## Intro

Type 1 diabetes mellitus (T1DM) is a life-long autoimmune disease that typically manifests itself in children between 10 and 14 years of age. That is not to say, however, that a diagnosis of T1DM is restricted to those individuals within this age range, and whilst it is rare for diagnosis to be made in individuals over 40, it is possible for a diagnosis of T1DM to occur at any age. T1DM is characterised by the destruction of the beta-cells within the pancreas which are responsible for producing insulin, a hormone used by the body to …. This loss of insulin producing beta cells means that there is an inability for T1DM patients to regulate their own blood glucose levels, which if left untreated can lead to hyperglycemia (blood glucose levels being too high). As such, T1DM patients are reliant on the subcutaneous injection of synthetic insulin to control the glycemic levels in their body and to maintain a healthy blood glucose level. Diagnosis of the condition is based on hyperglycemia, with a fasting blood glucose concentration above 7.0 mmol/L being characterised as enough to diagnose. Other symptoms commonly seen before diagnosis are going to the toilet a lot, constantly being thirsty and weight loss.

According to the World Health Organisation (WHO), blood glucose levels should be between 3.9 mmol/L and 5.6mmol/L for ordinary people and between 5 mmol/L and 9 mmol/L for diabetic patients. As a result of this, there is a constant need for diabetics to manage their blood glucose levels, with the aim of spending as much time within this range as possible and avoiding hyperglycemia and hypoglycemia (blood glucose levels too low). Both hyperglycemia and hypoglycemia are serious conditions that can lead to life threatening situations for patients; making precision when managing blood glucose levels a matter of life and death. Patients who spend extended periods of time in a hyperglycemic state can suffer from kidney damage, nerve damage (neuropathy), damage to the blood vessels of the retina leading to blindness, and complications with bones and joints. In extreme cases of hyperglycemia, diabetic ketoacidosis (DKA) can occur, which, if left untreated, can lead to a life-threatening diabetic coma. Meanwhile, hypoglycemia affords a range of different issues, ranging from slurred speech and numbness to seizures, coma and death in extreme cases if left untreated.

With an estimated 422 million people worldwide suffering from diabetes, making it the 3rd most common chronic illness in the world, it is clear that there is a need, and a great benefit to be gained, from using technology to help patients more accurately manage their blood glucose levels. Unfortunately, regulating blood glucose levels is not an exact science, and there are around 48 different factors that can impact a patient's blood glucose levels. The predominant factor is carbohydrates ingested by the patient. Carbohydrates are broken down by the amylase enzyme into glucose which is transferred into the blood stream to be absorbed by cells; clearly having a direct impact on the levels of glucose in the blood. Other major factors include any physical activity done by the patient, and any external insulin taken by the patient. Whilst these factors are more easily measurable for patients, and thus can be adapted to and controlled, other factors such as stress, illness and injury, and hormonal changes can also have a substantial impact on regulating blood glucose levels, whilst being much more difficult to manage. This makes the task of predicting how blood glucose levels will change and respond accordingly difficult for T1DM patients and machines alike - one small change can lead to large variations in behaviour.

The end vision of diabetes related technology is known as the Artificial Pancreas (AP) and consists of a fully closed-loop system where a T1DM patient would have to do no management of their blood glucose level at all apart from entering the amount of carbs the are eating; the AP would handle everything. Unfortunately, due to the problems in variation and some other issues we will cover below, the creation of this AP system still seems to be far in the future. However, that is not to say that technology has had no impact on the lives of T1DM patients, as there have been many advances which have made noticeable impacts on improving the quality of life of patients. Introduced in 1981, insulin pens are devices that allow patients to inject insulin directly into their body. Coupled together with small, easy to use glucometers which were introduced in the 1980s, patients were able to frequently monitor their blood glucose levels, and respond accordingly by either using their insulin pens, or ingesting more glucose, to control their blood glucose levels. This method of self-monitoring of blood glucose (SMBG) has significant drawbacks as blood is sampled intermittently, providing only short insights of glucose concentrations, whilst ignoring ongoing glucose fluctuations. With the idea of being able to monitor blood glucose at all times, Continuous Glucose Monitoring (CGM) devices were introduced that measure glucose concentrations at intervals (usually between 1-5 min), and then transmits/stores those values. Alongside the invention of the insulin pump, a device that allows programmable infusions of insulin over long, or short, periods of time, diabetic patients were given a much more reliable way of constantly monitoring and regulating their blood glucose levels.

The creation of CGM devices also lead to a new capability - the ability to create blood glucose datasets by accessing the data stored in the CGM; datasets which are needed to train accurate Machine Learning (ML) models. It is clear that the capability to accurately predict when one might suffer a hyperglycemic, or hypoglycemic, episode would be invaluable to a T1DM patient, as it would allow them to behave proactively to prevent the episode from ever occurring, rather than the current reactionary method, which requires a hypo/hyperglycemic episode to be taking place before we can detect its presence. We name the amount of time into the future which we forecast the Prediction Horizon (PH). Current prediction horizons available to T1DM patients are very short (~5 mins? Find a stat), and thus don't give ample time for patients to react to predicted issues. As such, much work has been done in applying ML techniques to improve the length of the PH to a length of hours instead of minutes, but as of yet no concrete solution to the problem has been found. This is due to a variety of reasons, such as the variability of a patient's BG level, due to day-to-day uncontrollable factors as mentioned earlier, as well as some other issues that will be discussed later when different approaches that have been experimented with to solve this very problem are explored.

## Main Bit

#### Intro

Machine learning is a field of Artificial Intelligence that focuses on learning through observing patterns in large quantities of past data, and iteratively improving accuracy at a given task. In the context of T1DM, the large quantities of past data refer to the patient's blood glucose levels over a period of months (as well as other points of interest such as exercise undertaken / carbohydrate intake / insulin intake), whilst the task to improve at is predicting the BG levels of the patient for a specified PH. There is interest in the use of Neural Networks (NN), a kind of Machine learning model that is able to learn very complex relationships between data by minimising the value of an error function, through iterative processes such as gradient descent. This error function can be specified to suit the model as needed and it is common, in the case of blood glucose forecasting, to see an error function comparing the absolute difference between forecasted blood glucose levels, and real recorded blood glucose levels. Much work has been done researching the applications of different kinds of NN to blood glucose forecasting, as it was believed that NNs would be able to solve the problem of inter-subject variability - by training a new network on each individual's data.

When training neural networks, there is much debate amongst researchers about what to include and what not to include within the input parameters to the network. Some believe that just the CGM data is sufficient, claiming that the inclusion of other measures, such as carbohydrate intake, would simply increase variance and decrease model performance. This is sharply contrasted by researchers who have experimented with both complex and simple physiological models to try and model the effect that factors such as exercise and insulin intake will have on future glucose levels. There is much debate about the correlation between past glucose levels and future levels, with some using only the current blood glucose levels as an input to the model, and others using many past readings to inform future predictions. Whilst there is much diversity amongst the methods employed, unfortunately, the overarching theme from most NN approaches is that they can be very accurate at shorter PH (~30 minutes), but that accuracy quickly drops off at larger PH (~6 hours).

### ANN

One such example of the application of NNs to the task of blood glucose prediction was by ((\*\*Paper 3, insert ref\*\*)), who attempted to use a NN trained using only CGM data. The model considered the preceding 20 minutes of CGM data and attempted to predict BG levels at 3 different PHs of 15, 30 and 45 minutes. The accuracy of the model was measured using the Root Mean Squared Error (RMSE) - a widely used metric for BG prediction models that takes the square root of the sum of the squares of all prediction errors - and the Prediction Delay (PD) - a metric that takes into account the delay between when certain features of the CGM data were seen in the original data and the predicted data (for instance the time gap between a peak appearing in the original and predicted data). When making a prediction, two different approaches were compared - a standard NN model that would make one prediction across the PH, and an AutoRegressive Model (ARM). In an ARM, the model makes the prediction one step at a time and uses a recursive algorithm to allow each prediction to inform subsequent predictions, trying to capture the relationships between previous and future BG levels more accurately. It was found that across the 15, 30 and 45 minute PHs there was a RMSE of ~10, 18, 27, mg/dL respectively. For upward trends in the data, there was a PD of 4, 9 and 14 minutes respectively, and for downward trends, a PD of 5, 15 and 26 minutes was seen.

This use of a basic NN highlights some of the main problems with this approach when applied to blood glucose prediction.First, the accuracy falls of very quickly, with 2.7 times worse performance over a short difference in time of 30 minutes, and second is that the model is much quicker at recognising upward than downward trends. This is a major issue, as downward trends lead to hypoglycemic episodes, which can be far more dangerous in the short term. To try and combat the first of these problems, ((\*\*Paper 5 insert ref\*\*)) included time-domain attributes along with their model. Time-domain features of the past CGM data such as minimum, maximum, mean, standard deviation etc. are passed as inputs to the model along with the data, in the hopes of improving accuracy. The approach was tested with the data of 12 different patients, across different PHs of 15, 30, 45 and 60 minutes. The accuracy found varied from patient to patient, with the best patient having RMSE values of 1.42, 3.61, 6.22 and 8.03 mg/dL respectively across the PHs and the worst obtaining values of 4.72, 11.32, 16.88 and 26.65 mg/dL respectively. Whilst this is a great improvement on the basic use of a NN, it highlights the large performance variability across patients - a variability that could result in medical emergencies if used currently.

#### RNN

Another method to try and better capture the relationship between previous CGM readings and future expected readings is a specific type of NN called a Recurrent Neural Network (RNN). An RNN differs from an ordinary 'feedforward' NN in the sense that in a NN, all inputs are passed from the input layer 'forwards' through each layer until they reach the output layer, the inputs are recursively passed into the network such that the value at a node in the network depends on both the current input and a hidden representation, which is calculated by all the previous inputs to the model. In this way, RNNs offer a way to calculate compound representations that depend on all the previous inputs to the network, rather than just passing in and handling the previous CGM data as separate datapoints. The most often seen implementation of RNNs is the Long Short Term Memory (LSTM) RNN, which implements an input, output and forget gate to control the value of the recursive hidden representation. This is done in such a way that it avoids the problems of exploding/vanishing gradient, which can affect training in vanilla RNNs.

((\*\*Paper 1 ref\*\*)) Robertson et al. [ref] used RNNs to perform BG prediction on data generated from the AIDA diabetes simulator. AIDA is a freeware diabetes simulator that models steady-state glucose-insulin interactions by describing the physiology of a person with T1DM. This model considers carbohydrate intake, insulin, kidney function and insulin sensitivity. Predictions were made on the data for both short (15-60 minutes) and long term (8-10 hours) PHs. Some points of note viewed from their experiment were that in short term prediction, there were errors between the predicted and actual Blood Glucose Levels (BGL) of over 1 mmol/L - an amount that is very unsafe in practice as the different between 3 mmol/L and 2 mmol/L is potentially dangerous. It was also noted that predictions directly after input events (e.g. meal eaten, insulin taken) were much less accurate; showing that the model wasn't able to correctly handle these events as it should. The model was also seen to be far more accurate at night, most likely due to the lack of events such as eating meals, which would lead to a dramatic change in BGL that the model cannot accurately predict. Overall, in the short term a RMSE of $0.15 \pm 0.04$ SD mmol/L was viewed and in the long term a RMSE of $0.14 \pm 0.16$ SD mmol/L was seen - again we note the trends that the model is far more unpredictable at longer term PH. Whilst these results are very promising, and the accuracy viewed is very good, it is also important to consider that the data was obtained from a mathematical model, and so is not hugely indicative of performance on real life data. The mathematical model only handles a few variables to enact changes on its output BGLs, whereas in free-living conditions there are many more factors whose fluctuations can impact BGLs in real-life patients.

The model developed by Robertson et al., was limited, performing poorly when significant events are taking place, such as the ingestion of carbs and/or insulin being delivered, which can result in rapid changes in BGL. This is a common issue in the research of BGL prediction and whilst some believe that by studying just trends in the CGM data of a patient to implicitly identify when events like this take place, and change our predictions accordingly, there is a separate approach of using Compartmental Models (CMs) to directly input the data of these events (to include carb and insulin data) into the network. CMs are mathematical functions that try and describe the processes that occur in the inaccessible parts of the human body. For instance, you would have one model describing how the carbohydrate intake at a given time would impact blood glucose levels over a given period of time afterwards. These CMs can range from simple bell curves to much more complex models such as the Hovorka model (\*\*Ref??\*\*), and there has been research done into the benefits and drawbacks of using simple or more complex CMs.

((\*\*Paper 4 Ref\*\*)) Mougiakakou et al. proposed a solution for Blood Glucose Forecasting involving the combination of CMs and RNNs. The input data is passed to three separate CMs, which estimate 1) the effect of short acting insulin intake, 2) the effect of fast acting insulin intake, and 3) the effect of carbohydrate intake on blood glucose levels. The outputs of these three CMs are then passed to the RNN along with past CGM data in order to predict subsequent blood glucose levels. Another point of note in this approach is that the RNN is trained with a Real-Time Recurrent Learning (RTRL) algorithm. Ordinarily, when training a network, you optimise the weights with a learning algorithm on a set of training data, then use those pre-determined optimised weights whenever making a prediction. Conversely, with Real Time Learning (RTL), the weights of the network continue to be improved during use, and each pair of predicted and real BGL values is used to iteratively update the network in real-time. This is thought be very applicable to BGL prediction, as you would be able to initially train a model on a batch of patient data, then continually improve the model as it was being used by the patient; permitting the model to become as personalised and accurate as possible for the user. The data for this model consisted of only 4 or 5 glucose measurements per day, over a period of ~70 days, making the data far sparser than that which has been seen in previous approaches. Overall, the approach obtained a RMSE of 41 mg/dL for predictions for each measurement; an impressively accurate result given the sparsity of the data. Mougiakakou et al. [ref] concluded that including other information in the model such as sex, age, and number of years a patient has had the condition could improve the performance of the model. Of specific interest was the potential inclusion of more CMs to capture patient physical activity.

The approach seen by Mougiakakou et al. [ref] used relatively complex, pre-defined mathematical models to describe the effects of the attributes that the CMs used were modelling. Munoz-Organero et al. [ref] ((\*\*Paper 6 ref\*\*)) attempted a different, more complex approach to the formation of their CMs, by training Deep Neural Networks to predict the impacts of Fast Insulin, Slow Insulin and Carbohydrate Intake on blood glucose levels. The output of these three models would be concatenated and fed as an input along with CGM data to give an overall prediction on the BG variation over the next time step. The deep physiological models were trained using 9 hour windows of data, and attempted to learn the BG variations over the time period, based on the three separate factors highlighted. It was noted that one of the key detriments to this kind of approach is the inconsistency of CGM datasets - some will note down every time a patient has a meal or takes insulin, whereas others will miss datapoints, which can lead to inaccuracies in the model. The models were trained and tested using data from both the AIDA diabetes simulator and the open source D1NAMO dataset (an open-source dataset collected from nine real life T1DM patients). Across a 30 minute PH, the model on the D1NAMO dataset achieved an RMSE of 6.42 mg/dL and with the simulated AIDA data, an RMSE of 3.45 mg/dL. These results show that the model was incredibly accurate over the data, with the AIDA data being the more accurate - as expected due to the added variability of real patient data. However, this accuracy is over a point of complete health for the patients, and the dynamics of insulin/carbohydrate absorption would change massively due to factors such as stress and illness. This means that, in practice, such complex deep physiological models could become very inaccurate during these times of imperfect health, and thus still are unproven and unsafe for use by real T1DM patients. It would require more real-life testing to verify whether this approach could feasibly work.

Whilst directly modelling the impacts of other factors such as carbohydrate intake with CMs is possible, there is another train of thought that we should be able to intrinsically detect these events and their impacts by looking at different signals in the CGM data. One example of this approach was taken by Wang et al. [ref] ((\*\*Paper 8 ref\*\*)) who combined Variational Mode Decomposition with an LSTM RNN optimised by Improved particle Swarm Optimisation (IPSO) to perform BG prediction. Imagine, for example, that a given T1DM eats lunch between 12-2 pm every day - we would expect to see a spike in their BG levels around this time of day every day. Through signal decomposition techniques such as VMD, we can extract the separate signals that combine to make the overall BG graph - in this case, one such signal would be a signal with a period of 1 day and a spike around the lunch time of the patient. Through this decomposition, it is thought to be possible to reduce the non-stationarity of the CGM data, and make it possible for the RNN to make more accurate predictions. To ensure the LSTM has optimal hyper parameters, Wang et al. [ref] employed an Improved Particle Swarm Optimisation algorithm - an algorithm that uses many 'particles' initialised at random points in the parameter space, then iteratively converges to the optimal solution by tracking the best particles in the current space, and initialising the next round of particles to be closer to these best ones.

To measure the success of their approach, Wang et al. [ref] used both the RMSE between predicted and actual BG levels, as well as the Clarke Error Grid (CEG) [ref]. The CEG is a tool for checking the accuracy of blood glucose monitors. It splits predictions into 5 separate zones: A, B, C, D and E. Zone A is defined as clinically accurate; Zone B indicates an incorrect, but benign, and therefore clinically acceptable prediction. Zone C represents values leading to inappropriate treatment, but without dangerous consequences for the patient. Zone D represents values leading to potentially dangerous failure to detect hypo or hyper-glycemic events and Zone E represents values leading to treat hypoglycemia instead of hyperglycemia and vice-versa. For a model to be clinically accurate, predictions should occur within zones A and B. Across a 30, 45 and 60 minute PH their model achieved an RMSE of 3.031, 5.432 and 5.716 mg/dL respectively. At the 60 minute PH, their model had 95.4% of predictions in Zone A, 3.8% in Zone B and 0.8% in Zone D. These results at short PH are very promising, but the model is so far untested at the more important long term PHs that are currently struggling to be accurately predicted over.

The approaches we have seen so far have all suffered from one common issue - a lack of data to train models on. In typical machine learning problems, datasets often consist of millions of datapoints which allows the creation of very complex and accurate models. In contrast, CGM datasets are difficult to create, and often consist of only thousands of datapoints which leads to a distinct shallowness in the data and inhibits the training of more accurate models. Aliberti et al. [ref]((\*\*Paper 9 ref\*\*)) sought to rectify this issue and instead of just training their model on the data of one patient, they trained their model on the data of many patients - whilst keeping a large chunk of the training data as the data of the main patient in question. They claimed that this would not only make more data available for the training of the model but also make their model more flexible. Past approaches that are only trained on a single patient's data require a lot of fine tuning and produce a model that is rigid and can only be used on one specific T1DM patient. In medical practice, it would be infeasible to produce an individual model for every T1DM patient and, as such, Aliberti et al. [ref] claimed that their model which would be trained across many different datasets, could offer a solution to this issue. They tried their attempt using both a Non-Linear Autoregressive NN (NAR), a network which computes the value of a signal $y$ at time $t$ using $n$ past values of $y$ as regressors, thus allowing it to model the dependency between sequential datapoints, and a LSTM RNN. Over 30, 45, 60 and 90 minute PHs, the NAR achieved RMSE of 18.2, 25.31, 33.12, 47.64 mg/dL respectively, and the LSTM received an RMSE of 5.93, 7.18, 13.21 and 28.57 mg/dL respectively. The LSTM was far more accurate than the NAR, and over a short PH this approach performed very well, but once again the performance very quickly dropped off at longer PH.

#### CNN + Others

So far all of the approaches have viewed the task of BG Prediction as a regression task where you are trying to predict a value for the BG variation over the PH. Zhu et al. [ref] ((\*\*Paper 7 ref\*\*)) instead formulated the problem as a classification problem, where the change between current glucose value and the future glucose value is quantised into 256 different classes to be predicted between. The data was sourced from the OhioT1DM dataset [ref], which is a dataset of real-life diabetes patients. Zhu et al. used a Convolution Neural Network (CNN) a type of neural network typically used in image recognition, due to its ability to reduce the dimensionality of input data. Over a 30 minute PH, the approach achieved an average RMSE of 21.72 mg/dL for its predictions, which is not as promising as other approaches we have seen involving regression via an RNN. Another approach not using the traditional RNN was experimented with by Wang et al. [ref] ((\*\*Paper 10 ref\*\*)), who built a model based on LightGBM - a gradient boosting algorithm that generates a decision tree to make predictions on the data.

### Conclusion

*currently incomplete*