**Data Source:** <https://archive.ics.uci.edu/ml/datasets/Diabetes+130-US+hospitals+for+years+1999-2008>

**Medical Code Referencing:**<http://www.icd9data.com/>

**Paper on the data**: <https://www.hindawi.com/journals/bmri/2014/781670/>

**Feature Descriptions:** <https://www.hindawi.com/journals/bmri/2014/781670/tab1/>

Table 1: List of features and their descriptions in the initial dataset (the dataset is also available at the website of Data Mining and Biomedical Informatics Lab at VCU (http://www.cioslab.vcu.edu/)).

Feature name Type Description and values % missing

Encounter ID Numeric Unique identifier of an encounter 0%

Patient number Numeric Unique identifier of a patient 0%

Race Nominal Values: Caucasian, Asian, African American, Hispanic, and other 2%

Gender Nominal Values: male, female, and unknown/invalid 0%

Age Nominal Grouped in 10-year intervals:

0, 10), 10, 20), …,

90, 100) 0%

Weight Numeric Weight in pounds. 97%

Admission type Nominal Integer identifier corresponding to 9 distinct values, for example, emergency, urgent, elective, newborn, and not available 0%

Discharge disposition Nominal Integer identifier corresponding to 29 distinct values, for example, discharged to home, expired, and not available 0%

Admission source Nominal Integer identifier corresponding to 21 distinct values, for example, physician referral, emergency room, and transfer from a hospital 0%

Time in hospital Numeric Integer number of days between admission and discharge 0%

Payer code Nominal Integer identifier corresponding to 23 distinct values, for example, Blue Cross/Blue Shield, Medicare, and self-pay 52%

Medical specialty Nominal Integer identifier of a specialty of the admitting physician, corresponding to 84 distinct values, for example, cardiology, internal medicine, family/general practice, and surgeon 53%

Number of lab procedures Numeric Number of lab tests performed during the encounter 0%

Number of procedures Numeric Number of procedures (other than lab tests) performed during the encounter 0%

Number of medications Numeric Number of distinct generic names administered during the encounter 0%

Number of outpatient visits Numeric Number of outpatient visits of the patient in the year preceding the encounter 0%

Number of emergency visits Numeric Number of emergency visits of the patient in the year preceding the encounter 0%

Number of inpatient visits Numeric Number of inpatient visits of the patient in the year preceding the encounter 0%

Diagnosis 1 Nominal The primary diagnosis (coded as first three digits of ICD9); 848 distinct values 0%

Diagnosis 2 Nominal Secondary diagnosis (coded as first three digits of ICD9); 923 distinct values 0%

Diagnosis 3 Nominal Additional secondary diagnosis (coded as first three digits of ICD9); 954 distinct values 1%

Number of diagnoses Numeric Number of diagnoses entered to the system 0%

Glucose serum test result Nominal Indicates the range of the result or if the test was not taken. Values: “>200,” “>300,” “normal,” and “none” if not measured 0%

A1c test result Nominal Indicates the range of the result or if the test was not taken. Values: “>8” if the result was greater than 8%, “>7” if the result was greater than 7% but less than 8%, “normal” if the result was less than 7%, and “none” if not measured. 0%

Change of medications Nominal Indicates if there was a change in diabetic medications (either dosage or generic name). Values: “change” and “no change” 0%

Diabetes medications Nominal Indicates if there was any diabetic medication prescribed. Values: “yes” and “no” 0%

24 features for medications Nominal For the generic names: metformin, repaglinide, nateglinide, chlorpropamide, glimepiride, acetohexamide, glipizide, glyburide, tolbutamide, pioglitazone, rosiglitazone, acarbose, miglitol, troglitazone, tolazamide, examide, sitagliptin, insulin, glyburide-metformin, glipizide-metformin, glimepiride-pioglitazone, metformin-rosiglitazone, and metformin-pioglitazone, the feature indicates whether the drug was prescribed or there was a change in the dosage. Values: “up” if the dosage was increased during the encounter, “down” if the dosage was decreased, “steady” if the dosage did not change, and “no” if the drug was not prescribed 0%

Readmitted Nominal Days to inpatient readmission. Values: “<30” if the patient was readmitted in less than 30 days, “>30” if the patient was readmitted in more than 30 days, and “No” for no record of readmission. 0%