
A Hybrid Intelligent System Design for Diabetes Risk Classification

Thirumalaimuthu Thirumalaiappan Ramanathan, B.E (AU)



Faculty of Education, Science, Technology and Mathematics

University of Canberra

November 2015

*A Dissertation Submitted in partial fulfillment of the requirements for the degree of
Master of Information Sciences (Research)*

© 2015

Thirumalaimuthu Thirumalaiappan Ramanathan

ALL RIGHTS RESERVED

Abstract

Risk classification is a major technical challenge in medical diagnosis and chronic illness management. Various computational techniques have been developed for risk classification in recent years with improvements in outcomes. This thesis investigates a novel approach combining support vector machine (SVM) and fuzzy modelling. The proposed hybrid model (called SVM-Fuzzy) is designed, implemented and evaluated on an available benchmark dataset.

Diagnosis and management of diabetes mellitus (also known as type 2 diabetes) are the motivating problem for the current investigation of risk classification. Type 2 diabetes is a chronic condition marked by elevated levels of blood glucose. The prevalence of diabetes is increasing at a fast pace due to obesity, in particular, central obesity, physical inactivity, and unhealthy dietary habits. In type 2 diabetes patients, a range of different organs are under stress and are influenced by the altered metabolic condition. Hence the ailment has to be properly managed within stipulated boundaries to minimize other health complications caused by type 2 diabetes and to assure longevity of life. In the proposed model, fuzzy reasoning is used to classify the level of risks from data. SVM is used to design the fuzzy rules. The well-known Pima Indian diabetes dataset (Frank, 2010) is used to train the SVM and for testing the fuzzy system. The goal is to evaluate the proposed design for better accuracy in risk classification and to investigate training the machine learning algorithm using sample real world data. Another goal is also to investigate efficiency in classification by optimizing selection of right sized datasets through appropriate dataset size selection from experiments. The experiments from the SVM-Fuzzy model show promising results on the benchmark Pima Indian diabetes dataset.

Early in the thesis the various soft computing approaches, such as artificial neural network, SVM, Bayesian network and fuzzy logic models which have been previously reported as relevant to diabetes risk classification are reviewed. The drawbacks in these models are analyzed and identified gaps from the previous computational models on risk classification are targeted in the research questions.

The main objective in risk classification is to handle uncertain and incomplete input data for risk classification and to make use of sample datasets using fuzzy reasoning. The experimental results from the proposed model (SVM-Fuzzy) from the type 2 diabetes risk classification problem demonstrates that it is able to handle uncertain input data by making use of sample datasets for fuzzy reasoning. The outcomes from the proposed

model on type 2 diabetes risk classification are summarized and compared with the previous computational models. These experimental results are encouraging. The SVM-Fuzzy model can be further extended to manage type 2 diabetes risks using its output risk values in planning lifestyle and diet for responsive management of type 2 diabetes conditions. The model can further be extended to a multi-agent learning and planning environment for solving classification problems in complex applications, such as extracting patterns in structure of proteins in bioinformatics.

Acknowledgement

I thank my God for giving me strength and knowledge to conduct the research and write this thesis. I would like to convey my gratitude to my supervisor Professor Dharmendra Sharma for his valuable contribution and expertise. I discussed all my ideas with him and he gave me specific direction to follow on. His guidance, motivation and encouragement helped in every step of this research work. My thanks to Professor Xu Huang, Associate Professor Dat Tran and Dr Chris Barnes provided for their guidance and support. I would also like to extend my appreciation to the Faculty of Education, Sciences, Technology and Mathematics for their support for my research. My special thanks to my parents and my uncle Professor Kaliappa Kalirajan (from the Australian National University) for keeping me motivated and for all their love and understanding throughout the journey.

CONTENTS

Abstract	iii
Acknowledgement	vii
List of Tables and Figures	xiii
Chapter 1: Introduction	1
1.1 Motivation	1
1.2 Research problem	2
1.3 Objectives	3
1.4 Research questions	3
1.5 Methodology	3
1.6 Thesis organization	4
Chapter 2: Advances in Computational Classification: A Review	5
2.1 Introduction	5
2.2 Artificial Neural Network	5
2.2.1 Artificial Neural Network for Classification	6
2.2.2 Drawbacks of Artificial Neural Network in Classification	6
2.3 Support Vector Machine	6
2.3.1 Intelligible SVM Model for Classification	7
2.3.2 Region based SVM algorithm for Classification	8
2.3.3 Drawbacks of SVM Models in Classification	9
2.4. Fuzzy logic	9

2.4.1 Mamdani-Style Fuzzy Inference	10
2.4.2 Fuzzy Logic for Classification	11
2.4.3 Drawbacks of Fuzzy System in Classification	11
2.5. Hybrid Models	11
2.5.1 Hybrid Model using Modified PSO and Least Squares SVM for Classification	12
2.5.1.1 Particle Swarm Optimization	12
2.5.1.2 Modified-PSO	12
2.5.1.3 Least Squares-SVM	13
2.5.1.4 Integrating Modified-PSO and Least Squares-SVM	13
2.5.2 A Soft Intelligent Binary Model for Classification	14
2.5.2.1 Multi-Layer Perceptron	14
2.5.2.2 Multi-Layer Perceptron using Fuzzy Logic	15
2.5.3 A Hybrid Genetic Algorithm for Classification	15
2.5.3.1 Genetic Algorithm	15
2.5.3.2 K-Nearest-Neighbor	16
2.5.3.3 The Weighted KNN Algorithm	16
2.5.3.4 The Distance-WKNN Algorithm	16
2.5.3.5 Combined use of Genetic Algorithm and K-Nearest-Neighbor	17
2.5.4 Drawbacks of Hybrid Models in Classification	17
2.6 Bayesian Network	17
2.6.1 Bayesian Network for Classification	18
2.6.2 Drawbacks of Bayesian Learning in Classification	18
2.7 A Neuro-Fuzzy System	18
2.7.1 Neuro-Fuzzy for Classification	19

2.7.2 Drawbacks of Neuro-Fuzzy in Classification	19
2.8 Summary	19
Chapter 3: Risk Classification: Proposed SVM-Fuzzy Model	21
3.1 Introduction	21
3.2 Foundation of Risk	21
3.3 Risk Classification	22
3.4 Importance of Dataset in Risk Classification	23
3.5 Proposed Algorithm: SVM-Fuzzy	24
3.6 Pima Indian Diabetes Dataset	25
3.7 Summary	25
Chapter 4: System Design for SVM-Fuzzy and testing results	27
4.1 Introduction	27
4.2 Architecture of SVM-Fuzzy	27
4.3 Fuzzy sets	28
4.4 Fuzzy Rules in SVM-Fuzzy	32
4.5 Defuzzification in SVM-Fuzzy	38
4.6 Summary	39
Chapter 5: Performance Analysis of SVM-Fuzzy Model	41
5.1 Introduction	41
5.2 Result Analysis for SVM-Fuzzy system	41
5.3 Classification accuracy	43
5.4 Summary	43
Chapter 6: Conclusion and Further Work	45

Appendix A: Publications	49
Appendix B: Pima Indian Diabetes Dataset	51
References	53

List of Tables and Figures

Table 2.1 Classification accuracy of reviewed models	20
Table 4.1 Ranges of input and output fuzzy sets	31
Table 5.1 SVM-Fuzzy Results	42
Figure 2.1 Generic architecture of a fuzzy expert system	10
Figure 4.1 SVM based Fuzzy expert system architecture	28
Figure 4.2 Fuzzy sets for the input BMI	29
Figure 4.3 Fuzzy sets for the input blood pressure	29
Figure 4.4 Fuzzy sets for the input glucose concentration	30
Figure 4.5 Fuzzy sets for the input serum insulin	30
Figure 4.6 Classification fuzzy set	31
Figure 4.7 Training from first group dataset	33
Figure 4.8 Training from second group dataset	33
Figure 4.9 Training from third group dataset	34
Figure 4.10 Rule viewer (rule 1-32)	35
Figure 4.11 Rule viewer (rule 33-66)	36
Figure 4.12 Rule viewer (rule 67-96)	37
Figure 4.13 Centroid of Gravity method	38