



LIST OF CONTENTS

APPROVAL PAGE	i
PLAGIARISM FREE STATEMENT PAGE	ii
PREFACE	iii
LIST OF CONTENTS	iv
NOMENCLATURE	viii
INTISARI	ix
ABSTRACT	1
CHAPTER I INTRODUCTION	1
1.1. Background	1
1.2. State the Art of Research	3
1.3. Problem Formulation	5
1.4. Objectives	5
1.5. Benefits	6
CHAPTER II RESEARCH STUDIES	7
2.1. Literature Studies	7
2.1.1. Architecture Development	7
2.1.2. Features Enrichment	8
2.1.3. Protein Language Modelling	8
2.2. Theoretical Studies	11
2.2.1. Protein	11
2.2.2. Protein Structure	12
2.2.3. Protein Structure Prediction	13
2.2.4. Protein Secondary Structure Prediction	14
2.2.5. Embedding	16
2.2.6. Protein Language Modelling	16
2.2.7. LIFT-SS Architecture	17
2.2.8. CNN	18
2.3. Research Hypotheses	18
CHAPTER III RESEARCH METHODS	20
3.1. Implementation Environment	20
3.2. System Overview	22



3.3. Dataset of the Experiments	22
3.3.1. Dataset Analysis	22
3.3.2. Data Processing	26
3.3.3. Data Processing Variation	27
3.4. Design of Experimentation	30
3.5. Architecture of Experimentation	30
3.5.1. Architectures Design	30
3.5.2. Architecture Implementation	32
3.6. Evaluation of the Experiment	33
3.6.1. PSSP Training Evaluation	33
3.6.2. Biomedical Use Case: COVID-19 Mutation	35
3.7. Research Timeline	36
CHAPTER IV RESULTS AND DISCUSSION	38
4.1. Results	38
4.1.1. LIFT-SS Adoptation Training Results	38
4.1.2. CNN for T5-based Embedding Training Results	46
4.1.3. Metrics Per Sequence Distribution on Sequence Length	48
4.1.4. Biomedical Use Case: COVID-19 Mutation	50
4.2. Discussion	51
4.2.1. Data Size Difference Sensitivity	51
4.2.2. "Split in Two" Masking Strategy	53
4.2.3. PLMs Comparison	54
4.2.4. PLM-based Model Comparison Training	55
4.2.5. PLM-based Model Architectures Different	56
CHAPTER V CONCLUSIONS	58
BIBLIOGRAPHY	59
APPENDICES	65
A.1 Loss calculation (Yang, Liu, and Li (2023))	66
A.2 Data Load and Pre-processing modified from Yang, Liu, and Li (2023)	69
A.3 Utilization modified from Yang, Liu, and Li (2023)	79
A.4 LIFT-SS training code modified from Yang, Liu, and Li (2023)	103
A.5 Network of LIFT-SS T5-based model modified from Yang, Liu, and Li (2023)	116
A.6 Function to set modified LIFT-SS network of ESM-based model modified from Yang, Liu, and Li (2023)	124
A.7 Function to set modified LIFT-SS network of Ankh-based model modified from Yang, Liu, and Li (2023)	132



A.8 Function to set CNN T5-based from Yang, Liu, and Li (2023) and El-	
naggar et al (2020)	137