**ABSTRACT**

Diabetes mellitus, known as diabetes, is a group of metabolic disorders and has affected hundreds of millions of people. The detection of diabetes is of great importance, concerning its severe complications. There have been plenty of research studies about diabetes identification, many of which are based on the Pima Indian diabetes data set. It’s a data set studying women in Pima Indian population started from 1965, where the onset rate for diabetes is comparatively high. Most of the research studies done before mainly focused on one or two particular complex technique to test the data, while a comprehensive research over many common techniques is missing. In this paper, we make a comprehensive exploration to the most popular techniques (e.g. DNN (Deep Neural Network), SVM (Support Vector Machine), etc.) used to identify diabetes and data preprocessing methods. Basically, we examine these techniques by the accuracy of cross-validation on the Pima Indian data set. We compare the accuracy of each classifier over several ways of data preprocessors and we modify the parameters to improve their accuracy. The best technique we find has 77.86% accuracy using 10-fold cross-validation. We also analyze the relevance between each feature with the classification result.