UNIVERSITAS
GADJAH MADA
GA-based Feature Selection for Protein Secondary Structure Prediction
FATHURROCHMAN H., Afiahayati, S.Kom., M.Cs, Ph.D
Universitas Gadjah Mada, 2021 | Diunduh dari http://etd.repository.ugm.ac.id/

TABLE OF CONTENTS

ACKNOWLEDGMENT i
TABLE OF CONTENTS ii
LIST OF FIGURES v
LIST OF TABLES vii
INTISARI viii
ABSTRACT ix
CHAPTER I INTRODUCTION 1
1.1. Background 1
1.2. Problem Formulation 3
1.3. Objectives 3
1.4. Scope and Limitation 3
1.5. Contribution 4
1.6. Research Outline 4
CHAPTER II RELATED WORK 6
CHAPTER III THEORETICAL BASIS 11
3.1. Protein 11
3.2. Protein Structure 11
3.3. Protein Structure Prediction 12
3.4. Protein Features 14
3.4.1. Amino Acid Sequences Features 14
3.4.2. Position-Specific Scoring Matrix (PSSM) 15
3.4.3. Physical Features 15
3.4.4. Conformation Parameters 15
3.5. Deep Learning Network 16
3.5.1. Convolutional Neural Network (CNN) 16

UNIVERSITAS
GADJAH MADA
GA-based Feature Selection for Protein Secondary Structure Prediction
FATHURROCHMAN H., Afiahayati, S.Kom., M.Cs, Ph.D
Universitas Gadjah Mada, 2021 | Diunduh dari http://etd.repository.ugm.ac.id/

3.5.2. Long Short Term Memory (LSTM) 17
3.5.3. Bi-directional LSTM (BLSTM) 19
3.5.4. Attention Model 19
3.6. Genetic Algorithm 20
CHAPTER IV PROPOSED METHOD 22
4.1. System Overview 22
4.2. Dataset Analysis 23
4.3. System Design 29
4.4. GA-based Feature Selection 31
4.4.1. Chromosome Representation 32
4.4.2. Fitness Function 32
4.4.3. Tournament Selection 34
4.4.4. Crossover and Mutation 35
4.5. Architecture Design 37
4.5.1. CNN Architecture 37
4.5.2. CNN-BLSTM Architecture 38
4.5.3. CNN-BLSTM with Attentions Layer 39
4.5.4. CNN-BLSTM with Highway Connections 39
4.6. Training Design 40
4.7. Evaluation Design 42
CHAPTER V IMPLEMENTATION 44
5.1. Implementation Environment 44
5.2. Data Preprocessing 45
5.3. Architecture Implementation 48
5.4. CNN Architecture 49
5.4.1. CNN-BLSTM Architecture 49
5.4.2. CNN-BLSTM with Attentions 50

UNIVERSITAS
GADJAH MADA
GA-based Feature Selection for Protein Secondary Structure Prediction
FATHURROCHMAN H., Afiahayati, S.Kom., M.Cs, Ph.D
Universitas Gadjah Mada, 2021 | Diunduh dari http://etd.repository.ugm.ac.id/

5.4.3. CNN-BLSTM with Highway 51
5.4.4. Genetic Algorithm Implementation 52
5.5. Evaluation Implementation 53
CHAPTER VI EXPERIMENTAL RESULT 57
6.1. Dataset 57
6.2. Finding Optimal Feature Subset 58
6.2.1. Experimental Parameters 58
6.2.2. Single Objective GA 59
6.2.3. Multiobjective GA 60
6.2.4. Optimal Feature Subsets Result 61
6.3. Evaluation on Optimal Feature Subset 62
6.4. Comparison with previous studies 64
CHAPTER VII CONCLUSIONS AND FUTURE WORKS 66
7.1. Conclusions 66
7.2. Future Works 67
REFERENCES 68
APPENDIX 71
A1. Predicted Result on CASP10 T0711 71
A2. Predicted Result on CASP10 T0651-D3 72
A3. Predicted Result on CASP10 T0716 73
A4. Predicted Result on CASP10 T0726-D2 74
A5. Predicted Result on CASP10 T0685-D2 75