

List of Figures:

Figure Number	Figure Title	Page Number
1.1	Central Dogma	10
1.2	Different Gene Expression Profile	12
1.3	Gene Expression Pattern	13
1.4	Working of the Proposed System	16
1.5	Proposed Ensemble Method	17
2.1	Central Dogma Detail	18
2.2	Genetic Information	18
2.3	Codon Part in mRNA	19
3.1	cDNA Microarray	26
3.2	cDNA Color Details	26
4.1	Cluster Validation	32
5.1	Connectivity Index (Yeast)	43
5.2	Dunn Index (Yeast)	44
5.3	Silhouette Index(Yeast)	44
5.4	APN(Yeast)	45
5.5	ADM(Yeast)	45
5.6	AD(Yeast)	46
5.7	FOM(Yeast)	46
5.8	Connectivity Index (Yeast sporulation)	51
5.9	Dunn Index (Yeast sporulation)	51
5.10	Silhouette Index(Yeast sporulation)	52
5.11	AD(Yeast sporulation)	52
5.12	ADM(Yeast sporulation)	53
5.13	APN(Yeast sporulation)	53
5.14	FOM(Yeast sporulation)	54
5.15	Connectivity Index (Rat-CNS)	59
5.16	Dunn Index (Rat-CNS)	59
5.17	Silhouette Index(Rat-CNS)	60
5.18	AD(Rat-CNS)	60
5.19	ADM(Rat-CNS)	61
5.20	APN(Rat-CNS)	61
5.21	FOM(Rat-CNS)	62
5.22	Connectivity Index (Homo Sapiens)	67
5.23	Dunn Index (Homo Sapiens)	67
5.24	Silhouette Index(Homo Sapiens)	68
5.25	AD(Homo Sapiens)	68
5.26	ADM(Homo Sapiens)	69
5.27	APN(Homo Sapiens)	69
5.28	FOM(Homo Sapiens)	70

List of Tables:

Table Number	Table Title	Page Number
1.1	Expression Level of Gene	11
2.1	Amino Acid	20
4.1	Validation Criteria	36
5.1	Data Sets	38
5.2	Internal Validation (Yeast)	39
5.3	Stability Validation (Yeast)	40
5.4	Optimal Score(Yeast)	42
5.5	Ensemble Result(Yeast)	43
5.6	Internal Validation (Yeast Sporulation)	47
5.7	Stability Validation (Yeast Sporulation)	48
5.8	Optimal Score(Yeast Sporulation)	50
5.9	Ensemble Result(Yeast Sporulation)	50
5.10	Internal Validation (Rat-CNS)	55
5.11	Stability Validation (Rat-CNS)	56
5.12	Optimal Score(Rat-CNS)	58
5.13	Ensemble Result(Rat-CNS)	63
5.14	Internal Validation (Homo Sapiens)	64
5.15	Stability Validation (Homo Sapiens)	66
5.16	Optimal Score(Homo Sapiens)	67
5.17	Ensemble Result(Homo Sapiens)	68

List of Abbreviations:

SA	Stability Analysis
IV	Internal Validation
APN	Average Proportion Non-Overlap
ADM	Average Distance between Mean
AD	Average Distance
FOM	Figure Of Merit
SI	Silhouette Index
DI	Dunn Index
CI	Connectivity Index
K-Means	K Medoid
PAM	Partitioning Around Medoid
Clara	Clustering Large Application
CNS	Central Nervous System
DNA	Deoxyribo Nucleic Acid
RNA	Ribo Nucleic Acid

Contents

Declaration	i
Acknowledgement	ii
Abstract	iii
List of Figures	iv
List of Tables & Abbreviation	v

1. Introduction

1.1 Biological Background	9
1.2 Genetic Background	10
1.3 Gene Expression	10
1.4 Measuring Gene Expression	11
1.5 Gene Expression Pattern	11
1.6 Characteristics of Gene Expression Data	12
1.7 Scale Transformation	12
1.8 Motivation	13
1.9 Research Issues	13
1.10 Existing Approaches	14
1.11 Data Mining	14
1.12 Clustering Technique	15
1.13 Analysis of Gene Expression Matrix	15
1.14 Objective and Aim of the Work	15
1.15 Assumption and Limitation	17
1.16 Organisation of the Dissertation	17

2. Central Dogma

2.1 Replication	20
2.2 Transcription	20
2.3 Translation	21
2.4 Codon Property	21
2.5 Transcription Factor Glossary	22
2.6 Discussion	23

3. Gene Coherent Pattern Identification

3.1 Microarray Technology	24
3.2 Gene Expression Data	24
3.3 Quantification of Gene Expression Data	25
3.3.1 cDNA Microarray	25
3.4 Proximity Measure	26
3.5 Normalization	27
3.6 Clustering Gene expression Data	27
3.7 Discussion	28

4. Gene Coherent Pattern Identification using Unsupervised

Approach

4.1 Different Clustering Techniques Used	29
4.1.1 K-Means	29
4.1.2 Hierarchical	30
4.1.3 PAM	30
4.1.4 SOM	31
4.1.5 CLARA	31
4.2 Validation of Clusters	32
4.2.1 Different Aspects of Cluster Validation	32
4.3 Classical Validation Approach	33
4.3.1 External Validation	33
4.3.2 Internal Validation	33
4.4 Related Works	33
4.4.1 Validity Analysis	33
4.4.2 Stability Measures	34
4.4.2.1 APN	35
4.4.2.2 AD	35
4.4.2.3 ADM	35
4.4.2.4 FOM	36
4.5 Criteria of Validation Measure	36
4.6 reliability of Clusters	36

5. Experimental Results and Validation

5.1 Environment used	37
5.2 Data set used and their description	38
5.3 Results	38-70

6. Conclusion and Future work

7. Bibliography