Contents

1 Introduction **1**
- 1.1 Motivation . . . 1
- 1.2 Hypothesis . . . 3
- 1.3 Objective . . . 3
  - 1.3.1 General Objective . . . 3
  - 1.3.2 Specific Objectives . . . 4
- 1.4 Methodology . . . 4
- 1.5 Thesis Structure . . . 5

2 Related Work **7**
- 2.1 Supervised Learning . . . 7
- 2.2 Interpretable methods . . . 8
  - 2.2.1 Decision Trees and Random Forest . . . 9
  - 2.2.2 Bayesian derived methods . . . 10
- 2.3 Interpretability indicators in Fuzzy Rule-Based Systems . . . 11
- 2.4 Interpretability in other methods . . . 12
- 2.5 Dempster Shafer Theory . . . 14
  - 2.5.1 Mass assignment functions . . . 14
  - 2.5.2 Belief and Plausibility . . . 15
  - 2.5.3 Probability estimation . . . 17
  - 2.5.4 Dempster Rule . . . 17
  - 2.5.5 Commonality functions . . . 19
- 2.6 Dempster Shafer Theory applications in supervised learning . . . 20
- 2.7 Mathematical methods . . . 22
  - 2.7.1 Optimization Methods . . . 22
  - 2.7.2 Automatic differentiation . . . 27

3 Proposed Model **33**
- 3.1 Classification . . . 33
  - 3.1.1 Dempster Shafer Implementation . . . 33
  - 3.1.2 Rules . . . 36
  - 3.1.3 Support . . . 36
  - 3.1.4 One-attribute Rule Generators . . . 38
  - 3.1.5 Two-attribute Rule Generators . . . 40
  - 3.1.6 Prediction . . . 40
- 3.2 Optimization . . . 42

3.2.1 Loss Functions . . . 43
3.2.2 Gradient Descent . . . 43
3.2.3 Projection . . . 45
3.2.4 Convergence Conditions . . . 47
3.2.5 Batching . . . 47
3.2.6 Pseudo-code for the optimization process . . . 48
3.3 Interpretability . . . 48
3.4 Improvements and Implementation notes . . . 50
3.4.1 Rule predicate precomputation . . . 50
3.4.2 Dempster Rule vectorization . . . 50
3.4.3 Commonality transformation . . . 51
3.4.4 Commonality and normalization . . . 52
3.4.5 Plausibility probability estimation and commonality . . . 53
3.5 Model Complexity . . . 53
3.5.1 Prediction . . . 54
3.5.2 Training . . . 54
3.5.3 Interpretability . . . 54

4 Results **55**
4.1 Controlled Scenarios . . . 55
4.1.1 1-D distribution . . . 55
4.1.2 Full random classification . . . 57
4.1.3 Imperfect classification . . . 60
4.1.4 2D distribution . . . 62
4.1.5 2D Gaussian distributions . . . 66
4.1.6 Multi-class classification . . . 67
4.1.7 Commonality transformation . . . 70
4.1.8 Multiple kinds of attributes . . . 72
4.1.9 Scikit learn classifier comparison datasets . . . 73
4.2 Traditional Datasets . . . 75
4.2.1 Iris Dataset . . . 77
4.2.2 Wine quality dataset . . . 77
4.2.3 Heart Disease Dataset . . . 77
4.2.4 Breast Cancer Dataset . . . 78
4.2.5 Handwritten Digits Dataset . . . 80
4.2.6 Gas sensor array drift . . . 82
4.2.7 Summary . . . 82

5 Application to Stroke risk prediction **83**
5.1 Motivation . . . 83
5.2 Data Description . . . 84
5.3 Data Exploration . . . 85
5.4 Embedding . . . 88
5.5 Rules and Configuration . . . 89
5.6 Model Results . . . 91
5.6.1 Comparison with other Machine Learning Models . . . 92
5.6.2 Comparison with Deep Learning methods . . . 95

5.6.3 Comparison with other Stroke risk assessment methods 97

6 Validation **99**
6.1 Interpretability Results 99
6.2 Explanation models 101
6.2.1 Decision Trees 101
6.2.2 Partial dependency plots 104
6.2.3 LIME 106
6.3 Contrasting the Model with Medical Literature 109
6.3.1 Cerebrovascular disease in the past 110
6.3.2 Low values of platelets 110
6.3.3 High values of Hemoglobin concentration 110
6.3.4 Diabetes condition 111
6.3.5 High level of body fat 111
6.3.6 High rates of HDL-C 111
6.3.7 High values of waist measurements 112
6.4 Expert Survey 112

7 Conclusion **115**
7.1 Future Work 116
7.2 Contributions 117

Bibliography **118**

Appendix **126**