

CHAPTER 1

INTRODUCTION

1.1. Background

The number of patients that had infected by Diabetes Mellitus (DM) has reached 415 million patients in 2015 and by 2040 this number will rise approximately up to 642 million patients (IDF, Diabetes Atlas 7th Edition, 2015). This number are significantly increase, due in 2013 number of patients that had infected by DM are 382 million (IDF, Diabetes Atlas 6th, 2013) (Lukmanto & Irwansyah, 2015).

One of the most common characteristic of DM is patient has a high blood glucose level caused by body deficiency or resistance to insulin which can lead to serious health complication (Beloufa & Chikh, 2013) (Varma, Rao, Lakshmi, & Rao, 2014). Meanwhile, as mentioned on previous statement that DM is a complex disease that can lead to serious health complications, such as: Coronary Artery disease, Peripheral Vascular disease, Peripheral Neuropathy, Liver dysfunction and Sexual dysfunction in woman (Feinglos & Bethel, 2008).

In medical expertise, there are several of test that can be used to detect the DM. The most common and well-known tests are referring to Random Plasma glucose test, Fasting Plasma glucose (FPG) and Oral Glucose Tolerance Test (OGTT) (Feinglos & Bethel, 2008). There is also another supplementary indicator that can be used to strengthen the blood glucose level measurement such as age and symptoms (ADA, 2010) (Lukmanto & Irwansyah, 2015).

Nowadays, many of DM care solutions are provided in the market. People can either go to the hospital or find the portable glucose meter at pharmacy. The challenge was people might not have sufficient money to go to hospital as well as the accuracy of portable glucose meters are depending on the quality. Rapid grow in IT has encourage scientist to optimize the prediction accuracy as well as to reduce the medical cost.

Based on PID (Pima Indian Diabetes) datasets, which is known as one of the most popular dataset that being used in DM research literature. there are 8 variables that can be used to indicate whether is there any potential of DM in someone (Hayashi & Yukita, 2016), including: Number of pregnancy (Times), plasma glucose concentration after 2 h, Diastolic blood pressure (mmHg), Triceps skin fold thickness (mm), 2-hour serum insulin (μ U/mL), BMI, Diabetes pedigree function and age.

Meanwhile, come up with huge amounts of labeled medical data that has a relevancy with DM is another benefit to fights against DM, one of the most well-known and powerful method in data processing is a classifier technique, which is currently has been identified that being used widely to transform on how to diagnose various of complex disease including DM (Zhu, Xie, & Zheng, 2015). Refers to some of well-known public DM datasets, there are some of inputs variable that can be used as a baseline for the classifier method to generate the output, afterwards the output will be classified based on the potential.

Having an issues with low output accuracy is a primary case that need to be solved in detecting DM, one of the effort to detect DM by merging FIS and AHP had been proposed by (Lukmanto & Irwansyah, 2015) aimed to reduce the number of rules in inference system without any data assessment step prior the inference process (Lukmanto & Irwansyah, 2015), unfortunately the accuracy is still below than 90%.

Detecting DM using multiple classifier system has been proposed by (Zhu, Xie, & Zheng, 2015), meanwhile data preprocessing is also unavailable prior the classification process and it also been suggested to implements genetic algorithm in the future (Zhu, Xie, & Zheng, 2015). Another effort to detect the DM using Fuzzy computation also been proposed by (Meza-Palacios, et al., 2017) also proposed a method with the implementation fuzzy expert system which collecting the data from the expert, meaning there is no data preprocessing step based on real labeled dataset and focus to be aimed to replicate on how the expert diagnose the DM.

One of the most popular classifier method is Fuzzy Inference System (FIS), which is being used in various area of research to perform the classification process. However, these systems have been regarded to have lower performance than Support Vector Machine (SVM). Therefore, in this study, we aim to investigate the hybridization of FIS and SVM, known as fuzzy SVM.

Fuzzy SVM computation as a part of computational intelligence implementation that can be used as a classifier method has an advantage to be adopted due it will optimize the classifier computation process by avoiding the sharp cut-off performing by crisp or classical classification method. This research will also leverage F-Score feature selection to increase the capability of Fuzzy SVM computation in executing the classification process, due the previous researches are still fully leverage Fuzzy SVM without any data measurement prior the classification step.

Hereinafter, the inputs variables of the datasets need to be selected before been chosen as an input variables of the Fuzzy inference, in this study, F-score selection method will be performed to measure the most usable and advantageous input variables by doing a selection process for input variables of the datasets, then the selected variables will be used as an input.

The key objective and motivation of this research is to discuss and share the potential of adopting machine learning to optimize the fuzzy rules in detecting Diabetes Mellitus without sacrificing the accuracy result to support medical analysis.

1.2. Research Problem

This section will describe the research problem, some of the identified problem will be included in this section as a baseline in this research, some of the problems are:

1. How pre-processing and feature selection can be leveraged in this research to select the most informative feature in dataset instead of using

full set of datasets, so that the process and result of the classification can be simplified and optimized?

2. How Fuzzy SVM will be synergized in this research to produce the fuzzy rules based on learning result instead of expert approach?

1.3. Research Objective

This section will describe the main objective of this research that will be linked to the research problem:

1. By eliminating the data through data pre-processing step and adopting the F-Score selection to identify the most informative feature, this research has been targeted to simplify the process of classification without sacrificing the accuracy result.
2. The synergy of Fuzzy SVM in this research has been targeted to avoid the sharp cut-off that can give an impact in reducing the output accuracy.

1.4. Scope of Research

This section will be showing up the scope of research to ensure the research are keep on track based on the research cases and meet the objective:

1. This research is proposed by using most common datasets as an input variable for the classifier method, PID as one of well-known DM datasets will be used in this research
2. F-Score features selection will be chosen as a method to measure the variables in datasets and retrieve the most appropriate variables to be used in inference process
3. This research will adopt Fuzzy SVM method to avoid sharp cut-off in classification process
4. All concepts and techniques that have been proposed and implemented in this research are only applicable in detecting Diabetes Mellitus