

CONTENTS

| | |
|---|-----|
| Acknowledgement | v |
| List of Figures | vii |
| List of Tables | x |
| Abstract | xii |
| 1. INTRODUCTION | 1 |
| 1.1. Data Mining and Bioinformatics | 1 |
| 1.2. Genomics | 6 |
| 1.3. Genetic Disorders and Muscular Dystrophy | 17 |
| 1.4. Review of Literature | 24 |
| 1.5. Objectives of the Research | 32 |
| 1.6. Organization of the Thesis | 33 |
| 2. MACHINE LEARNING | 35 |
| 2.1. Supervised Classification Algorithms | 38 |
| 2.1.1. Decision Tree Classifier | 38 |
| 2.1.2. Naïve Bayes Classifier | 42 |
| 2.1.3. Artificial Neural Network | 44 |
| 2.1.4. Support Vector Machine | 47 |
| 2.2. Ensemble Learning | 57 |
| 2.3. Training and Testing | 65 |
| 2.4. Summary | 70 |
| 3. DEEP LEARNING | 71 |
| 3.1. Feed Forward Neural Networks | 72 |

| | |
|--|------------|
| 3.2. Deep Neural Network Architecture | 74 |
| 3.3. Deep Neural Network Classifier | 76 |
| 3.4. Auto Encoder | 80 |
| 3.5. Stacked Auto Encoder | 82 |
| 3.6. Summary | 87 |
| 4. PROBLEM MODELING | 88 |
| 4.1. Data Acquisition through Positional Cloning | 88 |
| 4.2. Corpus Development | 89 |
| 4.3. Framework of Disease Identification Model | 94 |
| 4.3.1. Muscular Dystrophy Disease Identification Model through Shallow Learning Approach | 94 |
| 4.3.2. Muscular Dystrophy Disease Identification Model through Deep Learning Approach | 98 |
| 4.4. Summary | 99 |
| 5. MUSCULAR DYSTROPHY DISEASE IDENTIFICATION MODELS THROUGH SUPERVISED CLASSIFICATION TECHNIQUES | 100 |
| 5.1. Tools and Libraries | 101 |
| 5.2. Model I: Predicting Muscular Dystrophy Disease using Features related to Missense and Non sense (Non-Synonymous) Mutations | 103 |
| 5.3. Model II: Predicting Muscular dystrophy disease using Features related to Silent (Synonymous) Mutations | 115 |
| 5.4. Model III: Predicting Muscular Dystrophy Disease using Features related to Insertion and Deletion Mutations | 124 |

| | |
|---|------------|
| 5.5. Model IV: Predicting Muscular Dystrophy Disease using Features related to Splicing Mutations | 134 |
| 5.6. Model V: Predicting Muscular Dystrophy using Features related to Aggregated Mutational Descriptors | 143 |
| 5.7. Summary | 151 |
| 6. MUSCULAR DYSTROPHY DISEASE IDENTIFICATION MODELS THROUGH ENSEMBLE LEARNING | 153 |
| 6.1. Building Disease Identification Models through LibD3C Classifier | 153 |
| 6.2. Performance Evaluation Comparison with Supervised Learning Algorithms | 161 |
| 6.3. Summary | 163 |
| 7. MUSCULAR DYSTROPHY DISEASE IDENTIFICATION THROUGH DEEP LEARNING APPROACH | 164 |
| 7.1. TensorFlow | 165 |
| 7.2. Disease Identification Model using Deep Learning with Nucleotide Mapping Scheme | 166 |
| 7.3. Disease Identification Model using Deep Learning with Codon Mapping Scheme | 171 |
| 7.4. Comparison of Models based on ANN and DNN | 177 |
| 7.5. Summary | 181 |
| 8. CONCLUSION | 182 |
| REFERENCES | 185 |
| LIST OF PUBLICATIONS | 196 |
| Appendix | |
| A. Sample Gene Sequences | |

B. Feature Extraction and Feature Vectors

C. Python Code