Contents

1 Motivation and Background **1**
1.1 Introduction 2
1.1.1 Motivation 2
1.1.2 Problem Statement 7
1.1.3 Research Questions 18
1.1.4 Approach to Novelty Detection 19
1.1.5 Structure of the Thesis 21
1.2 Background and Methodology Foundations 22
1.2.1 Terms and Definitions 23
1.2.2 Notation & Mathematical Notation 28
1.2.3 Experimental Evaluation 35
1.2.4 Data Streams 50
1.2.5 Anomaly, Novelty, and Outlier Detection 51
1.2.6 Novelties and Concept Drift 57
2 Literature Review **63**
2.1 State of the Art 64
2.1.1 Surveys 65
2.1.2 Approaches 81
2.1.3 Concept Drift 86
2.1.4 Evaluation Methodology 88
2.1.5 Literature Study – Evaluation & Data Sets 90
2.2 Conclusion 94
3 Baseline Techniques **97**
3.1 Methodology 99
3.1.1 Experimental Setup 99

3.1.2 Data Sets 100
3.2 Competing Approaches 100
3.2.1 Decision Based 106
3.2.2 Isolation Based Detectors 114
3.2.3 Neighbour Based Detectors 123
3.2.4 Novelty Detectors 127
3.3 Results & Comparison 137
3.3.1 Analysis 139
3.3.2 Discussion 140
4 The CANDIES Methodology **145**
4.1 Combined Approach to Novelty Detection 147
4.1.1 Demarcation from Existing Approaches 147
4.1.2 The CANDIES Pattern 149
4.2 Probabilistic Model 152
4.2.1 Gaussian Mixture Models 153
4.2.2 Context Buffer 161
4.3 Anomaly Detection 163
4.3.1 HDR/LDR Separation 163
4.3.2 AlphaDetector 165
4.3.3 Discussion 168
4.4 HDR Detection 170
4.4.1 The Chi-squared HDR detector 171
4.4.2 Discussion 176
4.5 Locality Matching 179
4.5.1 GMM Locality 180
4.5.2 Discussion 181

4.6 Novelty Detection 184
4.6.1 LDR Novelty Detection 184
4.6.2 HDR Novelty Detection 197
4.6.3 Obsoleteness 201
4.6.4 Propagation 209
4.6.5 Discussion 211
4.7 Concept Drift 214
4.7.1 Normal-Wishart Drift Compensation 215
4.7.2 Discussion 222
4.8 Implementation 225
4.8.1 ndnet 226
4.8.2 mCANDIES 228
4.9 Conclusion 235
5 Evaluation **239**
5.1 Evaluation Methodology 241
5.1.1 Online Evaluation 244
5.1.2 Solution Proposal: FSM-Eval 249
5.1.3 Discussion 255
5.2 Synthetic Data Generation 257
5.2.1 Sampling Processes 259
5.2.2 Scenario Generation 274
5.3 Experimental Evaluation 276
5.3.1 NDT Detectors 277
5.3.2 Expresso – Experimentation Framework 279
5.3.3 Results on Baseline Data Sets 284
5.3.4 Synthetic Data: Anomaly Detection 289

5.3.5 Synthetic Data: Novelty Detection 295
5.4 Conclusion 313
6 Conclusion & Outlook **319**
6.1 Conclusion 320
6.1.1 Discussion 322
6.1.2 Scientific Contributions 324
6.2 Outlook 327
6.2.1 Future Work 328
Appendices **329**
A Data Sets Tables **331**
B Performance Tables **337**
B.1 Baseline Techniques 337
B.2 NDT Detectors - FDS Statistics 345
Bibliography **349**
Acronyms **371**
Index **377**