Chinese diabetes datasets

Introduction

Data of the diabetes mellitus patients is essential in the study of diabetes management, especially when employing the data-driven machine learning methods into the management.

Type 1 diabetes happens when the body's defense system mistakenly attacks the cells in the pancreas that make insulin, so people with this type need insulin shots for life, usually starting when they're young. Type 2 diabetes, which is more common, is tied to things like being overweight and not moving enough, making the body resistant to insulin or not making enough of it, but it can often be controlled with lifestyle changes, pills, or insulin injections if needed.

ShanghaiT1DM and ShanghaiT2DM Datasets are collected to promote and facilitate the research in diabetes management. The datasets comprise of Type 1 (n=12) and Type 2 (n=100) diabetic patients in Shanghai, China.

Details

The dataset file comprises two folders named “Shanghai_T1DM” and “Shanghai_T2DM” and two summary sheets named “Shanghai_T1DM_Summary” and “Shanghai_T2DM_ Summary”.

The “Shanghai_T1DM” folder and “Shanghai_T2DM” folder contain 3 to 14 days of CGM data corresponding to 12 patients with T1DM and 100 patients with T2DM, respectively. Of note, for one patient, there might be multiple periods of CGM recordings due to different visits to the hospital, which were stored in different excel tables. In fact, collecting data from different periods in one patient can reflect the changes of diabetes status during the follow-up.

The excel table is named by the patient ID, period number and the start date of the CGM recording. Thus, for 12 patients with T1DM, there are 8 patients with 1 period of the CGM recording and 2 patients with 3 periods, totally equal to 16 excel tables in the “Shanghai_T1DM” folder. As for 100 patients with T2DM, there are 94 patients with 1 period of CGM recording, 6 patients with 2 periods, and 1 patient with 3 periods, amounting to 109 excel tables in the “Shanghai_T2DM” folder.

The summary sheets summarize the clinical characteristics, laboratory measurements and medications of the patients included in this study, with each row corresponding to one excel table in “Shanghai_T1DM” and “Shanghai_T2DM” folders. Clinical characteristics include patient ID, gender, age, height, weight, BMI, smoking and drinking history, type of diabetes, duration of diabetes, diabetic complications, comorbidities as well as occurrence of hypoglycemia. Laboratory measurements contain fasting and 2-hour postprandial plasma glucose/C-peptide/insulin, hemoglobin A1c (HbA1c), glycated albumin, total cholesterol, triglyceride, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, creatinine, estimated

glomerular filtration rate, uric acid and blood urea nitrogen. Both hypoglycemic agents and medications given for other diseases before the CGM reading were also recorded.

Fields Explanation

Overall, the excel tables include CGM BG values every 15minutes, capillary blood glucose (CBG) values, blood ketone, self-reported dietary intake, insulin doses and non-insulin hypoglycemic agents.

Each excel table in the "Shanghai_T1DM" folder and "Shanghai_T2DM" folder contains the following data fields:

1. <Date>: Recording time of the CGM data.
2. <CGM>: Continuous Glucose Monitoring data recorded every 15 minutes.
3. <CBG>: Capillary Blood Glucose level measured by the glucose meter.
   (*It provides immediate feedback, allowing individuals to adjust insulin doses, diet, and activity levels. This monitoring can be done conveniently at home with a glucose meter, helping to maintain blood sugar within target ranges and prevent complications. Continuous glucose monitoring offers real-time data for those who need closer monitoring.*)
4. <Blood ketone>: Plasma-hydroxybutyrate measured with ketone test strips (Abbott Laboratories, Abbott Park, Illinois, USA). The blood ketone was measured when diabetic ketoacidosis(DKA) was suspected with a considerably high glucose level.
   (*Ketones are chemicals produced by the liver when there's insufficient insulin to help your cells use sugar for energy. Diabetic Ketoacidosis (DKA) is a severe, life-threatening complication primarily associated with T1D but can also occur in T2D in certain situations. It happens when there's a severe lack of insulin, leading to high blood sugar levels and the production of ketones.*).
5. <Dietary intake>: Self-reported time and weighed food intake.
6. <Insulin dose-s.c.>: **Subcutaneous** insulin injection with insulin pen.
7. <Insulin dose-i.v.>: Dose of **intravenous** insulin infusion.
8. <Non-insulin hypoglycemic agents>: Hypoglycemic agents other than insulin.
9. <CSII-bolus insulin>: Dose of insulin delivered before a meal through insulin pump.
10. <CSII-basal insulin>: The rate (iu/per hour) at which basal insulin was continuously infused through insulin pump.

Insulin administration includes continuous subcutaneous insulin infusion using insulin pump, multiple daily injections with insulin pen, and insulin that were given intravenously in case of an extremely high BG level. CSII is treatment for type 1 and type 2 inpatients, i.e., before discharged from hospital.