

**A NOVEL ALGORITHM FOR IMPROVED TRACTABILITY AND
EXPLAINABILITY OF INFERENCES ON HIGH-DIMENSIONAL
DATA USING CONDITIONED ATTENTION MECHANISM**

**ASHUZA KUDERHA RIFIN
(22PCG02636)**

**B.Sc Computer Science, Université Catholique de Bukavu, Bukavu, DR
Congo**

AUGUST, 2024

**A NOVEL ALGORITHM FOR IMPROVED TRACTABILITY AND
EXPLAINABILITY OF INFERENCES ON HIGH-DIMENSIONAL
DATA USING CONDITIONED ATTENTION MECHANISM**

BY

**ASHUZA KUDERHA RIFIN
(22PCG02636)**

**B.Sc. Computer Science, Université Catholique de Bukavu, Bukavu, DR
Congo**

**A DISSERTATION SUBMITTED TO THE SCHOOL OF
POSTGRADUATE STUDIES IN PARTIAL FULFILMENT OF THE
REQUIREMENT FOR THE AWARD OF MASTER OF SCIENCE
(M.Sc.) DEGREE IN COMPUTER SCIENCE IN THE DEPARTMENT
OF COMPUTER AND INFORMATION SCIENCES, COVENANT
UNIVERSITY, OTA, OGUN STATE**

AUGUST, 2024

ACCEPTANCE

This is to attest that this dissertation is accepted in partial fulfillment of the requirements for the award of the degree of Master of Science in Computer Science in the Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Ogun State, Nigeria.

Miss Adefunke F. Oyinloye
(Secretary, School of Postgraduate Studies)

Signature and Date

Prof. Akan B. Williams
(Dean, School of Postgraduate Studies)

Signature and Date

DECLARATION

I, ASHUZA KUDERHA Rifin (22PCG02636) declare that the research entitled “**A NOVEL ALGORITHM FOR IMPROVED TRACTABILITY AND EXPLAINABILITY OF INFERENCES ON HIGH-DIMENSIONAL DATA USING CONDITIONED ATTENTION MECHANISM**” was carried out by me under the supervision of Dr. Iheanetu Olamma of the Department of Computer and Information Sciences, Covenant University, Ota, Nigeria. I attest that this dissertation has not been submitted either wholly or partially for the award of any degree elsewhere. All data and scholarly information used in this dissertation are duly acknowledged.

ASHUZA KUDERHA Rifin

Signature and Date

CERTIFICATION

This is to certify that this dissertation titled “**A NOVEL ALGORITHM FOR IMPROVED TRACTABILITY AND EXPLAINABILITY OF INFERENCES ON HIGH-DIMENSIONAL DATA USING CONDITIONED ATTENTION MECHANISM**” is original research carried out by **ASHUZA KUDERHA Rifin (22PCG02636)** in the Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Ogun State, Nigeria under the supervision of **Dr. Iheanetu Olamma**. We have examined and founded this work acceptable as part of the requirements for the award of Master of Science (M.Sc.) in Computer Science.

Dr. Iheanetu Olamma
(Supervisor)

Signature and Date

Prof. Olufunke Oladipupo
(Head of Department)

Signature and Date

Prof. Afolashade O. Kuyoro
(External Examiner)

Signature and Date

Prof. Akan B. Williams
(Dean, School of Postgraduate Studies)

Signature and Date

DEDICATION

I dedicate this work to all the healthcare workers who dedicated their lives to save the lives others. May this work help to measure their contributions.

ACKNOWLEDGEMENTS

I would like to express my gratitude to God who made this journey in Nigeria possible. To my family, I say thank you for the support.

To all my lecturers, I say thank you for the training. Especially, I thank my Supervisor, Dr. Iheanetu Olamma, for her guidance and support.

I am also grateful for the support from CApIC-ACE to help me pursue my dreams. I appreciate the work of Prof. Emeka E.J.Iweala, Prof. E. Adetiba, Prof. Ezekiel Adebisi, Prof. Olubanke Ogulana and Ms. Helen Jevwegaga.

To all my fellows, I say thank you for the support during this Master's Program.

TABLE OF CONTENTS

| CONTENTS | PAGES |
|---|--------------|
| TITLE PAGE | I |
| COVER PAGE | II |
| ACCEPTANCE | III |
| DECLARATION | IV |
| CERTIFICATION | V |
| DEDICATION | VI |
| ACKNOWLEDGEMENTS | VII |
| TABLE OF CONTENTS | VIII |
| LIST OF FIGURES | XII |
| LIST OF TABLES | XIII |
| ABSTRACT | XIV |
| | |
| CHAPTER ONE: INTRODUCTION | 1 |
| 1.1 Background to the Study | 1 |
| 1.2 Statement of the Research Problem | 5 |
| 1.3 Research Questions | 7 |
| 1.4 Aim and Objectives of the Study | 8 |
| 1.5 Research Hypotheses | 8 |
| 1.6 Justification for the study | 9 |
| 1.7 Scope of the study | 9 |
| 1.8 Definition of Terms | 9 |
| | |
| CHAPTER TWO: LITERATURE REVIEW | 10 |
| 2.1 Preamble | 10 |
| 2.2 Conceptual Review | 10 |
| 2.2.1 Data Representation Format | 10 |
| 2.2.2 Knowledge Representation | 12 |
| 2.2.3 Inference | 14 |
| 2.3 Theoretical Review | 15 |
| 2.3.1 Knowledge Representation and Reasoning | 15 |
| 2.3.2 Decision Rule Learning | 15 |
| 2.3.4 Attention Mechanism | 16 |

| | | |
|-----------------------------------|--|-----------|
| 2.4 | Empirical Review | 18 |
| 2.4.1 | Existing Attention Mechanisms | 18 |
| 2.4.2 | Existing Methods for Decision Rule Learning under Predictive Data Mining | 25 |
| 2.5 | Challenges with the Existing Methods | 34 |
| 2.6 | Gaps in Literature | 35 |
| CHAPTER THREE: METHODOLOGY | | 36 |
| 3.1 | Preamble | 36 |
| 3.2 | Methods and Methodology | 37 |
| 3.2.1 | Data Preprocessing | 37 |
| 3.2.2 | Learning of the Representation | 38 |
| 3.2.3 | Extraction of Decision Rules from the Learned Representation | 42 |
| 3.3 | Algorithms | 49 |
| 3.3.1 | Pseudocode | 49 |
| 3.4 | Complexity Analysis | 52 |
| 3.4.1 | Learning of the Representation | 53 |
| 3.4.2 | Extraction of Decision Rules | 53 |
| 3.5 | Parallelization of the Computation | 53 |
| 3.6 | Benchmark | 54 |
| 3.6.1 | Rule Extraction and Classification Performance | 54 |
| 3.6.2 | Configuration | 55 |
| 3.6.3 | Performance Evaluation Metrics | 55 |
| 3.6.4 | Signification Analysis of the Results | 57 |
| 3.6.5 | Evaluation of Execution Time and Memory Footprint | 57 |
| 3.7 | Implementation | 58 |
| 3.7.1 | Data Structure | 58 |
| 3.7.2 | Modularity | 59 |
| 3.7.3 | Object Oriented Programming | 59 |
| 3.7.3 | Visualization of Attribute-value Pairs' importance | 60 |
| 3.7.4 | Visualization of Decision Rules | 60 |

| | | |
|---|--|-----------|
| 3.7.5 | Inference for New Samples | 60 |
| 3.7.6 | Integration of the Datasets | 61 |
| 3.7.7 | Deployment of the Package | 61 |
| 3.7.8 | Extensive Validation of the Approach on High-dimensional RNA-Sequencing Data | 61 |
| 3.8 | Learning the Representation of Gene Expressions and Different Cancer Types | 62 |
| 3.8.1 | Visualization of Cancer-gene Expression Relationships | 62 |
| 3.8.2 | Extraction of Decision Rules | 62 |
| 3.8.3 | Benchmarking of our Approach on RNA-Seq Gene Expression Classification | 62 |
| 3.9 | Ablation Study | 64 |
| 3.9.1 | Ablation of the Column-wise and/or Row-wise Normalization Step | 64 |
| 3.9.2 | Compression of the Embedding Vectors | 65 |
| CHAPTER FOUR: RESULTS AND DISCUSSION | | 66 |
| 4.1 | Preamble | 66 |
| 4.2 | Presentation of Results and Analysis | 66 |
| 4.2.1 | Results of the Benchmarking | 66 |
| 4.2.2 | Time Complexity | 74 |
| 4.2.3 | Results of RNA-Seq Analysis for Cancer Classification | 75 |
| 4.2.4 | Hyperparameter Tuning | 75 |
| 4.2.5 | Performance on the Classification Task | 77 |
| 4.2.6 | Results of the Ablation Study | 78 |
| 4.3 | Discussion of Results | 80 |
| 4.3.1 | The Learning of the Representation | 80 |
| 4.3.2 | Representation Power of the Approach | 80 |
| 4.3.3 | Generalization Capacity of the Models | 81 |
| 4.3.4 | Robustness of the Predictions | 81 |
| 4.3.5 | Robustness of the Architecture | 81 |
| 4.3.6 | Explainability Capacity | 81 |
| 4.4 | Implications of Findings (Theoretical and Empirical Findings) | 82 |

| | | |
|---------------------------------|---|------------|
| 4.4.1 | Development of a Novel Attention Mechanism for Tabular Data | 82 |
| 4.4.2 | Integration of the Proposed Novel Attention Mechanism in the Extraction of Decision Rules | 82 |
| 4.4.3 | Efficient Learning of Decision Rules and Inference Processes on High-dimensional Data | 82 |
| 4.4.4 | Development of a Visualization Tool that Assess the Contribution of Different Features on a Decision selected by the AI algorithm | 82 |
| 4.4.5 | Formalization of Decision Rule Learning as Energy-based Modeling | 83 |
| 4.4.6 | Reduction of the Gap between Explainability and Performance | 83 |
| CHAPTER FIVE: CONCLUSION | | 84 |
| 5.1 | Preamble | 84 |
| 5.2 | Summary | 84 |
| 5.3 | Conclusion | 84 |
| 5.4 | Contribution(s) to Knowledge | 85 |
| 5.5 | Limitations of the Study | 85 |
| 5.6 | Areas for Further Research | 86 |
| REFERENCES | | 87 |
| APPENDICES | | 100 |

LIST OF FIGURES

| FIGURES | TITLE OF FIGURES | PAGES |
|----------------|---|--------------|
| 3. 1: | Framework of the proposed CAHDRL approach | 37 |
| 3. 2: | Summarized version of the CAHDRL | 38 |
| 3. 3: | Model optimization 1 | 46 |
| 3. 4: | Model optimization 2 | 47 |
| 4. 1: | Comparison of the standard deviation of inference with the for different models (part 1) | 73 |
| 4. 2: | Comparison of the standard deviation of inference with the for different models (part 2) | 74 |
| 4. 3: | Significance analysis of the results with Wilcoxon Two-tailed Signed-Rank Test | 75 |
| 4. 4: | Analysis of the execution time | 76 |
| 4. 5: | Ablative analysis of the normalization component | 79 |
| 4. 6: | Ablative analysis of the dimension compression. | 80 |

LIST OF TABLES

| TABLE | TITLE OF TABLES | PAGES |
|--------------|---|--------------|
| 2. 1: | Attention techniques | 18 |
| 2. 2: | Decision rule learning methods | 27 |
| 3. 1: | Presentation of the first 8 datasets | 55 |
| 3. 2: | Presentation of the RNA-Seq datasets | 63 |
| 3. 3: | Existing methods for cancer types classification | 64 |
| 4. 1: | Evaluation results for support-count for the 8 datasets (Part 1) | 68 |
| 4. 2: | Evaluation results for support-count for the 8 datasets (Part 2) | 69 |
| 4. 3: | Evaluation results of the length of the rules for the 8 datasets (Part 1) | 70 |
| 4. 4: | Evaluation results of the length of the rules for the 8 datasets (Part 2) | 70 |
| 4. 5: | Evaluation results for the accuracy on a classification task (Part 1) | 72 |
| 4. 6: | Evaluation results for the accuracy on a classification task (Part 2) | 72 |
| 4. 7: | Hyperparameters for the RNA-Seq datasets | 76 |
| 4. 8: | Quality of decision rules for the dataset RNA-Seq-2086 | 78 |
| 4. 9: | Quality of decision rules for the dataset RNA-Seq-801 | 78 |
| 4. 10: | Performance for classification task on the dataset RNA-Seq-2086 | 78 |
| 4. 11: | Performance for classification task on the dataset RNA-Seq-801 | 79 |

ABSTRACT

The availability of a large amount of data that can be mined is a challenge for data mining techniques that cannot keep up with the exponentially increasing quantity of data. Most of the classical techniques have been reported inefficient to handle large amount of data. The aim of this work is to develop a novel attention mechanism for high-dimensional tabular data understanding for knowledge representation that can enhance tractability of decision rule learning. The conditioning theory was mathematically formalized as an attention mechanism that models an unconditioned response (or reflex response) to unconditioned stimulus by pairing conditioned stimulus with unconditioned stimulus. The formalization of the conditioning theory built was translated into an algorithm presented as conditioned attention mechanism and applied on input tabular dataset producing vector embeddings of attribute-value pairs. The learned representation of attribute-value pairs as vector embeddings are then exploited to build decision rules using a developed novel Coupling-Sorting-Branching technique (CoSoBra), formalized as energy-based model optimization. A Python-based compression algorithm was developed to reduce the dimension of the resulting very large vector embeddings while maintaining the quality of the attribute-value pairs representations. The developed novel Conditioned Attention High Dimensional Rule Learning (CAHDRL) algorithm, which captures both CoSoBra algorithm and Condition Attention mechanism was benchmarked against existing approaches for decision rule learning on ten datasets from six different domains. The efficacy of the CAHDRL approach was evaluated with respect to the effectiveness of the extracted decision rules and its efficiency during data processing. An ablation study was carried out to assess the efficiency CAHDRL. Results of the evaluation on the 8 first datasets showed that CAHDRL had higher accuracy of 0.85 in average against 0.8175 in average for the EAV-60 model, equal length of decision rule of 2.125 in average against 2.125 in average for the reduct-alpha-0.1 model and a better support of 392.5263 in average against 350.2663 in average for the reduct-alpha-0.1. t-distributed Stochastic Neighborhood Embedding (t-SNE) was used to visualize the embeddings on a 2D manifold. The results support literature findings on Mushroom dataset that “odor=creosote” pair is a good descriptor of the poisonous mushroom. The second visualization results using “Adult” dataset showed that USA natives are more likely to earn \$50,000.00 and above. Wilcoxon test of significance of efficacy results for CAHDRL validated its relatively superior performance against existing approaches, yielding values of p-values inferior to 0.05 for accuracy (3 of the seven comparisons), length of decision rules (6 of the seven comparisons) and support (4 of the five comparisons) respectively. CAHDRL can effectively find embedding vectors of attribute-value pairs and efficiently extract decision rules from diverse datasets for classification tasks, with a higher tractability and inherent explainability. CAHDRL also demonstrates a great scalability feature that allows for the efficient processing very high dimensional datasets by the parallelizable nature of CAHDRL.

Keywords: attention mechanism, decision rule, explainability, tractability, machine learning