

Performance key genes con: Train GSE68086, Test GSE156902

Mattia Manna

June 13, 2025

Contents

1	Classificazione con DEGS	1
1.1	Informazioni sulla numerosità dei dataset	1
1.2	Risultati performance	1
1.3	Summary DEGS	4
2	Classificazione con HUBS	5
2.1	Informazioni sulla numerosità dei dataset	5
2.2	Risultati performance	5
2.3	Summary HUBS	8
3	Classificazione con BETWEENNESS	9
3.1	Informazioni sulla numerosità dei dataset	9
3.2	Risultati performance	9
3.3	Summary BETWEENNESS	12
4	Classificazione con CLOSENESS	13
4.1	Informazioni sulla numerosità dei dataset	13
4.2	Risultati performance	13
4.3	Summary CLOSENESS	16
5	Classificazione con CLOSENESS_last5	17
5.1	Informazioni sulla numerosità dei dataset	17
5.2	Risultati performance	17
5.3	Summary CLOSENESS_last5	20
6	Classificazione con EIGENVECTOR	21
6.1	Informazioni sulla numerosità dei dataset	21
6.2	Risultati performance	21
6.3	Summary EIGENVECTOR	24
7	Classificazione con LCC	25
7.1	Informazioni sulla numerosità dei dataset	25
7.2	Risultati performance	25
7.3	Summary LCC	28
8	Classificazione con ElasticNet	29
8.1	Informazioni sulla numerosità dei dataset	29
8.2	Risultati performance	29
8.3	Summary ElasticNet	32

1 Classificazione con DEGS

1.1 Informazioni sulla numerosità dei dataset

Numerosità GSE68086 (train) all'importazione: (95, 862)

Numerosità classi: 'HC': 55, 'GBM': 40

Numerosità GSE156902 (test) all'importazione: (322, 4489)

Numerosità classi: 'HC': 252, 'GBM': 70

Numerosità GSE68086 (train): (95, 861). Ci sono 860 geni ed una variabile target.

Numerosità GSE156902 (test): (322, 4488). Ci sono 4487 geni ed una variabile target.

Numero di geni comuni tra i due dataset: 705

Seed utilizzato per la classificazione: 0

1.2 Risultati performance

1.2.1 Performance tramite Logistic Regression (LR)

Table 1: LR

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.863	0.676	0.714	0.694	0.905
HC	0.863	0.919	0.905	0.912	0.714

Table 2: Confusion Matrix LR

	Predicted GBM	Predicted HC
Actual GBM	50	20
Actual HC	24	228

1.2.2 Performance tramite Support Vector Classifier (SVC)

Table 3: SVC

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.783	0.500	0.614	0.551	0.829
HC	0.783	0.886	0.829	0.857	0.614

Table 4: Confusion Matrix SVC

	Predicted GBM	Predicted HC
Actual GBM	43	27
Actual HC	43	209

1.2.3 Performance tramite Support Vector Classifier (SVC) con KFold, k=5

Table 5: SVC KFold, k = 5

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.649	0.373	0.900	0.527	0.579
HC	0.649	0.954	0.579	0.721	0.900

Table 6: Confusion Matrix SVC KFold

	Predicted GBM	Predicted HC
Actual GBM	63	7
Actual HC	106	146

1.2.4 Performance tramite Random Forest Classifier

Table 7: RF

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.699	0.400	0.771	0.527	0.679
HC	0.699	0.914	0.679	0.779	0.771

Table 8: Confusion Matrix RF

	Predicted GBM	Predicted HC
Actual GBM	54	16
Actual HC	81	171

1.2.5 Performance tramite K Nearest Neighbors Classifier

Table 9: KNN

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.894	0.781	0.714	0.746	0.944
HC	0.894	0.922	0.944	0.933	0.714

Table 10: Confusion Matrix KNN

	Predicted GBM	Predicted HC
Actual GBM	50	20
Actual HC	14	238

1.2.6 Performance tramite Naive Bayes Binomial Classifier

Table 11: NB Binomial

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.795	0.667	0.114	0.195	0.984
HC	0.795	0.800	0.984	0.883	0.114

Table 12: Confusion Matrix NB Binomial

	Predicted GBM	Predicted HC
Actual GBM	8	62
Actual HC	4	248

1.2.7 Performance tramite Naive Bayes Gaussian Classifier

Table 13: NB Gaussian

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.581	0.296	0.671	0.410	0.556
HC	0.581	0.859	0.556	0.675	0.671

Table 14: Confusion Matrix NB Gaussian

	Predicted GBM	Predicted HC
Actual GBM	47	23
Actual HC	112	140

1.3 Summary DEGS

1.3.1 Summary DEGS train data

Table 15: Riassunto performance train, condizione di riferimento GBM

	Accuracy	Precision	Recall	F1-Score	Specificity
LR	0.895	0.969	0.775	0.861	0.982
SVC	0.874	0.938	0.750	0.833	0.964
RF	0.863	0.886	0.775	0.827	0.927
KNN	0.874	0.967	0.725	0.829	0.982
NB Binomial	0.737	0.826	0.475	0.603	0.927
NB Gaussian	0.779	0.757	0.700	0.727	0.836

1.3.2 Summary DEGS test data

Table 16: Riassunto performance test, condizione di riferimento GBM

	Accuracy	Precision	Recall	F1-Score	Specificity
LR	0.863	0.676	0.714	0.694	0.905
SVC	0.783	0.500	0.614	0.551	0.829
SVC KF	0.649	0.373	0.900	0.527	0.579
RF	0.699	0.400	0.771	0.527	0.679
KNN	0.894	0.781	0.714	0.746	0.944
NB Binomial	0.795	0.667	0.114	0.195	0.984
NB Gaussian	0.581	0.296	0.671	0.410	0.556

2 Classificazione con HUBS

2.1 Informazioni sulla numerosità dei dataset

Numerosità GSE68086 (train) all'importazione: (95, 44)

Numerosità classi: 'HC': 55, 'GBM': 40

Numerosità GSE156902 (test) all'importazione: (322, 4489)

Numerosità classi: 'HC': 252, 'GBM': 70

Numerosità GSE68086 (train): (95, 43). Ci sono 42 geni ed una variabile target.

Numerosità GSE156902 (test): (322, 4488). Ci sono 4487 geni ed una variabile target.

Numero di geni comuni tra i due dataset: 36

Seed utilizzato per la classificazione: 0

2.2 Risultati performance

2.2.1 Performance tramite Logistic Regression (LR)

Table 17: LR

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.764	0.474	0.786	0.591	0.758
HC	0.764	0.927	0.758	0.834	0.786

Table 18: Confusion Matrix LR

	Predicted GBM	Predicted HC
Actual GBM	55	15
Actual HC	61	191

2.2.2 Performance tramite Support Vector Classifier (SVC)

Table 19: SVC

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.913	0.828	0.757	0.791	0.956
HC	0.913	0.934	0.956	0.945	0.757

Table 20: Confusion Matrix SVC

	Predicted GBM	Predicted HC
Actual GBM	53	17
Actual HC	11	241

2.2.3 Performance tramite Support Vector Classifier (SVC) con KFold, k=5

Table 21: SVC KFold, k = 5

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.913	0.938	0.643	0.763	0.988
HC	0.913	0.909	0.988	0.947	0.643

Table 22: Confusion Matrix SVC KFold

	Predicted GBM	Predicted HC
Actual GBM	45	25
Actual HC	3	249

2.2.4 Performance tramite Random Forest Classifier

Table 23: RF

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.870	0.679	0.757	0.716	0.901
HC	0.870	0.930	0.901	0.915	0.757

Table 24: Confusion Matrix RF

	Predicted GBM	Predicted HC
Actual GBM	53	17
Actual HC	25	227

2.2.5 Performance tramite K Nearest Neighbors Classifier

Table 25: KNN

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.885	0.923	0.514	0.661	0.988
HC	0.885	0.880	0.988	0.931	0.514

Table 26: Confusion Matrix KNN

	Predicted GBM	Predicted HC
Actual GBM	36	34
Actual HC	3	249

2.2.6 Performance tramite Naive Bayes Binomial Classifier

Table 27: NB Binomial

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.786	0.600	0.043	0.080	0.992
HC	0.786	0.789	0.992	0.879	0.043

Table 28: Confusion Matrix NB Binomial

	Predicted GBM	Predicted HC
Actual GBM	3	67
Actual HC	2	250

2.2.7 Performance tramite Naive Bayes Gaussian Classifier

Table 29: NB Gaussian

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.575	0.297	0.700	0.417	0.540
HC	0.575	0.866	0.540	0.665	0.700

Table 30: Confusion Matrix NB Gaussian

	Predicted GBM	Predicted HC
Actual GBM	49	21
Actual HC	116	136

2.3 Summary HUBS

2.3.1 Summary HUBS train data

Table 31: Riassunto performance train, condizione di riferimento GBM

	Accuracy	Precision	Recall	F1-Score	Specificity
LR	0.905	0.919	0.850	0.883	0.945
SVC	0.916	0.971	0.825	0.892	0.982
RF	0.863	0.865	0.800	0.831	0.909
KNN	0.905	1.000	0.775	0.873	1.000
NB Binomial	0.779	0.952	0.500	0.656	0.982
NB Gaussian	0.832	0.833	0.750	0.789	0.891

2.3.2 Summary HUBS test data

Table 32: Riassunto performance test, condizione di riferimento GBM

	Accuracy	Precision	Recall	F1-Score	Specificity
LR	0.764	0.474	0.786	0.591	0.758
SVC	0.913	0.828	0.757	0.791	0.956
SVC KF	0.913	0.938	0.643	0.763	0.988
RF	0.870	0.679	0.757	0.716	0.901
KNN	0.885	0.923	0.514	0.661	0.988
NB Binomial	0.786	0.600	0.043	0.080	0.992
NB Gaussian	0.575	0.297	0.700	0.417	0.540

3 Classificazione con BETWEENNESS

3.1 Informazioni sulla numerosità dei dataset

Numerosità GSE68086 (train) all'importazione: (95, 44)

Numerosità classi: 'HC': 55, 'GBM': 40

Numerosità GSE156902 (test) all'importazione: (322, 4489)

Numerosità classi: 'HC': 252, 'GBM': 70

Numerosità GSE68086 (train): (95, 43). Ci sono 42 geni ed una variabile target.

Numerosità GSE156902 (test): (322, 4488). Ci sono 4487 geni ed una variabile target.

Numero di geni comuni tra i due dataset: 37

Seed utilizzato per la classificazione: 0

3.2 Risultati performance

3.2.1 Performance tramite Logistic Regression (LR)

Table 33: LR

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.742	0.445	0.757	0.561	0.738
HC	0.742	0.916	0.738	0.818	0.757

Table 34: Confusion Matrix LR

	Predicted GBM	Predicted HC
Actual GBM	53	17
Actual HC	66	186

3.2.2 Performance tramite Support Vector Classifier (SVC)

Table 35: SVC

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.916	0.831	0.771	0.800	0.956
HC	0.916	0.938	0.956	0.947	0.771

Table 36: Confusion Matrix SVC

	Predicted GBM	Predicted HC
Actual GBM	54	16
Actual HC	11	241

3.2.3 Performance tramite Support Vector Classifier (SVC) con KFold, k=5

Table 37: SVC KFold, k = 5

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.916	0.978	0.629	0.765	0.996
HC	0.916	0.906	0.996	0.949	0.629

Table 38: Confusion Matrix SVC KFold

	Predicted GBM	Predicted HC
Actual GBM	44	26
Actual HC	1	251

3.2.4 Performance tramite Random Forest Classifier

Table 39: RF

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.901	0.738	0.843	0.787	0.917
HC	0.901	0.955	0.917	0.935	0.843

Table 40: Confusion Matrix RF

	Predicted GBM	Predicted HC
Actual GBM	59	11
Actual HC	21	231

3.2.5 Performance tramite K Nearest Neighbors Classifier

Table 41: KNN

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.857	0.786	0.471	0.589	0.964
HC	0.857	0.868	0.964	0.914	0.471

Table 42: Confusion Matrix KNN

	Predicted GBM	Predicted HC
Actual GBM	33	37
Actual HC	9	243

3.2.6 Performance tramite Naive Bayes Binomial Classifier

Table 43: NB Binomial

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.783	0.500	0.014	0.028	0.996
HC	0.783	0.784	0.996	0.878	0.014

Table 44: Confusion Matrix NB Binomial

	Predicted GBM	Predicted HC
Actual GBM	1	69
Actual HC	1	251

3.2.7 Performance tramite Naive Bayes Gaussian Classifier

Table 45: NB Gaussian

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.602	0.316	0.714	0.439	0.571
HC	0.602	0.878	0.571	0.692	0.714

Table 46: Confusion Matrix NB Gaussian

	Predicted GBM	Predicted HC
Actual GBM	50	20
Actual HC	108	144

3.3 Summary BETWEENNESS

3.3.1 Summary BETWEENNESS train data

Table 47: Riassunto performance train, condizione di riferimento GBM

	Accuracy	Precision	Recall	F1-Score	Specificity
LR	0.916	0.944	0.850	0.895	0.964
SVC	0.926	1.000	0.825	0.904	1.000
RF	0.895	0.941	0.800	0.865	0.964
KNN	0.895	0.969	0.775	0.861	0.982
NB Binomial	0.768	0.950	0.475	0.633	0.982
NB Gaussian	0.853	0.861	0.775	0.816	0.909

3.3.2 Summary BETWEENNESS test data

Table 48: Riassunto performance test, condizione di riferimento GBM

	Accuracy	Precision	Recall	F1-Score	Specificity
LR	0.742	0.445	0.757	0.561	0.738
SVC	0.916	0.831	0.771	0.800	0.956
SVC KF	0.916	0.978	0.629	0.765	0.996
RF	0.901	0.738	0.843	0.787	0.917
KNN	0.857	0.786	0.471	0.589	0.964
NB Binomial	0.783	0.500	0.014	0.028	0.996
NB Gaussian	0.602	0.316	0.714	0.439	0.571

4 Classificazione con CLOSENESS

4.1 Informazioni sulla numerosità dei dataset

Numerosità GSE68086 (train) all'importazione: (95, 44)

Numerosità classi: 'HC': 55, 'GBM': 40

Numerosità GSE156902 (test) all'importazione: (322, 4489)

Numerosità classi: 'HC': 252, 'GBM': 70

Numerosità GSE68086 (train): (95, 43). Ci sono 42 geni ed una variabile target.

Numerosità GSE156902 (test): (322, 4488). Ci sono 4487 geni ed una variabile target.

Numero di geni comuni tra i due dataset: 37

Seed utilizzato per la classificazione: 0

4.2 Risultati performance

4.2.1 Performance tramite Logistic Regression (LR)

Table 49: LR

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.640	0.354	0.800	0.491	0.595
HC	0.640	0.915	0.595	0.721	0.800

Table 50: Confusion Matrix LR

	Predicted GBM	Predicted HC
Actual GBM	56	14
Actual HC	102	150

4.2.2 Performance tramite Support Vector Classifier (SVC)

Table 51: SVC

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.848	0.630	0.729	0.675	0.881
HC	0.848	0.921	0.881	0.901	0.729

Table 52: Confusion Matrix SVC

	Predicted GBM	Predicted HC
Actual GBM	51	19
Actual HC	30	222

4.2.3 Performance tramite Support Vector Classifier (SVC) con KFold, k=5

Table 53: SVC KFold, k = 5

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.463	0.271	0.871	0.414	0.349
HC	0.463	0.907	0.349	0.504	0.871

Table 54: Confusion Matrix SVC KFold

	Predicted GBM	Predicted HC
Actual GBM	61	9
Actual HC	164	88

4.2.4 Performance tramite Random Forest Classifier

Table 55: RF

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.441	0.267	0.900	0.412	0.313
HC	0.441	0.919	0.313	0.467	0.900

Table 56: Confusion Matrix RF

	Predicted GBM	Predicted HC
Actual GBM	63	7
Actual HC	173	79

4.2.5 Performance tramite K Nearest Neighbors Classifier

Table 57: KNN

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.795	0.520	0.729	0.607	0.813
HC	0.795	0.915	0.813	0.861	0.729

Table 58: Confusion Matrix KNN

	Predicted GBM	Predicted HC
Actual GBM	51	19
Actual HC	47	205

4.2.6 Performance tramite Naive Bayes Binomial Classifier

Table 59: NB Binomial

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.786	0.571	0.057	0.104	0.988
HC	0.786	0.790	0.988	0.878	0.057

Table 60: Confusion Matrix NB Binomial

	Predicted GBM	Predicted HC
Actual GBM	4	66
Actual HC	3	249

4.2.7 Performance tramite Naive Bayes Gaussian Classifier

Table 61: NB Gaussian

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.516	0.283	0.800	0.418	0.437
HC	0.516	0.887	0.437	0.585	0.800

Table 62: Confusion Matrix NB Gaussian

	Predicted GBM	Predicted HC
Actual GBM	56	14
Actual HC	142	110

4.3 Summary CLOSENESS

4.3.1 Summary CLOSENESS train data

Table 63: Riassunto performance train, condizione di riferimento GBM

	Accuracy	Precision	Recall	F1-Score	Specificity
LR	0.937	0.972	0.875	0.921	0.982
SVC	0.916	0.971	0.825	0.892	0.982
RF	0.853	0.842	0.800	0.821	0.891
KNN	0.905	0.970	0.800	0.877	0.982
NB Binomial	0.726	0.818	0.450	0.581	0.927
NB Gaussian	0.811	0.824	0.700	0.757	0.891

4.3.2 Summary CLOSENESS test data

Table 64: Riassunto performance test, condizione di riferimento GBM

	Accuracy	Precision	Recall	F1-Score	Specificity
LR	0.640	0.354	0.800	0.491	0.595
SVC	0.848	0.630	0.729	0.675	0.881
SVC KF	0.463	0.271	0.871	0.414	0.349
RF	0.441	0.267	0.900	0.412	0.313
KNN	0.795	0.520	0.729	0.607	0.813
NB Binomial	0.786	0.571	0.057	0.104	0.988
NB Gaussian	0.516	0.283	0.800	0.418	0.437

5 Classificazione con CLOSENESS_last5

5.1 Informazioni sulla numerosità dei dataset

Numerosità GSE68086 (train) all'importazione: (95, 44)

Numerosità classi: 'HC': 55, 'GBM': 40

Numerosità GSE156902 (test) all'importazione: (322, 4489)

Numerosità classi: 'HC': 252, 'GBM': 70

Numerosità GSE68086 (train): (95, 43). Ci sono 42 geni ed una variabile target.

Numerosità GSE156902 (test): (322, 4488). Ci sono 4487 geni ed una variabile target.

Numero di geni comuni tra i due dataset: 28

Seed utilizzato per la classificazione: 0

5.2 Risultati performance

5.2.1 Performance tramite Logistic Regression (LR)

Table 65: LR

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.516	0.293	0.871	0.439	0.417
HC	0.516	0.921	0.417	0.574	0.871

Table 66: Confusion Matrix LR

	Predicted GBM	Predicted HC
Actual GBM	61	9
Actual HC	147	105

5.2.2 Performance tramite Support Vector Classifier (SVC)

Table 67: SVC

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.469	0.283	0.943	0.436	0.337
HC	0.469	0.955	0.337	0.499	0.943

Table 68: Confusion Matrix SVC

	Predicted GBM	Predicted HC
Actual GBM	66	4
Actual HC	167	85

5.2.3 Performance tramite Support Vector Classifier (SVC) con KFold, k=5

Table 69: SVC KFold, k = 5

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.652	0.364	0.800	0.500	0.611
HC	0.652	0.917	0.611	0.733	0.800

Table 70: Confusion Matrix SVC KFold

	Predicted GBM	Predicted HC
Actual GBM	56	14
Actual HC	98	154

5.2.4 Performance tramite Random Forest Classifier

Table 71: RF

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.466	0.278	0.914	0.427	0.341
HC	0.466	0.935	0.341	0.500	0.914

Table 72: Confusion Matrix RF

	Predicted GBM	Predicted HC
Actual GBM	64	6
Actual HC	166	86

5.2.5 Performance tramite K Nearest Neighbors Classifier

Table 73: KNN

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.671	0.380	0.814	0.518	0.631
HC	0.671	0.924	0.631	0.750	0.814

Table 74: Confusion Matrix KNN

	Predicted GBM	Predicted HC
Actual GBM	57	13
Actual HC	93	159

5.2.6 Performance tramite Naive Bayes Binomial Classifier

Table 75: NB Binomial

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.786	0.667	0.029	0.055	0.996
HC	0.786	0.787	0.996	0.879	0.029

Table 76: Confusion Matrix NB Binomial

	Predicted GBM	Predicted HC
Actual GBM	2	68
Actual HC	1	251

5.2.7 Performance tramite Naive Bayes Gaussian Classifier

Table 77: NB Gaussian

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.463	0.273	0.886	0.418	0.345
HC	0.463	0.916	0.345	0.501	0.886

Table 78: Confusion Matrix NB Gaussian

	Predicted GBM	Predicted HC
Actual GBM	62	8
Actual HC	165	87

5.3 Summary CLOSENESS_last5

5.3.1 Summary CLOSENESS_last5 train data

Table 79: Riassunto performance train, condizione di riferimento GBM

	Accuracy	Precision	Recall	F1-Score	Specificity
LR	0.853	0.842	0.800	0.821	0.891
SVC	0.853	0.933	0.700	0.800	0.964
RF	0.853	0.882	0.750	0.811	0.927
KNN	0.853	0.933	0.700	0.800	0.964
NB Binomial	0.674	0.909	0.250	0.392	0.982
NB Gaussian	0.832	0.800	0.800	0.800	0.855

5.3.2 Summary CLOSENESS_last5 test data

Table 80: Riassunto performance test, condizione di riferimento GBM

	Accuracy	Precision	Recall	F1-Score	Specificity
LR	0.516	0.293	0.871	0.439	0.417
SVC	0.469	0.283	0.943	0.436	0.337
SVC KF	0.652	0.364	0.800	0.500	0.611
RF	0.466	0.278	0.914	0.427	0.341
KNN	0.671	0.380	0.814	0.518	0.631
NB Binomial	0.786	0.667	0.029	0.055	0.996
NB Gaussian	0.463	0.273	0.886	0.418	0.345

6 Classificazione con EIGENVECTOR

6.1 Informazioni sulla numerosità dei dataset

Numerosità GSE68086 (train) all'importazione: (95, 44)

Numerosità classi: 'HC': 55, 'GBM': 40

Numerosità GSE156902 (test) all'importazione: (322, 4489)

Numerosità classi: 'HC': 252, 'GBM': 70

Numerosità GSE68086 (train): (95, 43). Ci sono 42 geni ed una variabile target.

Numerosità GSE156902 (test): (322, 4488). Ci sono 4487 geni ed una variabile target.

Numero di geni comuni tra i due dataset: 38

Seed utilizzato per la classificazione: 0

6.2 Risultati performance

6.2.1 Performance tramite Logistic Regression (LR)

Table 81: LR

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.450	0.272	0.914	0.420	0.321
HC	0.450	0.931	0.321	0.478	0.914

Table 82: Confusion Matrix LR

	Predicted GBM	Predicted HC
Actual GBM	64	6
Actual HC	171	81

6.2.2 Performance tramite Support Vector Classifier (SVC)

Table 83: SVC

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.665	0.344	0.600	0.437	0.683
HC	0.665	0.860	0.683	0.761	0.600

Table 84: Confusion Matrix SVC

	Predicted GBM	Predicted HC
Actual GBM	42	28
Actual HC	80	172

6.2.3 Performance tramite Support Vector Classifier (SVC) con KFold, k=5

Table 85: SVC KFold, k = 5

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.711	0.376	0.500	0.429	0.770
HC	0.711	0.847	0.770	0.807	0.500

Table 86: Confusion Matrix SVC KFold

	Predicted GBM	Predicted HC
Actual GBM	35	35
Actual HC	58	194

6.2.4 Performance tramite Random Forest Classifier

Table 87: RF

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.360	0.244	0.929	0.387	0.202
HC	0.360	0.911	0.202	0.331	0.929

Table 88: Confusion Matrix RF

	Predicted GBM	Predicted HC
Actual GBM	65	5
Actual HC	201	51

6.2.5 Performance tramite K Nearest Neighbors Classifier

Table 89: KNN

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.677	0.345	0.543	0.422	0.714
HC	0.677	0.849	0.714	0.776	0.543

Table 90: Confusion Matrix KNN

	Predicted GBM	Predicted HC
Actual GBM	38	32
Actual HC	72	180

6.2.6 Performance tramite Naive Bayes Binomial Classifier

Table 91: NB Binomial

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.783	0.500	0.014	0.028	0.996
HC	0.783	0.784	0.996	0.878	0.014

Table 92: Confusion Matrix NB Binomial

	Predicted GBM	Predicted HC
Actual GBM	1	69
Actual HC	1	251

6.2.7 Performance tramite Naive Bayes Gaussian Classifier

Table 93: NB Gaussian

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.630	0.310	0.571	0.402	0.647
HC	0.630	0.845	0.647	0.733	0.571

Table 94: Confusion Matrix NB Gaussian

	Predicted GBM	Predicted HC
Actual GBM	40	30
Actual HC	89	163

6.3 Summary EIGENVECTOR

6.3.1 Summary EIGENVECTOR train data

Table 95: Riassunto performance train, condizione di riferimento GBM

	Accuracy	Precision	Recall	F1-Score	Specificity
LR	0.758	0.730	0.675	0.701	0.818
SVC	0.821	0.848	0.700	0.767	0.909
RF	0.842	0.821	0.800	0.810	0.873
KNN	0.811	0.867	0.650	0.743	0.927
NB Binomial	0.768	0.950	0.475	0.633	0.982
NB Gaussian	0.832	0.800	0.800	0.800	0.855

6.3.2 Summary EIGENVECTOR test data

Table 96: Riassunto performance test, condizione di riferimento GBM

	Accuracy	Precision	Recall	F1-Score	Specificity
LR	0.450	0.272	0.914	0.420	0.321
SVC	0.665	0.344	0.600	0.437	0.683
SVC KF	0.711	0.376	0.500	0.429	0.770
RF	0.360	0.244	0.929	0.387	0.202
KNN	0.677	0.345	0.543	0.422	0.714
NB Binomial	0.783	0.500	0.014	0.028	0.996
NB Gaussian	0.630	0.310	0.571	0.402	0.647

7 Classificazione con LCC

7.1 Informazioni sulla numerosità dei dataset

Numerosità GSE68086 (train) all'importazione: (95, 15)

Numerosità classi: 'HC': 55, 'GBM': 40

Numerosità GSE156902 (test) all'importazione: (322, 4489)

Numerosità classi: 'HC': 252, 'GBM': 70

Numerosità GSE68086 (train): (95, 14). Ci sono 13 geni ed una variabile target.

Numerosità GSE156902 (test): (322, 4488). Ci sono 4487 geni ed una variabile target.

Numero di geni comuni tra i due dataset: 7

Seed utilizzato per la classificazione: 0

7.2 Risultati performance

7.2.1 Performance tramite Logistic Regression (LR)

Table 97: LR

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.736	0.394	0.400	0.397	0.829
HC	0.736	0.833	0.829	0.831	0.400

Table 98: Confusion Matrix LR

	Predicted GBM	Predicted HC
Actual GBM	28	42
Actual HC	43	209

7.2.2 Performance tramite Support Vector Classifier (SVC)

Table 99: SVC

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.770	0.469	0.429	0.448	0.865
HC	0.770	0.845	0.865	0.855	0.429

Table 100: Confusion Matrix SVC

	Predicted GBM	Predicted HC
Actual GBM	30	40
Actual HC	34	218

7.2.3 Performance tramite Support Vector Classifier (SVC) con KFold, k=5

Table 101: SVC KFold, k = 5

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.780	0.491	0.400	0.441	0.885
HC	0.780	0.842	0.885	0.863	0.400

Table 102: Confusion Matrix SVC KFold

	Predicted GBM	Predicted HC
Actual GBM	28	42
Actual HC	29	223

7.2.4 Performance tramite Random Forest Classifier

Table 103: RF

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.755	0.433	0.414	0.423	0.849
HC	0.755	0.839	0.849	0.844	0.414

Table 104: Confusion Matrix RF

	Predicted GBM	Predicted HC
Actual GBM	29	41
Actual HC	38	214

7.2.5 Performance tramite K Nearest Neighbors Classifier

Table 105: KNN

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.814	0.600	0.429	0.500	0.921
HC	0.814	0.853	0.921	0.885	0.429

Table 106: Confusion Matrix KNN

	Predicted GBM	Predicted HC
Actual GBM	30	40
Actual HC	20	232

7.2.6 Performance tramite Naive Bayes Binomial Classifier

Table 107: NB Binomial

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.792	0.800	0.057	0.107	0.996
HC	0.792	0.792	0.996	0.882	0.057

Table 108: Confusion Matrix NB Binomial

	Predicted GBM	Predicted HC
Actual GBM	4	66
Actual HC	1	251

7.2.7 Performance tramite Naive Bayes Gaussian Classifier

Table 109: NB Gaussian

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.590	0.288	0.600	0.389	0.587
HC	0.590	0.841	0.587	0.692	0.600

Table 110: Confusion Matrix NB Gaussian

	Predicted GBM	Predicted HC
Actual GBM	42	28
Actual HC	104	148

7.3 Summary LCC

7.3.1 Summary LCC train data

Table 111: Riassunto performance train, condizione di riferimento GBM

	Accuracy	Precision	Recall	F1-Score	Specificity
LR	0.779	0.788	0.650	0.712	0.873
SVC	0.789	0.955	0.525	0.677	0.982
RF	0.758	0.743	0.650	0.693	0.836
KNN	0.716	0.697	0.575	0.630	0.818
NB Binomial	0.695	0.923	0.300	0.453	0.982
NB Gaussian	0.768	0.765	0.650	0.703	0.855

7.3.2 Summary LCC test data

Table 112: Riassunto performance test, condizione di riferimento GBM

	Accuracy	Precision	Recall	F1-Score	Specificity
LR	0.736	0.394	0.400	0.397	0.829
SVC	0.770	0.469	0.429	0.448	0.865
SVC KF	0.780	0.491	0.400	0.441	0.885
RF	0.755	0.433	0.414	0.423	0.849
KNN	0.814	0.600	0.429	0.500	0.921
NB Binomial	0.792	0.800	0.057	0.107	0.996
NB Gaussian	0.590	0.288	0.600	0.389	0.587

8 Classificazione con ElasticNet

8.1 Informazioni sulla numerosità dei dataset

Numerosità GSE68086 (train) all'importazione: (95, 60)

Numerosità classi: 'HC': 55, 'GBM': 40

Numerosità GSE156902 (test) all'importazione: (322, 4489)

Numerosità classi: 'HC': 252, 'GBM': 70

Numerosità GSE68086 (train): (95, 59). Ci sono 58 geni ed una variabile target.

Numerosità GSE156902 (test): (322, 4488). Ci sono 4487 geni ed una variabile target.

Numero di geni comuni tra i due dataset: 37

Seed utilizzato per la classificazione: 0

8.2 Risultati performance

8.2.1 Performance tramite Logistic Regression (LR)

Table 113: LR

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.916	0.877	0.714	0.787	0.972
HC	0.916	0.925	0.972	0.948	0.714

Table 114: Confusion Matrix LR

	Predicted GBM	Predicted HC
Actual GBM	50	20
Actual HC	7	245

8.2.2 Performance tramite Support Vector Classifier (SVC)

Table 115: SVC

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.910	0.956	0.614	0.748	0.992
HC	0.910	0.903	0.992	0.945	0.614

Table 116: Confusion Matrix SVC

	Predicted GBM	Predicted HC
Actual GBM	43	27
Actual HC	2	250

8.2.3 Performance tramite Support Vector Classifier (SVC) con KFold, k=5

Table 117: SVC KFold, k = 5

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.910	0.902	0.657	0.760	0.980
HC	0.910	0.911	0.980	0.945	0.657

Table 118: Confusion Matrix SVC KFold

	Predicted GBM	Predicted HC
Actual GBM	46	24
Actual HC	5	247

8.2.4 Performance tramite Random Forest Classifier

Table 119: RF

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.907	0.955	0.600	0.737	0.992
HC	0.907	0.899	0.992	0.943	0.600

Table 120: Confusion Matrix RF

	Predicted GBM	Predicted HC
Actual GBM	42	28
Actual HC	2	250

8.2.5 Performance tramite K Nearest Neighbors Classifier

Table 121: KNN

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.898	0.951	0.557	0.703	0.992
HC	0.898	0.890	0.992	0.938	0.557

Table 122: Confusion Matrix KNN

	Predicted GBM	Predicted HC
Actual GBM	39	31
Actual HC	2	250

8.2.6 Performance tramite Naive Bayes Binomial Classifier

Table 123: NB Binomial

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.811	0.846	0.157	0.265	0.992
HC	0.811	0.809	0.992	0.891	0.157

Table 124: Confusion Matrix NB Binomial

	Predicted GBM	Predicted HC
Actual GBM	11	59
Actual HC	2	250

8.2.7 Performance tramite Naive Bayes Gaussian Classifier

Table 125: NB Gaussian

	Accuracy	Precision	Recall	F1-Score	Specificity
GBM	0.829	0.660	0.443	0.530	0.937
HC	0.829	0.858	0.937	0.896	0.443

Table 126: Confusion Matrix NB Gaussian

	Predicted GBM	Predicted HC
Actual GBM	31	39
Actual HC	16	236

8.3 Summary ElasticNet

8.3.1 Summary ElasticNet train data

Table 127: Riassunto performance train, condizione di riferimento GBM

	Accuracy	Precision	Recall	F1-Score	Specificity
LR	0.979	1.000	0.950	0.974	1.000
SVC	0.958	1.000	0.900	0.947	1.000
RF	0.905	0.943	0.825	0.880	0.964
KNN	0.926	0.971	0.850	0.907	0.982
NB Binomial	0.811	0.893	0.625	0.735	0.945
NB Gaussian	0.863	0.846	0.825	0.835	0.891

8.3.2 Summary ElasticNet test data

Table 128: Riassunto performance test, condizione di riferimento GBM

	Accuracy	Precision	Recall	F1-Score	Specificity
LR	0.916	0.877	0.714	0.787	0.972
SVC	0.910	0.956	0.614	0.748	0.992
SVC KF	0.910	0.902	0.657	0.760	0.980
RF	0.907	0.955	0.600	0.737	0.992
KNN	0.898	0.951	0.557	0.703	0.992
NB Binomial	0.811	0.846	0.157	0.265	0.992
NB Gaussian	0.829	0.660	0.443	0.530	0.937