### TABLE OF CONTENTS

|  |  |
| --- | --- |
| ABSTRACT | **iv** |
| **ACKNOWLEDGMENTS** | **v** |
| **TABLE OF CONTENTS** | **vi** |
| **LIST OF FIGURES** | **viii** |
| **LIST OF TABLES** | **ix** |

|  |  |  |
| --- | --- | --- |
| Chapter 1: Introduction  **[Font Size 13]** | | **1** |
|  | 1.1 Overview …………………………………………………………..... | 1 |
|  | 1.2 Problem Definition …………………………………………………. | 2 |
|  | 1.3 Planted (l,d) Motif …………………………………………………... | 2 |
|  | 1.4 Motivation for Finding Motif………………………………………... | 4 |
|  | 1.5 Biological Background………………………………………………. | 4 |
|  | 1.5.1 DNA………………………………………………………….. | 4 |
|  | 1.5.2 Sequence Feature: Gene………………………………………. | 6 |
|  | 1.5.2 Sequence Feature: Regulatory Sites……………………………. | 6 |
|  | 1.5.2 Sequence Feature: Repetitive Elements………………………... | 8 |
|  | 1.6 Biosequence Analysis………………………………………………... | 8 |
|  | 1.6.1 Motivation…………………………………………………….. | 8 |
|  | 1.6.2 Sequence Evolution…………………………………………… | 10 |
|  | 1.6.3 Sequence Homology…………………………………………... | 10 |
|  | 1.6.4 Motif………………………………………………………….. | 12 |
|  | 1.6.5 Variation in Motif……………………………………………... | 13 |
|  | 1.7 Overview of the thesis……………………………………………….. | 13 |

|  |  |  |
| --- | --- | --- |
| Chapter 2: Literature Review | | **14** |
|  | 2.1 Overview……………………………………………………………. | 14 |
|  | 2.2 Representation of Motif……………………………………………... | 14 |
|  | 2.2.1 Consensus Sequence………………………………………… | 15 |
|  | 2.2.2 Weight Matrix Model-WMM………………………………… | 16 |
|  | 2.3 Approaches To Solve the Problem…………………………………... | 18 |
|  | 2.3.1 Enumerative Algorithm……………………………………... | 18 |
|  | 2.3.2 Local Search Based Algorithm………………………………. | 19 |
|  | 2.4 Conclusion…………………………………………………………... | 24 |
|  |  |  |
|  |  |  |
| **Chapter 3: Modifying SPSTAR for Faster Response** | | **26** |
|  | 3.1 Overview……………………………………………………………. | 26 |
|  | 3.2 The SPSTAR Algorithm……………………………………………... | 26 |
|  | 3.3 Implementation Details of SPSTAR Algorithm……………………… | 27 |
|  | 3.4 Modified SPSTAR Implementation………………………………….. | 28 |
|  | 3.5 Performance Comparison of SPSTAR and Modified SPSTAR………. | 31 |
|  | 3.6 Conclusion…………………………………………………………... | 32 |

|  |  |  |
| --- | --- | --- |
| **Chapter 4: Finding Motif Using Suffix Tree** | | **33** |
|  | 4.1 Overview……………………………………………………………. | 33 |
|  | 4.2 Definition of Suffix Tree and Basic Properties………………………. | 33 |
|  | 4.3 Algorithm for Constructing Suffix Tree and Its Complexities………... | 34 |
|  | 4.4 Implementation Details of Suffix Tree………………………………. | 36 |
|  | 4.5 Motif Finding Using Suffix Tree…………………………………….. | 37 |
|  | 4.6 Conclusion…………………………………………………………... | 41 |

|  |  |  |
| --- | --- | --- |
| **Chapter 5: Performance Measurement of The Proposed Algorithm** | | **42** |
|  | 5.1 Overview…………………………………………………………….. | 42 |
|  | 5.2 Defining the Performance…………………………………………… | 42 |
|  | 5.3 Factors Effecting Performance of a Motif Finder……………………. | 43 |
|  | 5.4 Test Data……………………………………………………………. | 43 |
|  | 5.5 Performance Comparison of SPSTAR and Proposed Method Using  Suffix Tree…………………………………………………………... | 43 |
|  | 5.6 Conclusion…………………………………………………………... | 46 |

|  |  |  |
| --- | --- | --- |
| **Chapter 6: Conclusion** | | **47** |
|  | 6.1 Discussion…………………………………………………………… | 47 |
|  | 6.2 Future Work…………….…………………………………………… | 48 |

|  |  |
| --- | --- |
| **APPENDIX A** | **49** |
| **APPENDIX B** | **55** |
| **APPENDIX C** | **63** |
| **BIBLIOGRAPHY** | **67** |