

**Supplemental Table S1: Illustration of adapting 7-NN classification model to perform risk assessment<sup>@</sup>.**

Risk assignment	Split (number of training points closest to the test point)	Count of diabetic patients assigned the risk specified in column 1	Count of nondiabetic participants assigned the risk specified in column 1
<b>Split Schema 1</b>			
“Low” risk	(0 to 1)	353	5501
“Borderline” risk	(2 to 3)	811	1558
“High” risk	(4 to 7)	1689	720
Total participants		2853	7779
<b>Split Schema 2</b>			
“Low” risk	(0 to 2)	718	6500
“Borderline” risk	(2 to 3)	1064	991
“High” risk	(4 to 7)	1071	288
Total participants		2853	7779

<sup>@</sup>,Two exemplary ‘split’ schemes are presented. It is expected that a high fraction of diabetic patients are assigned “high” risk and a low fraction are assigned “low” risk. Similarly, it is expected to see that a high fraction of nondiabetic participants are assigned “low” risk and that a low fraction are assigned “high” risk. It is clear that with split schema 2, more diabetics are assigned low and borderline classes, while more non-diabetics are being assigned to lower risk classes. Since it is not advisable that diabetic patients go undetected, split schema 1 is more (medically) acceptable, and it is the one that is used in the reported study. Similar experiments suggest using the same split for hypertension risk assessment as well.