Personalised Glucose Prediction via Deep Multitask Networks

John Daniels and **Pau Herrero** and **Pantelis Georgiou**[1]

Abstract. Glucose control is an essential requirement in primary therapy for diabetes management. Digital approaches to maintaining tight glycaemic control, such as clinical decision support systems and artificial pancreas systems rely on continuous glucose monitoring devices and self-reported data, which is usually improved through glucose forecasting. In this work, we develop a multitask approach using convolutional recurrent neural networks (MTCRNN) to provide short-term forecasts using the OhioT1DM dataset which comprises 12 participants. We obtain the following results - 30 min: 19.79±0.06 mg/dL (RMSE); 13.62±0.05 mg/dL (MAE) and 60 min: 33.73±0.24 mg/dL (RMSE); 24.54±0.15 mg/dL (MAE). Multitask learning facilitates an approach that allows for learning with the data from all available subjects, thereby overcoming the common challenge of insufficient individual datasets while learning appropriate individual models for each participant.

1 INTRODUCTION

In recent years, the proliferation of biosensors and wearable devices has facilitated the ability to perform continuous monitoring of physiological signals. In diabetes management, this has come with the increasing use of continuous glucose monitoring (CGM) devices for helping with glucose control. The current literature on clinical impact of CGM devices shows that continuously monitoring blood glucose concentration levels has benefit in maintaining tight glycaemic control [5, 2]. As a next step, glucose prediction offers an opportunity to further improve glucose control by taking actions to avert adverse glycaemic events, such as suspension of insulin delivery in closed-loop systems to avert hypoglycaemia.

The general work in this area has typically involved collecting data covering physiological variables such as glucose concentration levels, heart rate, and self-reported data covering exercise, sleep, stress, illness, insulin, and meals. However, public datasets covering ambulatory monitoring of T1DM population are not widely available.

Deep learning [6] facilitates learning the optimal features and has been shown to perform better than other methods involving hand crafted features that have been employed in recent times for predicting glucose concentration levels. However, typically these models require relatively large amounts of data to converge on an appropriate model.

In this work, we employ a multitask learning [1] approach in order to improve the performance of the glucose forecasting in a neural network, where each individual is viewed as a task, using shared layers to enable learning form other individuals.

[1] Imperial College London, United Kingdom, email: jsd111@imperial.ac.uk, p.herrero-vinias@imperial.ac.uk, pantelis@imperial.ac.uk

2 RELATED WORK

Glucose prediction has been a long-standing area of focus in the diabetes community. As a result, many approaches have existed in order to provide near-time glucose concentration level forecasts.

Early work in this area have focused on physiological models and traditional machine learning methods in predicting glucose concentration levels [12, 3]. Recent work as seen in the 2018 Blood Glucose Predictive Challenge has seen a move towards deep learning methods with more impressive results [11, 9, 14, 8]. These have used convolutional architectures, recurrent architectures, or a combination of both to model the task of glucose prediction.

3 DATASET AND DATA PREPROCESSING

In this section, we detail the transformations that are performed on the data prior to training and testing the model for each T1DM participant.

3.1 OhioT1DM Dataset 2020

The OhioT1DM dataset 2020 [10] is a dataset comprising 12 unique participants that cover eight weeks of daily living. The participants are given IDs as the data is anonymised. This data comprises physiological data gathered using a continuous glucose monitor (blood glucose concentration levels) and wristband device (heart rate, skin conductance, skin temperature), activity data (acceleration, step count), and self-reported data (meal intake, insulin, exercise, work, sleep, and stressors).

3.2 Dealing with Missing Values

A non-trivial aspect of the datasets used for developing glucose prediction models is the aspect of missingness. This is evident in the Ohio T1DM dataset with missingness present in both physiological variables and self-reported data [4].

Linear Interpolation: The blood glucose values that are missing in this dataset are typically missing at random. This could be attributed to issues around replacing glucose sensors and/or transmitters, or dealing with faulty communication. As a result, we employ linear interpolation in the training set to handle imputation of missing blood glucose concentration levels in the dataset over a period of one hour. In the samples where more than an hour of CGM data is missing the sample is discarded from the training set. This is illustrated with an example sequence in (C) of Fig.1

On the other hand, features which comprise self-reported data the assumption is made that any missing values represent an absence of
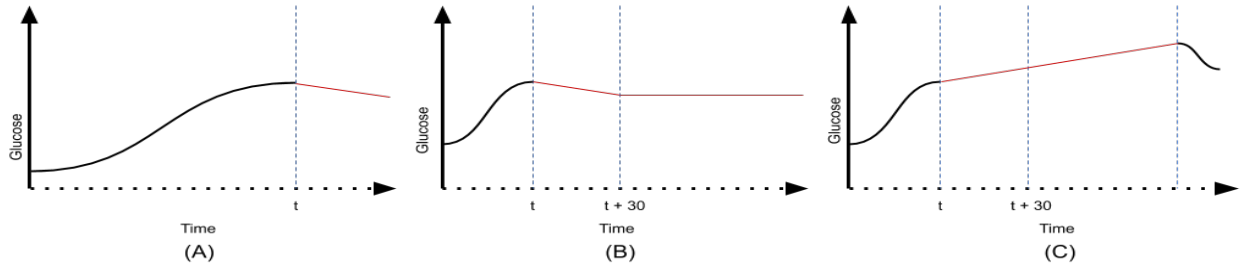
Figure 1. A visualisation of the imputation methods employed in this work. In (A) the input sequence has at least 30 minutes of recent values missing (eg. linear extrapolation). (B) shows the imputation scheme during testing for longer than 30 minutes of recent values missing (zero-order hold). Finally (C) shows the imputation scheme when the missing values of the input sequence are located between real values (linear interpolation).

said feature. Therefore all missing values in insulin, meal intake and reported exercise are imputed with zero.

The missingness in features from the self-reported data in the testing set is tackled similarly as in the training set. However, this is not the case for blood glucose concentration levels as interpolation when a current value at a given timestep is missing would lead to an inaccurate evaluation of model performance.

Extrapolation: In order to accurately evaluate the performance of the model we cannot always rely on interpolation at test time as this may require, in a real-time setting, an unknown future value to perform interpolation. Consequently, we need to rely on other methods of extrapolation to impute the missing glucose concentration levels. In this scenario (A), for gaps of data less than 30 minutes, we impute missing values with predicted values from the trained model. For missing recent values longer than 30 minutes as in (B), we pad the remaining values with the last computed value. In cases where, a gap larger than 30 minutes is evident in historical data and a current value is present at the given timestep, linear interpolation was then employed instead to provide a more accurate imputation.

3.3 Standardisation

To enable training the proposed model effectively, we perform transformation of the relevant input features (blood glucose concentration, insulin bolus, meal(carbohydrate) intake, and reported exercise). The blood glucose concentration levels are scaled down by a factor of 120. Similarly, the insulin bolus is scaled by 100 and meal intake values are scaled by 200 in the same range between features. The exercise values are transformed to a simple binary representation of the presence or absence of exercise, from the recorded exercise intensity on a range from 1-10.

4 METHODS

In this section we detail the machine learning technique that is used to provide the means of learning personalised models with the entire dataset. We detail the approach to develop the deep multitask network for personalisation. We provide a summary of the hyperparameters used in training as well and setting up the input for personalised multitask learning.

4.1 Multitask Learning

Multitask learning is an approach in machine learning that can be broadly described as a method of learning multiple tasks simultane-

ously with the aim of improving generalisation [1].

Multitask learning for personalisation has been used mainly in affective computing [13] with early work in diabetes management focusing on using multitask learning for developing prediction models for clustered groups of Type 1, Type 2, an non-diabetic participants [7] rather than leveraging similarities within groups such as gender, for personalised glucose predictions.

As seen Figure 2, the output from the shared layers are now fed into the individual(task)-specific fully connected layers of each user.

In a multitask setting of this kind, a multiplicative gating approach is used to ensure that the input corresponding to the particular user trains on just that user in the individual-specific layers. In that sense, at each iteration a batch that consists of data from a particular individual is used to train the shared layers and the layers specific to the individual.

4.2 CRNN Model

The deep learning model trained in the multitask learning setting is a convolutional recurrent neural network (CRNN) proposed by Li et. al [8] to perform short-term glucose prediction. This forms the basis of the single-task (STL) model. The convolutional recurrent model consists initially of a 3 temporal convolutional layers that perform a 1-D convolution with a Gaussian kernel over the sequence of input to extract features of various rates of appearance, followed by a max pooling layer after each convolution operation. The input is a 4-dimensional sequence that takes a 2-hour window of historical data.

The output from the convolutional layers performs feature extraction and feeds into a recurrent long short-term memory (LSTM) layer that is able to better model the temporal nature of the task.

The output from the shared layers feed into the fully connected layers of each user and to then provide the change in glucose value over the prediction horizon. This is then added to the current glucose value to provide the forecast glucose concentration level.

4.3 Loss Function

The loss function used for converging to the appropriate model for the glucose forecasting is the mean absolute error. This is expressed below as:

$$L(y, \hat{y}) = \frac{1}{N_{batch}} \sum_{k=1}^{N} |y - \hat{y}| \tag{1}$$
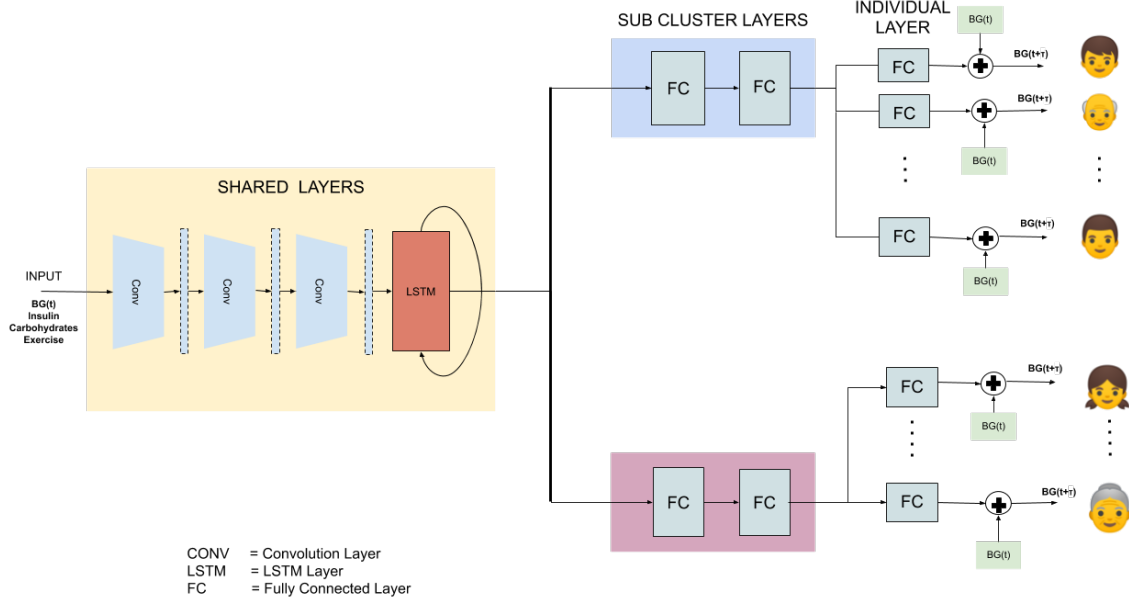
Figure 2. A detailed look at the formulation of convolutional recurrent networks in a multitask setting. In this setting, each user is represented as a task. In addition, the initial layers (convolutional and recurrent layers) are shared between each user, the next two (dense) layers are shared based on gender, and the last (dense) layer is specific to each user.

where $\hat{y}$ denotes the predicted results given the historical data and y denotes the reference change in glucose concentration over the relevant glucose prediction, and N_{batch} refers to the batch size.

4.4 Hyperparameters

The following table details provides the details of the hyperparameters used for the model architecture at each layer.

Table 1. A table detailing the size and dimensions of layers in the multitask CRNN model (MTCRNN)

Layer Description (layer)	Output Dimensions	No. of Parameters
Shared Convolutional Layers (Batch×Steps×Channels)		
(1) 1×4 conv	$128(1) \times 24 \times 8$	104
max_pooling, size 2	$128(1) \times 12 \times 8$	–
(2) 1×4 conv	$128(1) \times 12 \times 16$	528
max_pooling, size 2	$128(1) \times 6 \times 16$	–
(3) 1×4 conv	$128(1) \times 6 \times 32$	2080
max_pooling	$128(1) \times 3 \times 32$	–
Shared Recurrent Layer (Batch×Cells)		
(4) lstm	$128(1) \times 64$	24832
Sub-cluster Dense Layers (Batch×Units)		
(5) dense	$128(1) \times 256$	16640
(6) dense	$128(1) \times 32$	8224
Individual-Specific Dense Layers (Batch×Units)		
(7) dense	$128(1) \times 1$	33

The optimiser used for this work is Adam. The learning rate is 0.0053. The model is trained for 200 epochs. This value was obtained through grid search optimisation.

The model is developed on Keras 2.2.2, with a Tensorflow 1.5 backend. The training is performed on an NVIDIA GTX 1050 GPU.

The repository for the code accompanying the paper can be found at: https://github.com/jsmdaniels/ecai-bglp-challenge

5 RESULTS

5.1 Evaluation Metrics

The model is tested on data from six participant IDs: 540, 544, 552, 567, 584, 596.

The evaluation of the model is based on two metrics: root mean square error (RMSE) and the mean absolute error (MAE). The extrapolated points are not considered in calculating these metrics. The formulation of these metrics are provided below:

$$RMSE = \sqrt{\frac{1}{N} \sum_{k=1}^{N} (y - \hat{y})^2}, \tag{2}$$

$$MAE = \frac{1}{N} \sum_{k=1}^{N} |y - \hat{y}|. \tag{3}$$

where $\hat{y}$ denotes the predicted results given the historical data and y denotes the reference glucose measurement, and N refers to the data size.

In order to undertake a comprehensive evaluation of the model performance, the subsequent criteria for assessment are followed:

- **Performance evaluation over 30-minute and 60-minute prediction horizon (PH):** The RMSE and MAE for each participant is analysed for a the same length of values for both prediction horizons.
- **Comparison of training setting:** The performance of the multitask learning (MTL) approach is evaluated in the context of comparison with the performance of a single task learning (STL) approach which uses only patient specific data.

- **Multiple runs for each participant ID:** The multitask CRNN (MTCRNN) model uses randomly initialised weights at the start of training. Given the variable nature of this training procedure, the results reported are the average of 5 model runs.

The unit for results reported below is mg/dL. The best performance is in **bold**.

Table 2. A table showing prediction performance for 30 minutes the RMSE and MAE results of the six participants over 5 runs (CRNN)

	MTL		STL	
ID	RMSE	MAE	RMSE	MAE
540	**21.19±0.07**	**15.17±0.06**	22.45±0.39	16.21±0.34
544	**16.82±0.09**	**11.72±0.06**	18.63±1.59	12.57±0.23
552	**16.30±0.12**	**11.92±0.03**	17.11±0.24	12.68±0.49
567	**24.12±0.17**	**15.55±0.03**	24.73±0.45	16.01±0.71
584	**23.66±0.20**	**15.77±0.08**	24.30±0.48	16.20±0.23
596	**16.63±0.15**	**11.59±0.09**	16.78±0.20	12.00±1.77
Average	**19.79±0.06**	**13.62±0.05**	20.67±0.32	14.28±0.19

Table 3. A table showing the prediction performance for 60 minutes RMSE and MAE results of the six participants over 5 runs (CRNN)

|	MTL		STL	
ID	RMSE	MAE	RMSE	MAE
540	**38.29±0.29**	**28.60±0.17**	41.06±0.24	30.33±0.69
544	**28.97±0.24**	20.77±0.20	29.60±0.37	**20.52±0.17**
552	**29.35±0.27**	**22.07±0.13**	30.32±0.10	22.53±0.13
567	40.19±0.79	28.77±0.13	**40.09±0.64**	**27.71±0.13**
584	37.82±0.78	26.88±0.37	**37.22±0.34**	**26.64±0.41**
596	**27.74±0.11**	**20.12±0.14**	28.13±0.48	20.30±0.41
Average	**33.73±0.24**	**24.54±0.08**	34.40±0.14	24.67±0.14

6   DISCUSSION

As seen in Table 3, the results shown provide a comprehensive evaluation of the model predictive performance.
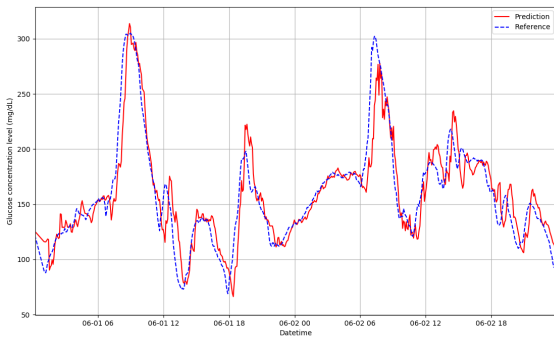
Figure 3. A graph showing the predictive performance of the model on participant ID:596 at a 30 minute predictive horizon.

Evidently, the model performance at PH = 30 minutes is better than the model performance at PH = 60 minutes, given that prediction at 60 minutes is a more complex task than prediction at 30 minutes.
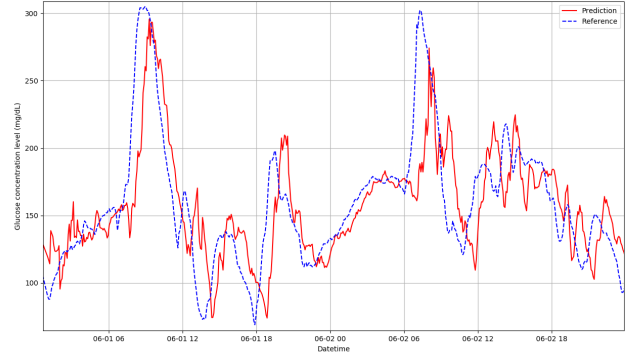
Figure 4. A graph showing the predictive performance of the model on participant ID:596 at a 60 minute predictive horizon.

Figures 3 and 4 exhibit the differences in performance as seen in the specific window for participant 596. The increased lag and reduced predictive performance can also be attributed to the higher chance of external activities (insulin, meals, exercise) that influence the blood glucose trajectory occurring over the prediction horizon.

The best predictive performances were achieved by the model with IDs 544, 552, 596 whereas, IDs 540, 567, and 584 exhibited worse performances over both 30 and 60 minute prediction horizons. An investigation of the glycaemic variability, using the coefficient of variation (CV) [2], of the training set of the former set of participants are stable (CV≤36%) whereas the latter group are labile (CV>36%). The multitask learning approach definitely performs better over the single task approach over a 30-minute prediction horizon. However, the performance improvement of the MTL approach over a 60-minute prediction is not consistent across each participant and metric.

One potential issue with multitask learning is the issue of negative transfer. This can be described as a scenario in which one or more of the tasks (individuals) or sampled batches during training are not strongly correlated, degrading the learning in the shared layers, and subsequently the performance at test time.

7   CONCLUSION

In this work, we have presented a multitask convolutional recurrent neural network that is capable of performing short-term personalised predictions - 19.79±0.06mg/dL (RMSE) and 13.62±0.05mg/dL (MAE) at 30 minutes, as well as 33.73±0.24mg/dL (RMSE) and 24.54±0.15mg/dL (MAE) at 60 minutes. We work towards leveraging population data while still learning a personalised model. In the future, we hope to address further challenges such as negative transfer during learning that could improve the accuracy of individual models. This approach would enable more accurate models to be deployed in the face of limited personal data.

REFERENCES

[1] Rich Caruana, 'Multitask Learning', *Machine Learning*, **28**(1), 41–75, (July 1997).

[2] Antonio Ceriello, Louis Monnier, and David Owens, 'Glycaemic variability in diabetes: clinical and therapeutic implications', *The Lancet Diabetes & Endocrinology*, (August 2018).

[3] E. I. Georga, V. C. Protopappas, D. Ardigò, M. Marina, I. Zavaroni, D. Polyzos, and D. I. Fotiadis, 'Multivariate Prediction of Subcutaneous Glucose Concentration in Type 1 Diabetes Patients Based on Support Vector Regression', *IEEE Journal of Biomedical and Health Informatics*, **17**(1), 71–81, (January 2013).

[4] Marzyeh Ghassemi, Tristan Naumann, Peter Schulam, Andrew L. Beam, and Rajesh Ranganath, 'Opportunities in Machine Learning for Healthcare', *arXiv:1806.00388 [cs, stat]*, (June 2018). arXiv: 1806.00388.

[5] Giacomo Cappon, Giada Acciaroli, Martina Vettoretti, Andrea Facchinetti, and Giovanni Sparacino, 'Wearable Continuous Glucose Monitoring Sensors: A Revolution in Diabetes Treatment', *Electronics*, **6**(3), 65, (September 2017).

[6] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning*, MIT Press, 2016. http://www.deeplearningbook.org.

[7] Weixi Gu, Zimu Zhou, Yuxun Zhou, Miao He, Han Zou, and Lin Zhang, 'Predicting Blood Glucose Dynamics with Multi-time-series Deep Learning', in *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems - SenSys '17*, pp. 1–2, Delft, Netherlands, (2017). ACM Press.

[8] K. Li, J. Daniels, C. Liu, P. Herrero-Vinas, and P. Georgiou, 'Convolutional Recurrent Neural Networks for Glucose Prediction', *IEEE Journal of Biomedical and Health Informatics*, 1–1, (2019).

[9] Kezhi Li, Chengyuan Liu, Taiyu Zhu, Pau Herrero, and Pantelis Georgiou, 'GluNet: A Deep Learning Framework for Accurate Glucose Forecasting', *IEEE Journal of Biomedical and Health Informatics*, **24**(2), 414–423, (February 2020).

[10] Cindy Marling and Razvan Bunescu, 'The OhioT1DM Dataset for Blood Glucose Level Prediction', *In: The 5th International Workshop on Knowledge discovery in healthcare data.*, (2020). CEUR proceeding in press. Available at http://smarthealth.cs.ohio.edu/bglp/OhioT1DM-dataset-paper.pdf.

[11] John Martinsson, Alexander Schliep, Björn Eliasson, and Olof Mogren, 'Blood Glucose Prediction with Variance Estimation Using Recurrent Neural Networks', *Journal of Healthcare Informatics Research*, **4**(1), 1–18, (March 2020).

[12] C. Pérez-Gandía, A. Facchinetti, G. Sparacino, C. Cobelli, E.j. Gómez, M. Rigla, A. de Leiva, and M.e. Hernando, 'Artificial Neural Network Algorithm for Online Glucose Prediction from Continuous Glucose Monitoring', *Diabetes Technology & Therapeutics*, **12**(1), 81–88, (January 2010).

[13] Sara Ann Taylor, Natasha Jaques, Ehimwenma Nosakhare, Akane Sano, and Rosalind Picard, 'Personalized Multitask Learning for Predicting Tomorrow's Mood, Stress, and Health', *IEEE Transactions on Affective Computing*, 1–1, (2017).

[14] Taiyu Zhu, Kezhi Li, Jianwei Chen, Pau Herrero, and Pantelis Georgiou, 'Dilated Recurrent Neural Networks for Glucose Forecasting in Type 1 Diabetes', *Journal of Healthcare Informatics Research*, (April 2020).