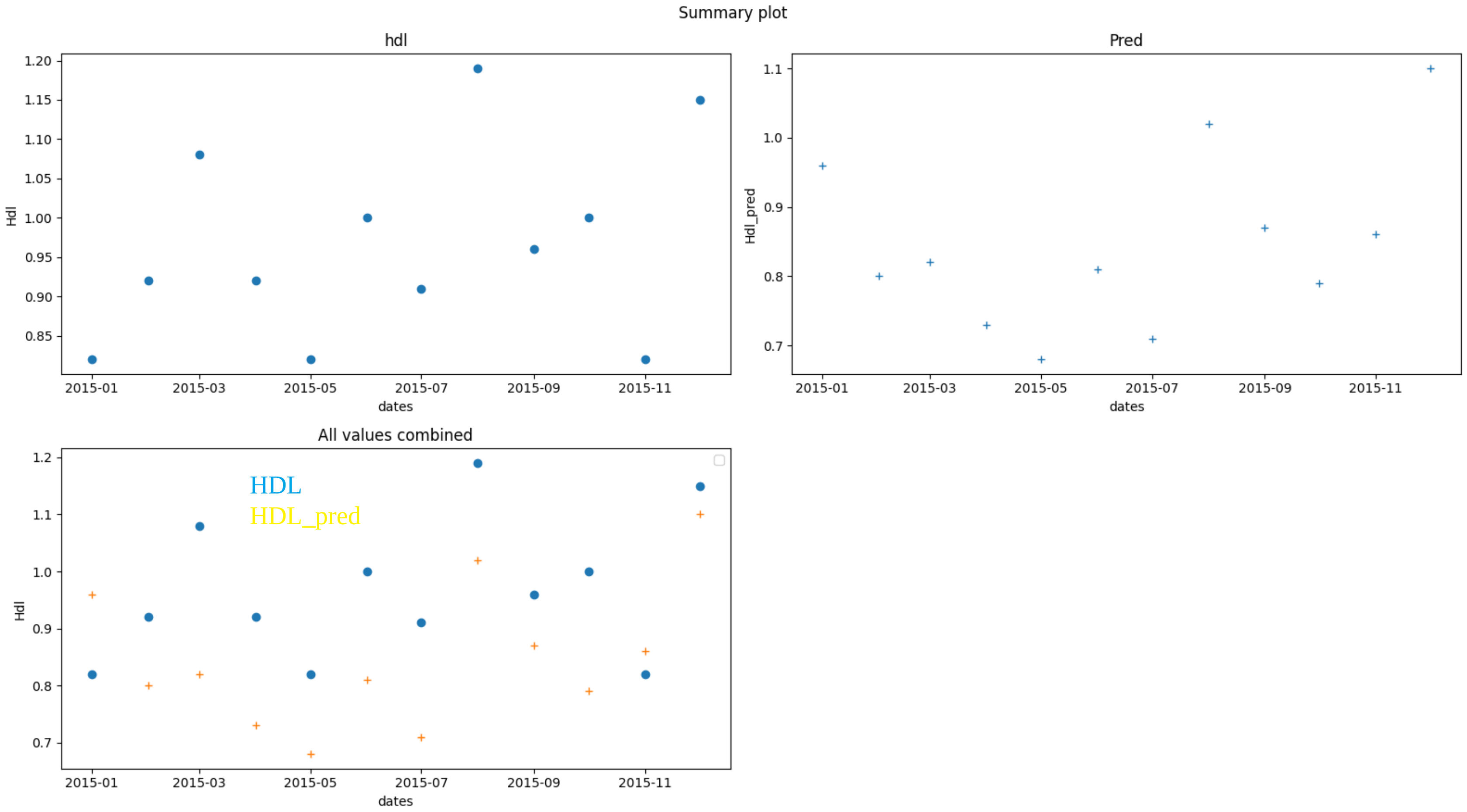
- Deep learning framework: Keras (TF backend)
- Number of training epochs: 150
- Optimizer: Adam (learning rate=0.0001)
- Loss: Mean Squared Error
- Metrics: Mean absolute error, Root mean square error
- Training strategies:
  - Model with best validation performance is saved

Results



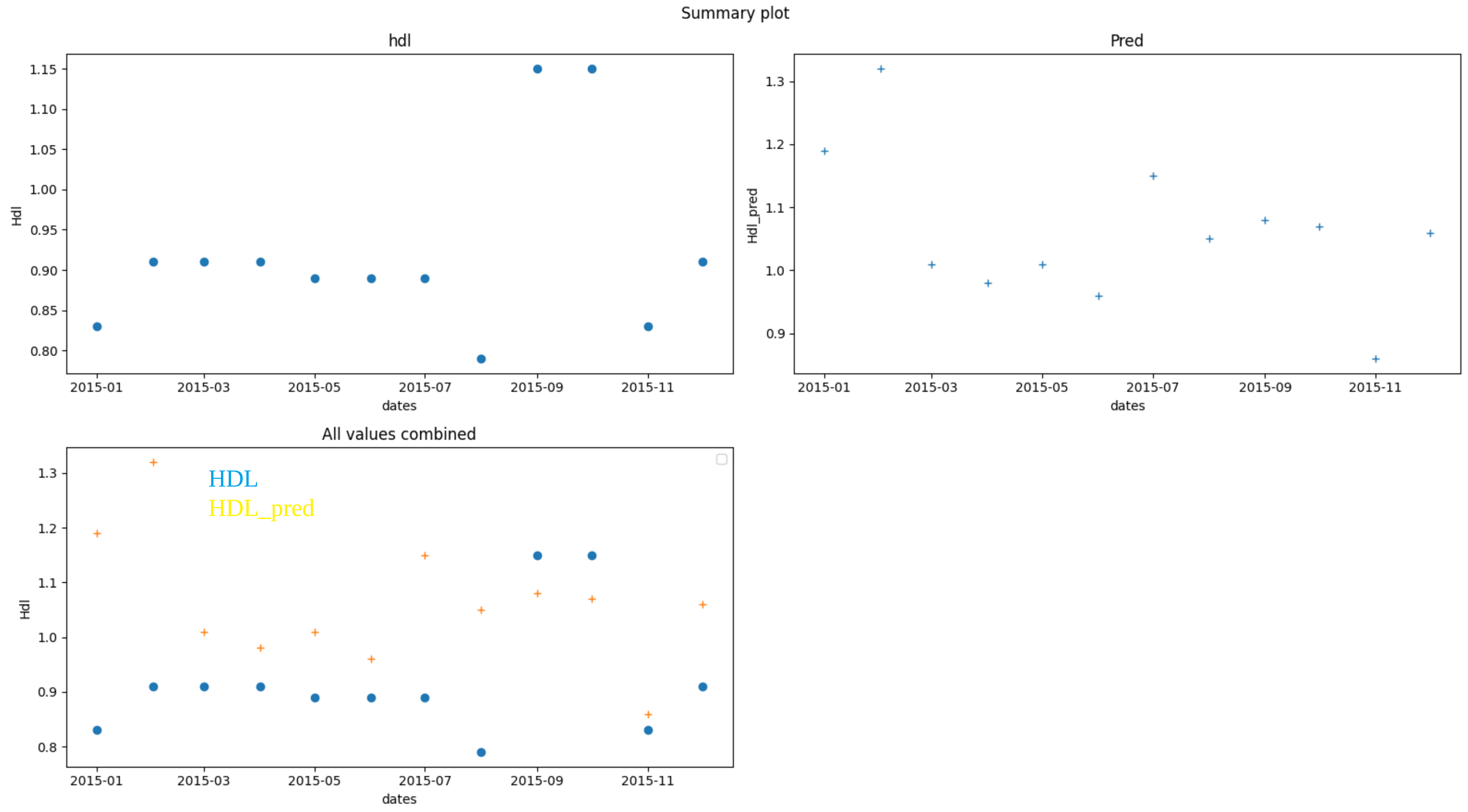
RMSE:0.071
MAE: 0.051

Results



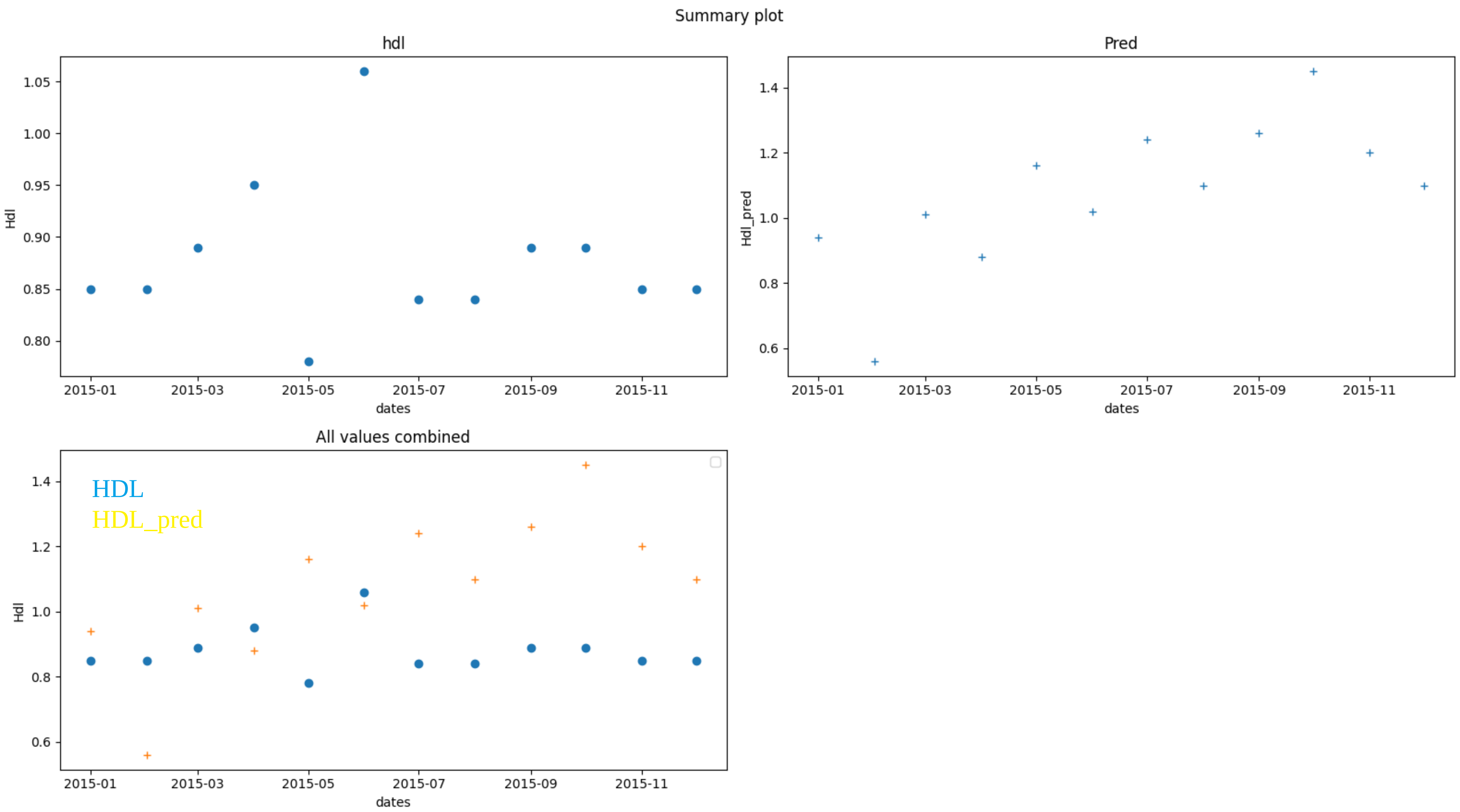
RMSE:0.34
MAE: 0.351

Results



RMSE:0.228
MAE: 0.247

Results



RMSE:0.171
MAE: 0.178

Related work

- 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.
- Karsanti, H. T., Ardiyanto, I., & Nugroho, L. E. (2019). Deep Learning-Based Patient Visits Forecasting Using Long Short Term Memory. Proceeding - 2019 International Conference of Artificial Intelligence and Information Technology, ICAIIT 2019, 344–349. https://doi.org/10.1109/ICAIIT.2019.8834634
- Maggiolo, M., & Spanakis, G. (2019). Autoregressive convolutional recurrent neural network for univariate and multivariate time series prediction. arXiv preprint arXiv:1903.02540.

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

Experiment 2
Multivariate Forecasting
(GoDARTS data)

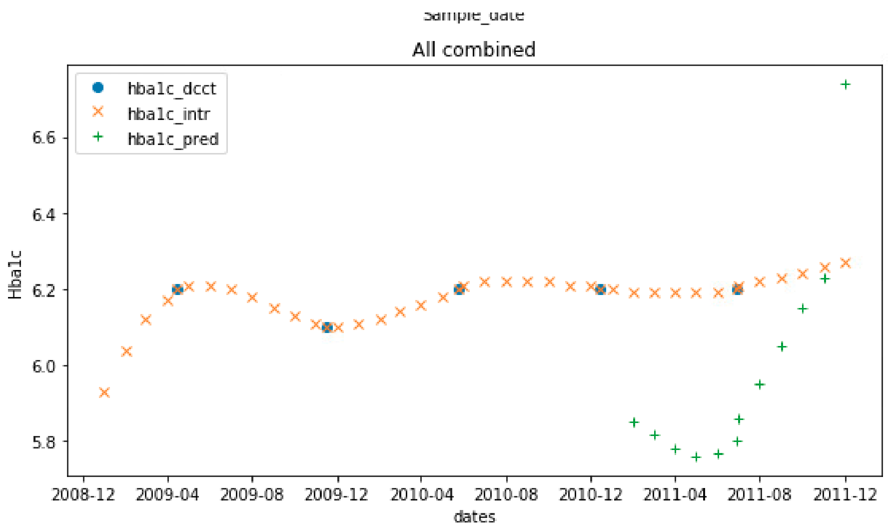
Data

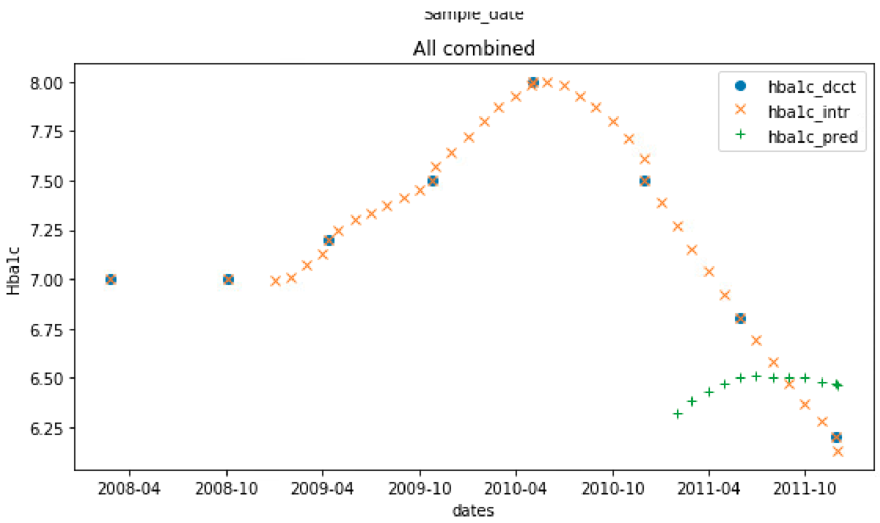
- Experiment 2
  - We select 3-year period for each patient with 36 measurements placed 1 month apart.
  - Number of patients: 1200
  - Target variable: HbA1c
  - Other variables: Hdl, non-Hdl, BMI

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

Li, K., Daniels, J., Liu, C., Herrero, P., & Georgiou, P. (2020). Convolutional Recurrent Neural Networks for Glucose Prediction. IEEE Journal of Biomedical and Health Informatics, 24(2), 603–613. https://doi.org/10.1109/JBHI.2019.2908488

Results

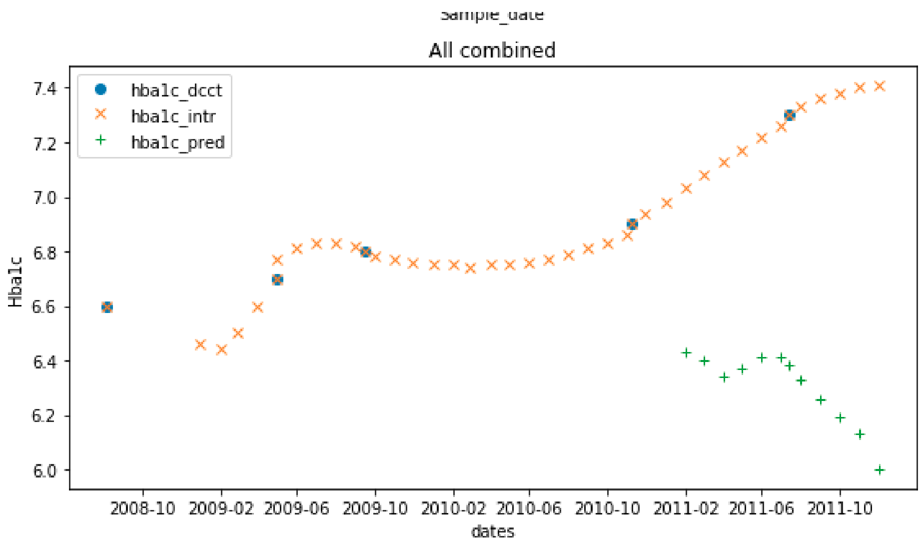


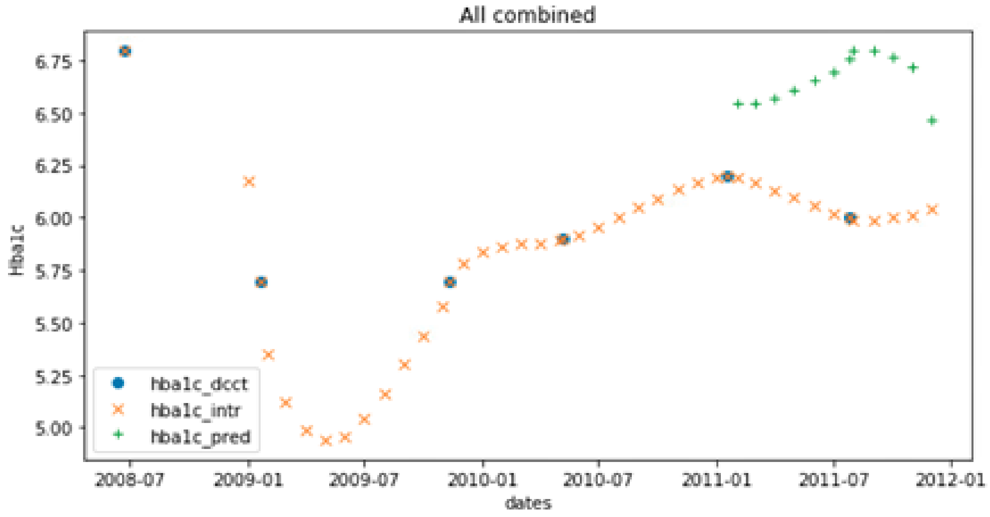


MAE: 0.176, RMSE:0.263

MAE: 0.25, RMSE:0.256

Results

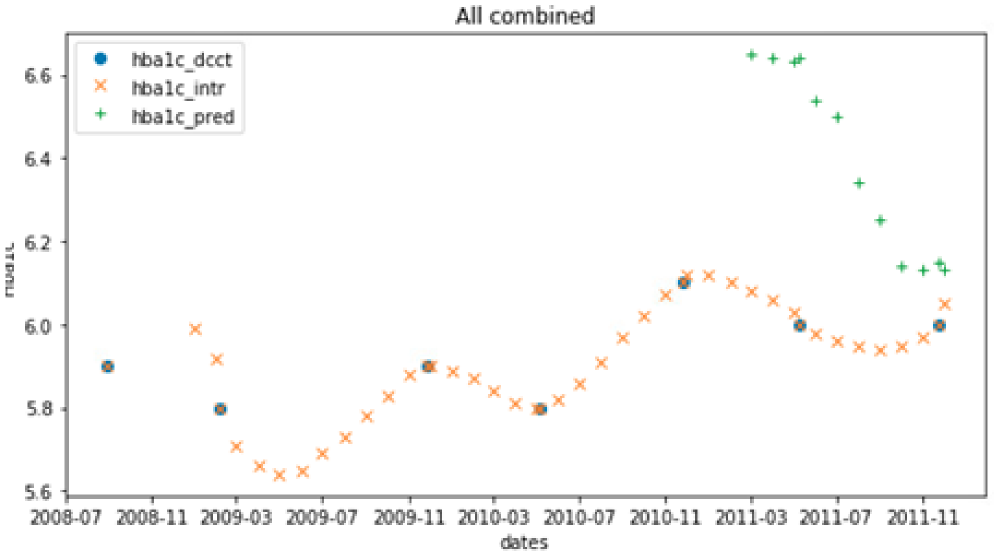


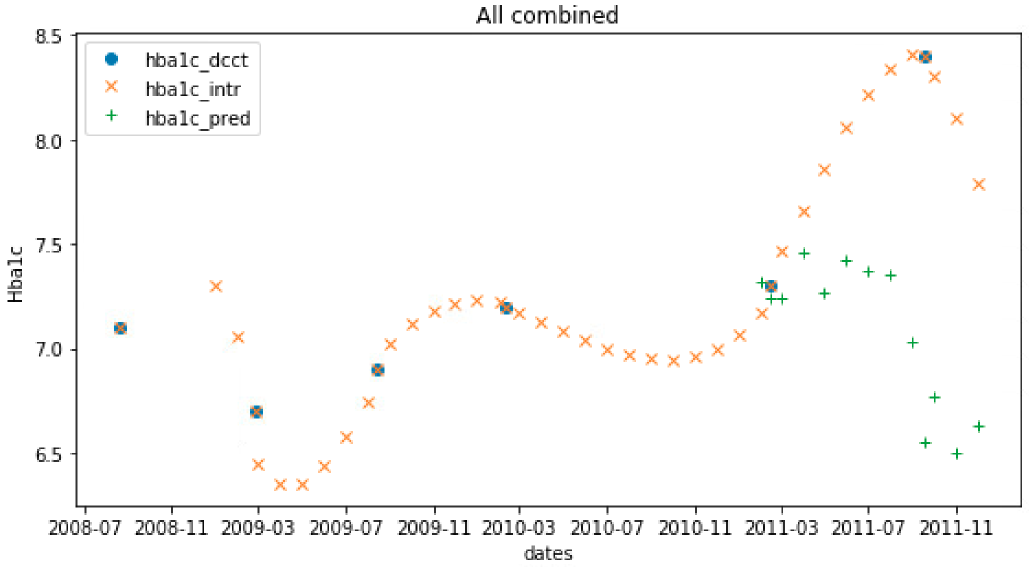


MAE: 0.608, RMSE:0.653

MAE: 0.564, RMSE:0.577

Results

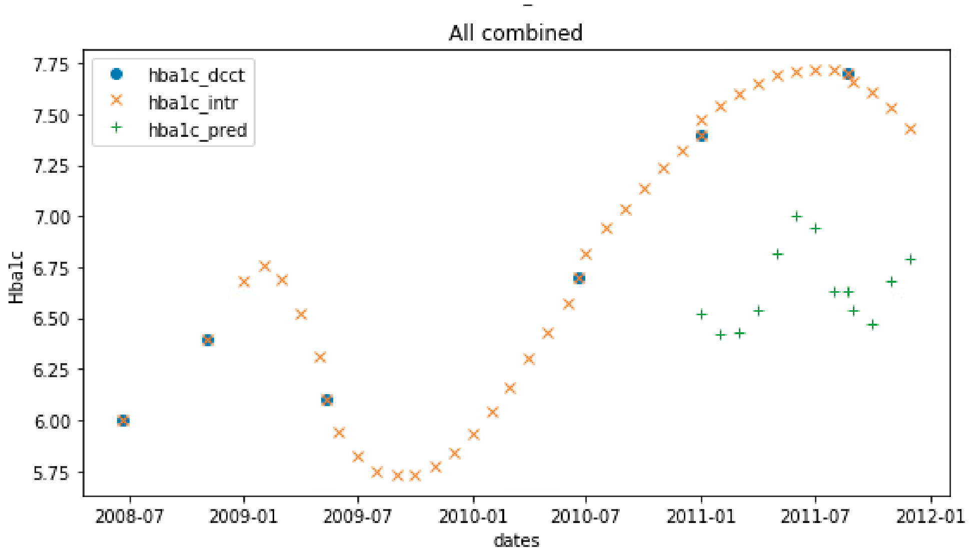


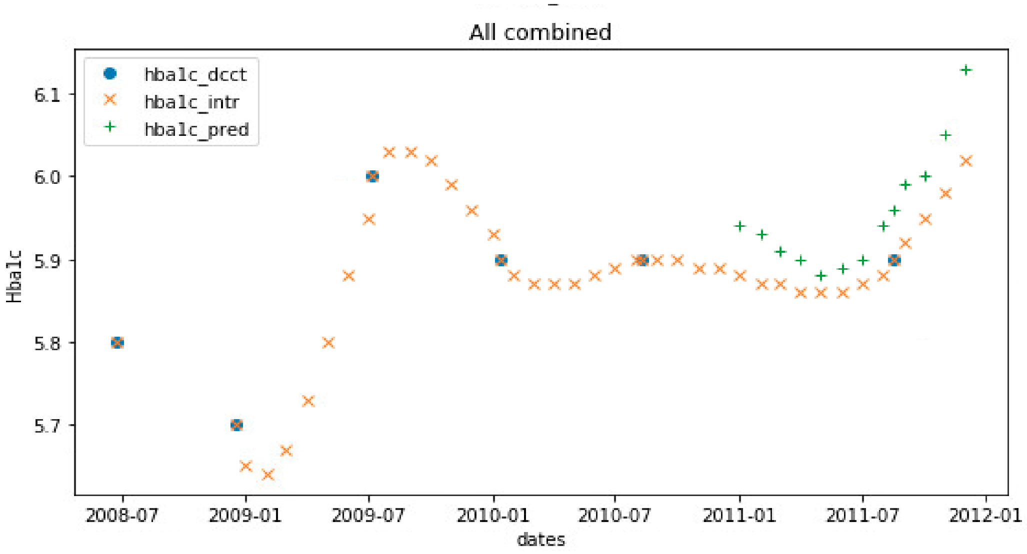


MAE: 0.331, RMSE:0.361

MAE: 0.864, RMSE:1.08

Results





MAE: 0.171, RMSE:0.178

MAE: 0.165, RMSE:0.199

Related work

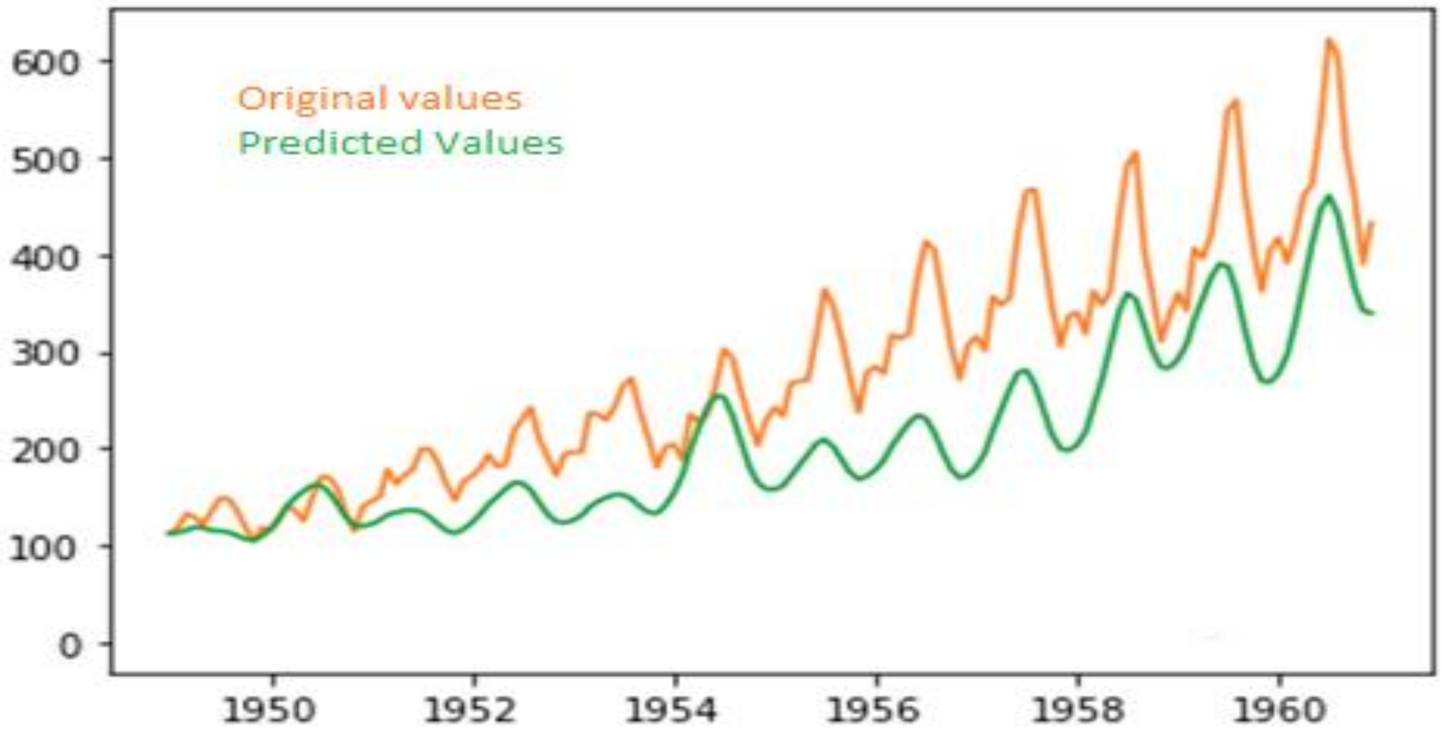
- Pham, T., Tran, T., Phung, D., & Venkatesh, S. (2016). DeepCare: A deep dynamic memory model for predictive medicine. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9652 LNAI*(i), 30–41. https://doi.org/10.1007/978-3-319-31750-2_3
- Pham, T., Tran, T., Phung, D., & Venkatesh, S. (2017). Predicting healthcare trajectories from medical records: A deep learning approach. Journal of Biomedical Informatics, 69, 218–229. https://doi.org/10.1016/j.jbi.2017.04.001
- Balasooriya, K., & Nanayakkara, N. D. (2020). Predicting short-term changing blood glucose level of diabetes patients using noninvasive data. IEEE Region 10 Annual International Conference, Proceedings/TENCON, 2020-Novem, 31–36. https://doi.org/10.1109/TENCON50793.2020.9293823
- Albers, D. J., Levine, M., Gluckman, B., Ginsberg, H., Hripcsak, G., & Mamykina, L. (2017). Personalized glucose forecasting for type 2 diabetes using data assimilation. In PLoS Computational Biology (Vol. 13, Issue 4). https://doi.org/10.1371/journal.pcbi.1005232

Experiment 3
ARIMA
(Univariate Forecasting)

Data

- Experiment 3
  - Air passenger's dataset
  - Test to check data stationarity
  - Synthetic dataset – In progress (Individual patient forecasting)

Results



AirPassengers dataset
ADF Statistic: -2.717
p-value: 0.071

Related work

- Alzahrani, S. I., Aljamaan, I. A., & Al-Fakih, E. A. (2020). Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. *Journal of Infection and Public Health*, *13*(7), 914–919. https://doi.org/10.1016/j.jiph.2020.06.001
    - COVID -19 progression
    - 500 patients
- Siami-Namini, S., Tavakoli, N., & Siami Namin, A. (2019). A Comparison of ARIMA and LSTM in Forecasting Time Series. Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018, 1394–1401. https://doi.org/10.1109/ICMLA.2018.00227

Current and Future work

- Article – “Predicting parameter progression in a dieting diabetic population using Recurrent Neural Networks” – In progress
- ARIMA univariate forecasting implementation on GoDARTS
- Implementation of Many-Many forecasting on GoDARTS

Acknowledgements

- Participants of GoDarts
- Prof. E Pearson
- Dr. Louise Donnelly
- INSPIRED Group
- 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?