

# Applying Machine Learning Techniques for Type2 Diabetes Readmission Prediction Based on Retrospective Data

Abhishek Sharma<sup>1</sup>, Nan Kong<sup>2,5</sup>, Paul M. Griffin<sup>3,5</sup>, Andrés García-Arce<sup>4</sup>, Munirul Haque<sup>5</sup>  
Computer Science, Purdue University<sup>1</sup>, Biomedical Engineering, Purdue University<sup>2</sup>, Industrial Engineering,  
Purdue University<sup>3</sup>, Geisinger Health Systems<sup>4</sup>, Regenstrief Center for Healthcare Engineering (RCHE), Purdue  
University<sup>5</sup>

## ABSTRACT

Roughly 9.3% of US population suffer from diabetes and the 30 day readmission rate for diabetes patients range between 14.4 to 22.7%. Hence identifying the risk of readmission is a crucial information for the service providers to not only reduce the healthcare cost but also improve the quality of patient care. This paper models machine learning algorithms to compute probability of 30-day hospital readmission for type2 diabetes patients. Along with novel pre-processing techniques to identify the challenges of noisy and non-homogenized medical data, we used methods to downsize the feature vector size without sacrificing prediction accuracy. Our method has been implemented on a publicly available dataset from University of California Irvine at <https://archive.ics.uci.edu/ml/datasets/diabetes+130-us+hospitals+for+years+1999-2008> which summarizes data from 130 US hospitals within span of 10 years with roughly 100,000 patients.

## KEYWORDS

Hospital Readmission, Type2 Diabetes, Healthcare and Machine Learning.