

Contents

1	Word Embedding Methods	1
1.1	Word2Vec	2
1.1.1	Continuous Bag of Words (CBoW).....	3
1.1.2	Skipgram	7
1.1.3	Word2Vec with Negative Sampling.....	9
1.2	FastText	12
1.3	Global Vector (GloVe).....	14
1.4	Basic Concepts of Graph Theory	17
1.5	DeepWalk	19
1.6	Node2Vec	21
1.7	Struc2Vec	25
1.8	Large Scale Information Network Embedding	32
1.9	GraRep: Learning Graph Representation with Global Structural Information	36
1.10	HOPE: High Order Proximity Preserved Embedding.....	42
1.11	Embeddings from Language Models (ELMo)	50
1.11.1	Character-Level Word Embedding Generation.....	51
1.11.2	Pre-training of Bidirectional LSTM	53
1.12	A High Level Overview of Remaining Word Embedding Methods.	55
1.12.1	Deep Learning Based Methods	55
1.12.2	Knowledge Graph Distance Based Scoring Methods	56
1.12.3	Walk Based Methods	57
1.12.4	MetaPath Based Methods	58
1.12.5	Mathematical Methods	59
1.12.6	Entire Graph as One Embedding Methods	60
2	Large Language Models	61
2.1	ULMFiT Language Model	63
2.2	Transformer: A Self-Attention Network.....	67
2.2.1	Input: Token and Position Embeddings.....	70
2.2.2	Causal or Backward-Looking Self-Attention	72

2.2.3	Multi-Head Attention.....	75
2.2.4	Normalization Layer.....	77
2.2.5	Residual Connection	78
2.2.6	Feed-Forward Network.....	78
2.2.7	Transformer Block's Residual Stream Perspective	79
2.2.8	Language Modeling Head.....	83
2.3	BERT Language Model	86
2.3.1	Tokenization Layer	88
2.3.2	Input Embedding Layer	89
2.3.3	BERT Encoding Layer	90
2.3.4	Pre-training of BERT Language Model	91
2.3.5	Fine Tuning	92
2.4	AlphaFold	92
2.4.1	Input Module.....	94
2.4.2	Evoformer.....	96
2.4.3	Structure Module	103
2.5	OmegaFold.....	106
2.6	RNAFormer	111
2.7	A High Level Overview of 18 Distinct Language Models.....	113
3	AI-Driven Insights into DNA Sequence Analysis Landscape	123
3.1	Research Methodology	124
3.1.1	Article Searching	124
3.1.2	Article Screening and Filtering	126
3.2	DNA Sequence Analysis Benchmark Datasets.....	126
3.3	A Brief Look on Word Embeddings and Language Models Predictive Pipelines.....	135
3.4	Open-Source DNA Sequence Analysis Predictive Models	136
3.5	Predictive Pipelines Performance Analysis.....	140
3.6	Publisher and Journal-Wise Distribution of Research Articles	174
3.7	Discussion.....	176
4	AI-Driven Insights into RNA Sequence Analysis Landscape	179
4.1	Research Methodology	180
4.1.1	Article Searching	180
4.1.2	Article Screening and Filtering	182
4.2	RNA Sequence Analysis Benchmark Datasets.....	182
4.3	A Brief Look on Word Embeddings and Language Models Predictive Pipelines.....	190
4.4	Open-Source RNA Sequence Analysis Predictive Pipelines	191
4.5	RNA Sequence Analysis Predictive Pipelines Performance Analysis	200
4.6	Publisher and Journal-Wise Distribution of Research Articles	239
4.7	Discussion.....	241

5	AI-Driven Insights into Protein Sequence Analysis Landscape.....	245
5.1	Research Methodology	246
5.1.1	Article Searching	247
5.1.2	Article Screening and Filtering	247
5.2	Word Embeddings Based AI-Driven Protein Sequence Analysis Predictive Pipelines	248
5.3	Language Models Based AI-Driven Protein Sequence Analysis Predictive Pipelines	251
5.4	Protein Sequence Analysis Benchmark Datasets.....	253
5.5	Open-Source Protein Sequence Analysis Predictive Models	268
5.6	Protein Sequence Analysis Predictive Pipelines Performance Analysis	285
5.7	Publisher and Journal-Wise Distribution of Research Articles	364
5.8	Discussion.....	369
6	AI-Driven Revolution in Peptide Classification Landscape	371
6.1	Biological Foundations and Classification of Peptides	373
6.2	A Look-Back into Existing Review Studies	378
6.3	Peptide Classification Using Experimental Techniques	383
6.4	A High Level Overview of AI Based Peptides Identification Predictive Pipelines	385
6.5	Research Methodology	387
6.5.1	Article Identification.....	387
6.5.2	Article Screening and Filtering	389
6.6	Peptide Classification Datasets.....	389
6.7	Representation Learning Techniques of Peptides Identification Predictors	393
6.8	Classification Methodologies of Peptides Identification Predictors .	397
6.9	Unmasking Peptides Identification Predictors Evaluation Strategies	399
6.10	State-of-the-Art Performance Values of Peptides Identification Predictors	400
6.11	Open Source Tools and Libraries Potential for Peptides Identification Pipelines	412
6.12	Peptides Classification Predictors Performance Analysis	419
6.12.1	Anti-fungal Peptide Classification: Predictive Pipelines Design and Performance Evaluation	419
6.12.2	Antimicrobial Peptide Classification: Predictive Pipelines Design and Performance Evaluation	423
6.12.3	Cell Penetrating Peptide Classification: Predictive Pipelines Design and Performance Evaluation	439
6.12.4	Bitter Peptide Classification: Predictive Pipelines Design and Performance Evaluation	446
6.12.5	AntiViral Peptides Classification: Predictive Pipelines Design and Performance Evaluation	450

6.12.6	Signal Peptides Classification: Predictive Pipelines Design and Performance Evaluation	456
6.12.7	Neuro Peptides Classification: Predictive Pipelines Design and Performance Evaluation	462
6.12.8	Anti-hypertensive Peptides Classification: Predictive Pipelines Design and Performance Evaluation	467
6.12.9	Anti-tubercular Peptides Classification: Predictive Pipelines Design and Performance Evaluation	471
6.12.10	Anti-bacterial Peptides Classification: Predictive Pipelines Design and Performance Evaluation	475
6.12.11	Blood Brain Barrier Peptides Classification: Predictive Pipelines Design and Performance Evaluation	477
6.12.12	Tumor-Homing Peptides Classification: Predictive Pipelines Design and Performance Evaluation	481
6.12.13	Anti-angiogenic Peptides Classification: Predictive Pipelines Design and Performance Evaluation	485
6.12.14	Dipeptidyl Peptidase-IV Peptide Classification: Predictive Pipelines Design and Performance Evaluation	488
6.12.15	Anti-cancer Peptides Classification: Predictive Pipelines Design and Performance Evaluation	491
6.12.16	Anti Inflammatory Peptides Classification: Predictive Pipelines Design and Performance Evaluation	502
6.12.17	Anti-MRSA, Insecticidal, Anti-sepsis, Biofilm Inhibitory, Anti-biofilm, and Hormone Peptide Classification: Predictive Pipelines Design and Performance Evaluation	507
6.12.18	Peptide Multi-label Classification: Predictive Pipelines Design and Performance Evaluation	512
6.13	Discussion	519
6.14	Publisher and Journal-Wise Distribution of Research Papers	520
References		523