Deeper into Type-1 Diabetes

Abinaya Elanchezhian[1] Jatin Bindra[1] Mehardeep Singh[1] Tanisha Chhabra[1]

[1]School of computer science, University of Windsor, Canada

i. Abstract

Type-1 diabetes is one of the major problems that are faced along the globe. With the advancement of technology more and more research labs are working on it for collecting the data, analyzing the data and applying machine learning and deep neural networks for prediction of insulin. In this project, we collected the data from a live firebase JSON data by taking a snapshot of the data. Further, using python scripts the data was analyzed and prediction was done to determine the blood glucose after 2 hours. The novelty of the project lies in the uniqueness of the dataset, and the fact that each patient would generate 3 data points per day leading to about 1000 data points at max. Creating a project especially around the data analysis of such a small data would usually take 4-5 years for any significant results. We took up this challenge and performed analysis on it. The accuracy we got was 72.92 % for Dinner, 59.35% for Breakfast, and 58.41% for lunch. Further, we wish to expand this in future by making an end to end pipeline for direct reinforcement learning of the live dataset.

ii. Introduction and Motivation

The World Health Organization states that, "[1]In 2017 there were 9 million people with type 1 diabetes; the majority of them live in high-income countries". The causes and the prevention methods of type-1 diabetes are not yet discovered. With such humongous number of people being affected by type-1 diabetes, the proposed project's focal point is to find relevant correlation between various available type - 1 diabetic data entities and predict insulin intake levels which would be an economical alternative for the presently used insulin pumps. The project uses real time patient data obtained from the android application 'Kalpavriksh' that has been developed solely to procure type 1 diabetic patient data such as food intake level, blood glucose level before intake, blood glucose level after intake, etc. The project focuses on understanding the obtained real time data and analyzing the data to draw meaningful conclusions.

iii. Related work

Many attempts have been made to apply Artificial Intelligence in predicting the insulin in the past or for managing the blood glucose. The major problem is that the body of an individual is always changing over time. People become fat-thin, sometimes do more exercise and sometimes do less. All these factors have an impact on the amount of insulin the body needs. Taylor and Forlenza[2] made an attempt to solve this problem by using a closed system insulin pump to inject insulin into the body of the patients.

An attempt has been made by Angehrn et al.[3] to study different ways in which this problem is analyzed and solved. They analyzed and realized that after applying different machine learning models, mathematical formulas, some sort of clinical support system or involvement of an expert is required for the system.
Woldaregay et al[4] conducted a very detailed research on the ways the problem was solved to tackle the prediction of insulin and blood glucose. They concluded that due to the complexity of the human body and the way in which insulin demands changes due to change in body shape, exercise and other day to day activities, it is extremely hard to predict the blood glucose after taking insulin or the insulin itself.

Given the complexity of the problem and the fact that no significant results or universal solution to this problem has been found out, this problem is extremely hard to solve.

iv. Proposed Model

Firebase is used as the database for storing the data from the application. The database supports data in JSON format. To process the JSON data following tools and libraries will be used:

Python: Python language will be primarily used for
- data fetching
- data arrangement
- data preprocessing
- application of ML model.

We have used python because it offers great libraries for dealing with data science applications.
Using python helped us to code the entire application end to end without switching the language as analysis can only be done in Python or R and firebase official documentation only supports Python and not R. Thus, we had no major choice or confusion in picking up python. Also, due to ease of use and simple syntax, Python is widely used in scientific and research communities by people without engineering backgrounds.

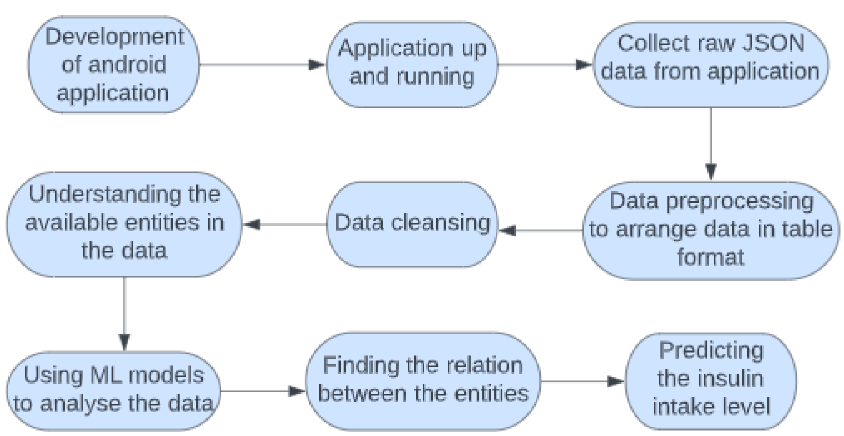

Fig.1 Application working diagram

Data fetching: In the project the data was fetched from firebase using the firebase API db.reference(). The data fetched was stored in a file .The JSON file generated was utilized by another python script for conversion into csv files.

Data arrangement: The data in JSON format was arranged in the form of CSV files that was used for data processing analysis.

Data preprocessing: The data was processed using python and analysis was done on it. All the data in the string format was converted to numbers for processing. For instance, the gender was converted to 0's and 1's representing males and females, the food intake was converted to 0-1-2 depending on breakfast, lunch and dinner.

Application of ML: We predicted the insulin intake based on all other parameters like age, gender, and carbs using models like linear regression,svm.

Sklearn:
For implementation, sklearn library was used to run the models. Sklearn was used because:

- Sklearn is free and open-source software for applying machine learning models.
- There are tons of built-in functions, ranging from calculating scores to printing confusion matrices, making it easy to use

v. Figure/diagram or result-

The results were analyzed by plotting the graphs. The graphs for each meal is plotted. The blue line shows the ground truth values and the red line shows the values predicted by the graph. The accuracy for each of the dinner, lunch and breakfast are 72.92%, 58.41% and 59.35% respectively.

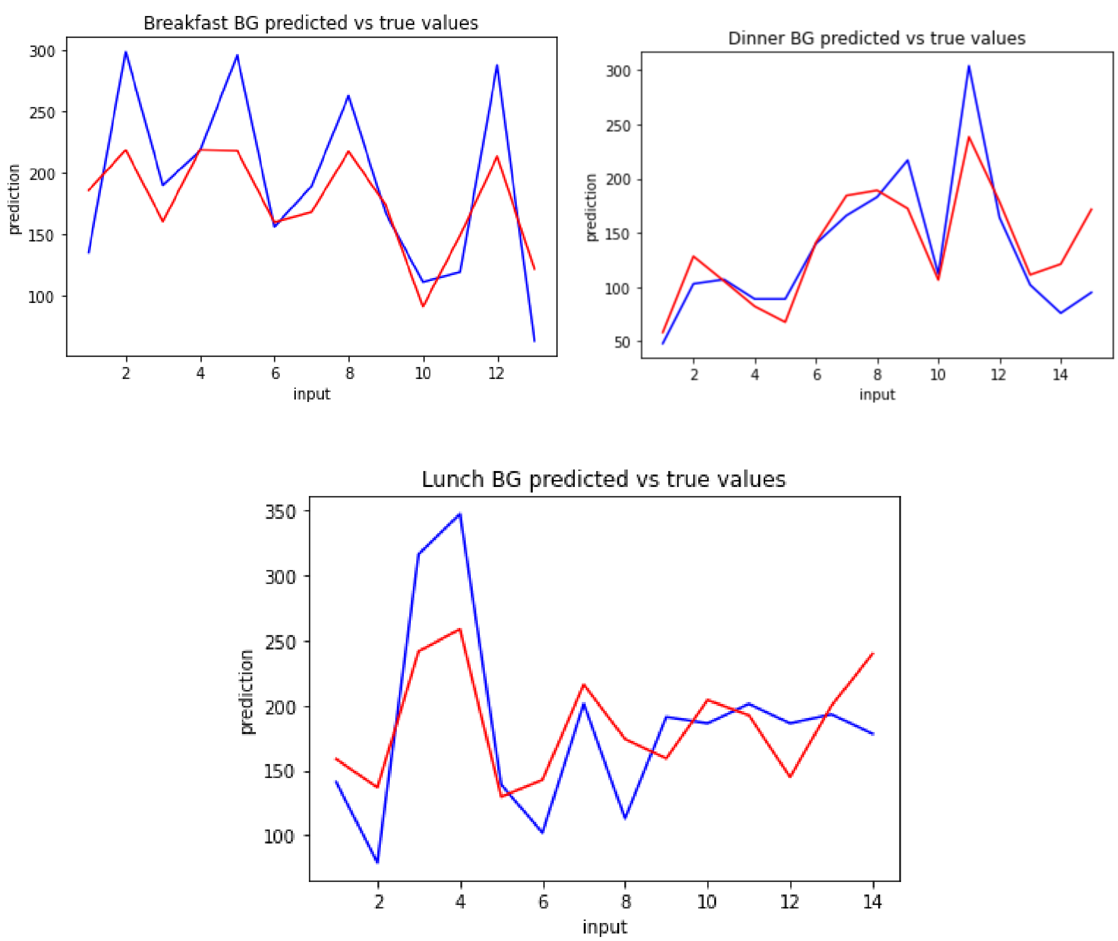

Fig. 2 Results plotted by matplotlib

The screenshots of the accuracy along with the linear regression coefficients are attached herewith for each of the lunch, breakfast and dinner.

Accuracy for the dinner:

Accuracy for the lunch:

Accuracy for the breakfast:

vi. Limitations or Challenges

The major challenges of the project are:

1. The data used for the project is obtained from the 'Kalpavriksh application. Although the application is up and running since 2018, the data is very limited for the analysis.
2. The raw data collected from the database of the 'Kalpavriksh' application was in the JSON format. It had to be pre-processed and converted into a table format before analyzing.
3. The collected data consisted of a few NULL values. This had to be altered during data cleansing.
4. The food intake values entered by the patients were "1 Indian plain Paratha", "Half cup Kadai Paneer", etc. These text values had to be carefully interpreted and the carbs value calculated.

The limitations of the project are:

LIMITATION	REASON	ALTERNATE APPROACH
Data collection and analysis from the 'Kalpavriksh' application is not real-time.	A snapshot of the collected data is obtained and the analysis is done.	Performing real-time data analysis by fetching data from the application. Real-time data analysis can be done with tools like Apache Samza, Spark Streaming, etc.
The entire analysis is done only on a single patient's data over a period of time. The accuracy and the end result is subject to vary if multiple patient data are analyzed together.	As per medical terminology, the body's metabolism behaves differently for different meals and it varies according to age, gender etc. Thus, no two patients can be clustered closely.	By trying the experiment on thousands of patients we might find two or more patients that are clustered closely enough for new comparisons, insights, and outcomes. For example, the data of one male and one female patient can be analyzed to find out if there are any significant changes in the outcome based on gender.
Fewer data points(3 per day)	A patient enters only three data every day namely, for breakfast, lunch, and dinner.	Data collection takes time. Has to wait for a few years to get a thousand data points

Table. 1 Table representing the limitations

vii. Conclusions and Future Work

Initially, we had raw data of one patient with – records. Some preprocessing was done on this dataset by scaling by using MinMaxScaler Library. For this regression problem, we used the Sklearn library to train the model by keeping the train-test ratio of 4:1. We achieved an accuracy of around 50%. As per medical terminology, the body's metabolism behaves differently for different meals, therefore Data was split on the basis of Category which enhanced the accuracy to 72.92 % for Dinner, 59.35% for Breakfast, and 58.41% for lunch. A graph was for true values vs actual values for all three categories where blue color depicted actual value and red depicted predicted value.

In the future, we can study deep learning and look to utilize neural networks in order to achieve better efficiency for this problem.

Our model can be integrated with a live database and hosted over the cloud in order to provide a real-time and scalable solution to the end-user.
After we develop a feasible model, it can be exposed to more patient data. This model can be then modified if it leads to better accuracy.

viii. References

[1]"In 2017 there were 9 million people with type 1 diabetes; the majority of them live in high-income countries"https://www.who.int/news-room/fact-sheets/detail/diabetes#:~:text=Type%201%20diabetes%20(previously%20known, live%20in%20 high%2Dincome%20 countries [Accessed on 28th March,2022].
[2] Taylor, K. A., & Forlenza, G. P. (2020). Use of Machine Learning and Hybrid Closed Loop Insulin Delivery at Diabetes Camps. *Diabetes Technology & Therapeutics*, *22*(7), 535-537.
[3] Angehrn, Z., Haldna, L., Zandvliet, A. S., Gil Berglund, E., Zeeuw, J., Amzal, B., ... & Heckman, N. M. (2020). Artificial intelligence and machine learning applied at the point of care. *Frontiers in Pharmacology*, *11*, 759.
[4] Woldaregay, A. Z., Årsand, E., Walderhaug, S., Albers, D., Mamykina, L., Botsis, T., & Hartvigsen, G. (2019). Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes. *Artificial intelligence in medicine*, *98*, 109-134.