TABLE OF CONTENTS

LIST OF SYMBOLS XV

LIST OF ABBREVIATIONS XVI

LIST OF KEYWORDS XVII

LIST OF FIGURES XVIII

LIST OF TABLES XX

PREFACE XXII

CHAPTER 1: INTRODUCTION 1

1.1. INTRODUCTION 1

1.2. PREDICTION OF PROTEIN FUNCTIONS AND ITS IMPORTANCE 4

1.3. PROBLEM DESCRIPTION AND MOTIVATION OF THE WORK 4

1.3.1. GENERAL FRAMEWORK OF THE COMPUTATIONAL INTELLIGENCE TECHNIQUES USED FOR PROTEIN FUNCTION PREDICTION 6

1.4. OBJECTIVE AND CONTRIBUTIONS TO THE THESIS 7

1.5. ORGANIZATION OF THE THESIS 8

CHAPTER 2: THEORETICAL BACKGROUND 11

2.1. LITERATURE REVIEW 11

2.1.1. COMPUTATIONAL INTELLIGENCE TECHNIQUES IN THE PREDICTION OF ION CHANNELS AND THEIR TYPES 12

2.1.2. COMPUTATIONAL INTELLIGENCE TECHNIQUES IN THE PREDICTION OF ENZYME FUNCTIONS 13

2.1.3. COMPUTATIONAL INTELLIGENCE TECHNIQUES IN PREDICTION OF NUCLEAR RECEPTORS AND THEIR SUBFAMILIES 17

2.1.4. Computational intelligence techniques in prediction of G-protein coupled receptors and their subfamilies 19
2.2. Features extraction of protein sequences 21
2.2.1. Amino acid composition (AAC) 23
2.2.2. Dipeptide composition (DC) 23
2.2.3. Correlation factors (CF) 24
2.2.4. Composition, transition and distribution features (CTD) 25
2.2.5. Sequence order descriptors 25
2.2.6. Pseudo amino acid composition (PAAC) 26
2.3. Feature selection techniques 27
2.3.1. Filter method 27
2.3.2. Wrapper method 28
2.3.3. Hybrid method 28
2.4. Computational intelligence techniques 28
2.4.1. Artificial neural network (ANN) 28
2.4.2. Naive Bayes classifier 29
2.4.3. Support vector machine (SVM) 29
2.4.4. k-nearest neighbor (k-NN) 30
2.4.5. Decision trees 30
2.4.6. Bagging 31
2.4.7. Boosting 31
2.4.8. Random subspace method 31
2.4.9. Random forests 32
2.5. Performance measures 32
2.6. Conclusion 34

CHAPTER 3: A MULTI STAGE APPROACH FOR THE PREDICTION OF ION CHANNELS AND THEIR SUBFAMILIES 35

3.1. Background 36
3.1.1. Material and methods 36
3.1.2. Features extraction of protein sequences 38
3.2. Proposed method and model 38
3.2.1. Feature subset selection 39

3.2.2. CLASSIFICATION OF ION CHANNELS AND THEIR TYPES 40
3.3. RESULTS AND PERFORMANCE ANALYSIS 42
3.3.1. PERFORMANCE MEASURES 42
3.3.2. RESULTS AND ANALYSIS 43
3.3.2.1. Prediction of ion channels and non-ion channels 45
3.3.2.2. Prediction of voltage and ligand gated ion channels 47
3.3.2.3. Prediction of subfamilies of voltage gated ion channels 49
3.3.2.4. Prediction of subfamilies of ligand gated ion channels 52
3.3.2.5. Prediction of subfamilies of voltage gated calcium ion channels 54
3.3.2.6. Prediction of subfamilies of voltage gated potassium ion channels 57
3.3.2.7. Prediction of subfamilies of voltage gated sodium ion channels 60
3.3.2.8. Prediction of subfamilies of voltage gated chloride ion channels 63
3.4. COMPARATIVE ANALYSIS 66
3.5. CONCLUSION 68

CHAPTER 4: A TOP DOWN APPROACH FOR THE PREDICTION OF ENZYME FUNCTIONAL CLASSES AND SUBCLASSES 69

4.1. BACKGROUND 70
4.1.1. MATERIAL AND METHODS 70
4.1.2. FEATURES EXTRACTION OF PROTEIN SEQUENCES 70
4.2. PROPOSED METHOD AND MODEL 72
4.2.1. PREDICTION OF ENZYME FUNCTIONAL CLASSES AND SUBCLASSES 72
4.2.2. PSEUDO CODE FOR THE ROTATION RANDOM FOREST 73
4.3. RESULTS AND PERFORMANCE ANALYSIS 77
4.3.1. PERFORMANCE MEASURES 77
4.3.2. RESULTS AND COMPARATIVE ANALYSIS 77
4.3.2.1. Prediction of enzymes and non-enzymes 77
4.3.2.2. Prediction of enzymes functional classes 78
4.3.2.3. Prediction of functional subclasses of enzymes 80
4.4. CONCLUSION 82

CHAPTER 5: AN EFFICIENT APPROACH FOR PREDICTION OF NUCLEAR RECEPTORS AND THEIR SUBFAMILIES 83

5.1. BACKGROUND 84

5.1.1. MATERIAL AND METHODS: 84

5.1.2. FEATURES EXTRACTION OF PROTEIN SEQUENCES 85

5.2. PROPOSED METHOD AND MODEL 85

5.2.1. FEATURE SUBSET SELECTION 86

5.2.2. CLASSIFICATION OF NUCLEAR RECEPTORS AND THEIR SUBFAMILIES 87

5.3. RESULT AND PERFORMANCE ANALYSIS 89

5.3.1. PERFORMANCE MEASURES 89

5.3.2. RESULT ANALYSIS 89

5.3.3. COMPARATIVE ANALYSIS 92

5.4. CONCLUSION 94

CHAPTER 6: AN EFFICIENT AND ROBUST APPROACH FOR THE PREDICTION OF G-PROTEIN COUPLED RECEPTORS AND THEIR SUBFAMILIES 95

6.1. BACKGROUND 96

6.1.1. MATERIAL AND METHODS 97

6.1.2. FEATURES EXTRACTION OF PROTEIN SEQUENCES 98

6.2. PROPOSED METHOD AND MODEL 98

6.2.1. FEATURE SUBSET SELECTION 99

6.2.2. CLASSIFICATION OF G-PROTEIN COUPLED RECEPTORS AND THEIR SUBFAMILIES 103

6.3. RESULT AND PERFORMANCE ANALYSIS 104

6.3.1. PERFORMANCE MEASURES 106

6.3.2. RESULTS AND COMPARATIVE ANALYSIS 106

6.3.2.1. Prediction of GPCRs and non-GPCRs 106

6.3.2.2. Prediction of subfamilies of G-protein coupled receptors 109

6.3.2.3. Prediction of subfamilies of class A G-protein coupled receptors 111

6.4. CONCLUSION 114

CHAPTER 7: CONCLUSION AND SCOPE FOR FUTURE WORK 115

7.1. CONCLUDING REMARKS 115

7.2. SCOPE FOR FUTURE WORKS 119

REFERENCES.....120

LIST OF PAPERS PUBLISHED/PRESENTED/COMMUNICATED 132

COPIES OF MANUSCRIPTS/ REPRINTS OF THE PAPERS COMMUNICATED/ ACCEPTED/ PUBLISHED.....134

PERSONAL PROFILE