Dataset name	Variables	Source	N=?	Outcome and summary
GDM prediction	Id Age Ethnicity (0 = white; 1 = not white) DM status Diastolic BP Systolic BP Central fat Gestational age Pregnancies Fasting glucose Prepreg BMI Delivery type (0 = vaginal birth; 1 = cesarian section) Child birth wt GDM diagnosis (0 = no; 1 = yes)	https://www.kaggle.com/datasets/kamyababedi/visceral-adipose	133	This dataset has visceral fat measurements, demographic, and other body measurements that can be used to predict the occurrence of GDM in pregnant women. This is a binary classification outcome, but other regression predictions can be made using other variables.
SEER Breast Cancer	Age Race Marital status T stage N stage 6th stage Grade A stage Tumor size Estrogen status Regional node exam Regional node positive Survival months Living status	https://ieee-dataport.org/open-access/seer-breast-cancer-data	4024	This dataset has demographic, cancer staging, and tumor specifics that can be used to predict the mortality of a breast cancer patient. This is a binary classification outcome of living or dying.
Drug Class Prediction	Age Sex BP Cholesterol Na/K ratio Drug type	https://www.kaggle.com/datasets/prathamtripathi/drug-classification	200	This dataset has demographic and cardiovascular measurements. This is a classification outcome of 5 drug types (A,B,C, X,Y).
Stroke Data	Gender Age Hypertension (0=no,	https://www.kaggle.com/datasets/zze	201	This dataset has demographic and cardiovascular measurements. This is a binary classification outcome of having a

	1=yes) Heart disease (0=no, 1=yes) Ever married Work type Residence Avg glucose BMI Smoking status Stroke incidence	ttrkalpakbal/full-filled-brain-stroke-dataset		stroke or not.
Cirrhosis	ID N of Days Status Drug Age Sex Ascites Hepatomegaly Spider Edema Bilirubin Cholesterol Albumin Copper Alk_Phos SGOT Triglycerides Platelets Prothrombin Stage	https://www.kaggle.com/datasets/fedesoriano/cirrhosis-prediction-dataset	418	This dataset has demographic, blood work, and hepatitis measurements. This is a binary classification outcome, but other regression predictions can be made using other variables.