

List of Tables

2.1	Comparison with related work in gene feature selection	15
2.2	Comparison with related work in cancer classification	18
3.1	Training and testing sample size using gene expression	28
3.2	Training and testing samples size using miRNA expression	28
3.3	HCC and lung cancer active learning results	32
3.4	Breast cancer ER+/ER- and HER2+/HER2- active learning results	33
3.5	Dataset sizes for the related work in [1] and [2]	37
3.6	Comparison to [1] using 10-fold cross-validation	38
3.7	Comparison to [2] using 10-fold cross-validation	39
4.1	Training and testing samples size for breast cancer and HCC subtypes using miRNA expression	50
4.2	Training and testing sample size for breast cancer, HCC and lung cancer subtypes using gene expression	50
4.3	Sample size and gene/miRNA numbers of unlabeled sets	51
4.4	Precision, recall and F1-measure for breast cancer subtypes RF classifiers using miRNA expression and gene expression dataset	54
4.5	Precision, recall and F1-measure for breast cancer subtypes SVM and LDS classifiers using miRNA expression and gene expression dataset	55
4.6	Results of HCC RF subtype classifiers using gene/miRNA expression dataset	57
4.7	Results of HCC subtype SVM and LDS classifiers using miRNA expression dataset	58
4.8	The number of overlapping miRNAs and genes between initial datasets and added datasets in breast cancer	58
4.9	The number of overlapping miRNAs and genes between initial datasets and added datasets in HCC	59
4.10	Results of lung cancer RF subtypes classifiers using gene expression dataset	60

List of Figures

2.1	Genes-Expression Matrix	7
2.2	Deep Belief Nets Structure [3]	9
3.1	Multilevel Feature Selection Approach for Gene Expression Sets	22
3.2	miRNAs Feature Selection based on Multilevel Gene Feature Selection Approach	23
3.3	Comparison to Classical Feature Selection using Gene Expression Sets in HCC (Highest value is marked with a red circle) . . .	30
3.4	Comparison to Classical Feature Selection using Gene Expression Sets in Lung Cancer (Highest value is marked with a red circle)	30
3.5	Comparison to Classical Feature Selection using Gene Expression Sets in Breast Cancer ER+/ER- (Highest value is marked with a red circle)	31
3.6	Comparison to Classical Feature Selection using Gene Expression Sets in Breast Cancer HER2+/HER2- (Highest value is marked with a red circle)	31
3.7	Comparison to Classical Feature Selection using MiRNA Expression Sets in HCC part 1 (Highest value is marked with a red circle)	34
3.8	Comparison to Classical Feature Selection using MiRNA Expression Sets in HCC part 2 (Highest value is marked with a red circle)	35
3.9	Comparison to Classical Feature Selection using MiRNA Expression Sets in Breast Cancer ER+/ER- part 1 (Highest value is marked with a red circle)	35
3.10	Comparison to Classical Feature Selection using MiRNA Expression Sets in Breast Cancer ER+/ER- part 2 (Highest value is marked with a red circle)	36
3.11	Comparison to Classical Feature Selection using MiRNA Expression Sets in Breast Cancer HER2-/HER2+ part 1 (Highest value is marked with a red circle)	36
3.12	Comparison to Classical Feature Selection using MiRNA Expression Sets in Breast Cancer HER2-/HER2+ part 2 (Highest value is marked with a red circle)	37
4.1	Self-Learning Approach Overview	43
4.2	miRNAs and Their Target Genes are Related by a Many to Many Relationship. The first column represents miRNAs and the second column represents target genes ids.	46
4.3	Training data size comparison of initial RF classifier compared to adapted self-learning and co-training for breast cancer using miRNA expression sets. (Class 0 is ER+/Her2-, class 1 is ER-/Her2+, class 2 is ER-/Her2- and class3 is ER+/Her2+)	53

List of Acronyms

ALL	Acute Lymphoblastic Leukemia.	11
AML	Acute Myeloid Leukemia.	12
DBNs	Deep Belief Nets.	6
DFL	Discrete Function Learning.	14
DNA	Deoxyribonucleic Acid.	4
ER+	Estrogen Receptor Positive.	26
ER-	Estrogen Receptor Negative.	26
FCBF	Fast Correlation Based Filter.	11
HCC	Hepatocellular Carcinoma.	ii
HER2+	Human Epidermal Growth Factor Receptor 2 Positive.	26
HER2-	Human Epidermal Growth Factor Receptor 2 Negative.	26
KNN	K-Nearest Neighbors.	11
LDS	Low Density Separation.	15
MiRNAs	MicroRNAs.	5
MLFS	Multilevel Feature Selection.	ii
MLL	Mixed Lineage Leukemia.	12
mRNA	Messenger RNA.	14
PAM	Prediction Analysis of Microarrays.	14
RBM s	Restricted Boltzmann Machines.	6
RFE	Recursive Feature Elimination.	12
RF s	Random Forests.	24
RNA	Ribonucleic Acid.	5
SNP	Single Nucleotide Polymorphism.	59
SRBCT	Small Round Blue Cell Tumors.	35
SVM	Support Vector Machine.	11