### Informazioni sui dati e modelli

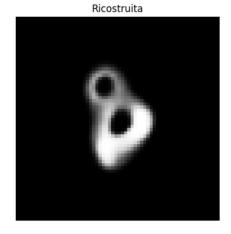
Divisione dati autoencoder: 0,2 validation e 0,2 test

2517 immagini train aumentate a 15102 787 validation 630 Test

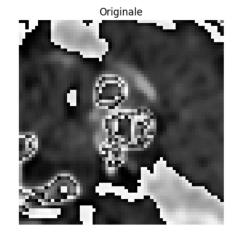
I modelli sono stati allenati sulle immagini ottenute tagliando il 30% dell'area

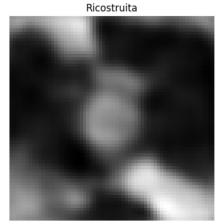
L'autoencoder che usa le immagini complete è leggermente modificato rispetto a quello utilizzato per le immagini con sfondo nero (dropout, kernels più grandi, custom loss)

Originale con Maschera



Esempio di maschera per ssim





Esempio di immagine originale (le due immagini sopra e sotto sono la stessa placca)

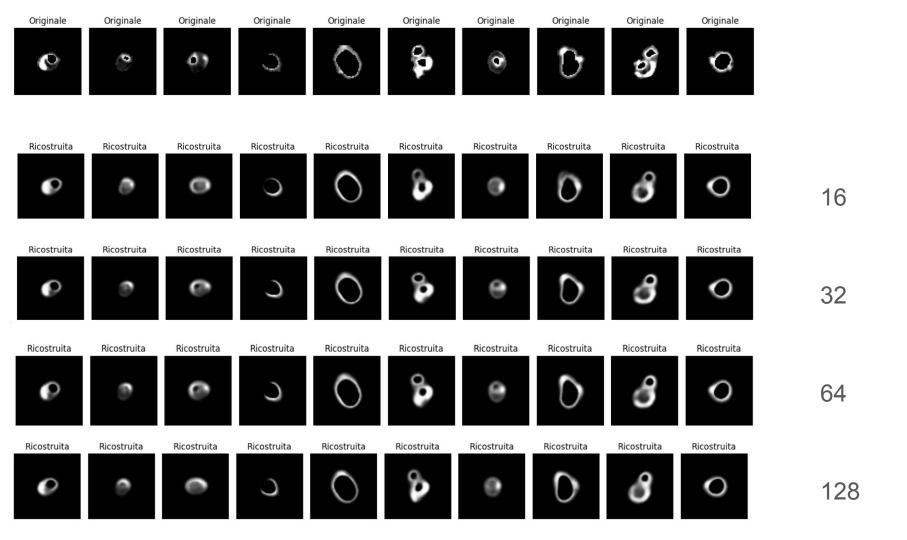






### Comparing similarity indexes for different autoencoders

Autoencoder	Mean SSIM (sfondo nero)	Mean SSIM Mask (sfondo nero)	Mean Cosine Similarity (sfondo nero)	Mean SSIM (originale)
16	0.9213	0.5995	0.8635	Non testato
32	0.9431	0.7126	0.9160	0.1114
64	0.9454	0.7239	0.9056	0.1324
128	0.9403	0.6987	0.8989	0.1913





Di seguito i risultati ottenuti facendo prima feature correlation (0.85), poi mantenendo le features con p-value (calcolato con t-test indipendente) inferiore a 0.1 (ipotesi di significatività relativa al gruppo 0 e 1), poi si confrontano i 5 classifiers (SVM, Logistic, MLP, RandomForerest, XgBoost) a cui si applica feature selection (RandomForest, Logistic, mrmr, lasso e p\_value (applicato partendo dalla sola feature più significativa e considerandone una in più a ogni iterazione)

69 pazienti train set 30 nel validation set 25 nel test set

Numero immagini partenza	solo area >= 30%	solo area >= 50%	solo area >= 70%
5049	3934	2991	1929

# Risultati immagini con sfondo nero

#### Val sopra, Test sotto nel caso shuffle di train e Aree 0.3

_

43

majority

voting

		vai su	ora, res	i Solio II	ei caso	Shume	ıı li aiii e	Alee U	.5	
N.features prima e dopo	Auc	Pr Auc	Precision	Recall	f1	confusion matrix	Accuracy	classificator e/Selector	mode	N-features selected by p_value (0.1)
32 -> 4	0.777	0.731	0.737	1	0.848	11 5 0 14	0.833	XgBoost/Lo giatic	majority voting	12
64 -> 19	0.741	0.6997	0.666	1	0.8	9 7 0 14	0.766	MLP/p_valu e	majority voting	31
128 -> 7	0.817	0.767	0.8235	1	0.903	13 3 0 14	0.9	Logistic/R andomFor est	majority voting	43
									1	
32 -> 4	0.471	0.34	0.333	0.625	0.435	7 10 3 5	0.48	XgBoost/Lo gistic	majority voting	12
64 -> 19	0.392	0.28	0.294	0.625	0.4	5 12 3 5	0.4	MLP/p_val ue	majority voting	31

0.3478

6 11

0.4

Logistic/R

andomFor

est

0.5

0.266

0.2564

0.2941

128 ->

N.features prima e dopo	Precision	Recall	f1	confusion matrix	Accuracy	classificatore/ Selector	mode
32 -> 4	0.25	0.5	0.333	5 12 4 4	0.36	XgBoost/Logi stic	majority voting
64 -> 19	0.357	0.625	0.4545	8 9 3 5	0.52	MLP/p_valu e	majority voting
128 -> 7	0.214	0.375	0.2727	6 11 5 3	0.36	Logistic/Ra ndomForest	majority voting

### Val sopra, Test sotto nel caso NON shuffle di train e val 0.3

N.features prima e dopo	Auc	Pr Auc	Precision	Recall	f1	confusion matrix	Accuracy	classificator e/Selector	mode	N-features selected by p_value (0.1)
32 -> 15	0.55	0.3855	0.55	1	0.71	10 9 0 11	0.7	MLP/Logis tic	majority voting	19
64 -> 21	0.445	0.4	0.478	1	0.647	7 12 0 11	0.6	MLP/Rando mForest	majority voting	30
128 -> 8	0.77	0.63	0.588	0.909	0.714	12 7 1 10	0.733	MLP/p_valu e	mean	48

32 -> 15	0.507	0.338	0.2727	0.375	0.3158	9 8 5 3	0.48	MLP/Logis tic	majority voting	19
64 -> 21	0.44	0.32	0.286	0.5	0.363	7 10 4 4	0.44	MLP/Rand omForest	majority voting	30
128 -> 8	0.338	0.27	0.333	0.5	0.4	9 8 4 4	0.4	MLP/p_val ue	mean	48

N.features prima e dopo	Precision	Recall	f1	confusion matrix	Accuracy	classificatore/ Selector	mode
32 -> 15	0.277	0.625	0.3846	4 13 3 5	0.36	MLP/Logisti c	majority voting
64 -> 21	0.4	1	0.571	5 12 0 8	0.52	MLP/Rando mForest	majority voting
128 -> 8	0.266	0.5	0.3478	6 11 4 4	0.4	MLP/p-valu e	mean

#### Val sopra, Test sotto nel caso shuffle di train e aree 0.5

		vai oo	pia, ico	t ootto ii	Ci dado	orianic (	ai tiaiii t	J GICC V	. •	
N.features prima e dopo	Auc	Pr Auc	Precision	Recall	f1	confusion matrix	Accuracy	classificator e/Selector	mode	N-features selected by p_value (0.1)
32 -> 9	0.714	0.593	0.666	1	0.8	9 7 0 14	0.766	RandomFor est/mrmr	mean	10
64 -> 6	0.7678	0.699	0.6842	0.9286	0.7878	10 6 1 13	0.766	MLP/mrmr	majority voting	23
128 -> 13	0.8527	0.7458	0.7368	1	0.8484	11 5 0 14	0.8333	MLP/mrmr	mean	44
32 -> 9	0.4853	0.3144	0.375	0.75	0.5	7 10 2 6	0.52	RandomFor est/mrmr	mean	10
64 -> 6	0.4044	0.2933	0.222	0.25	0.235	10 7 6 2	0.48	MLP/mrmr	majority voting	23

0.4348

7 10

0.48

MLP/mrmr

mean

44

0.2986

0.4412

128 ->

13

0.333

0.625

N.features prima e dopo	Precision	Recall	f1	confusion matrix	Accuracy	classificatore/ Selector	mode
32 -> 9	0.3125	0.625	0.4166	6 11 3 5	0.44	RandomFore st/mrmr	mean
64 -> 6	0.4166	0.625	0.5	10 7 3 5	0.6	MLP/mrmr	majority voting
128 -> 13	0.2149	0.375	0.2727	6 11 5 3	0.36	MLP/mrmr	mean

Val sopra, Test sotto nel caso NON shuffle di train e aree 0.5

N.features prima e dopo	Auc	Pr Auc	Precision	Recall	f1	confusion matrix	Accuracy	classificator e/Selector	mode	N-features selected by p_value
32 -> 15	0.5454	0.398	0.5	0.909	0.645	9 10 1 10	0.633	MLP/p_val	majority voting	20
64 -> 2	0.6124	0.5519	0.4583	1	0.6286	6 13 0 11	0.566	XgBoost/p_ value	mean	27
128 -> 7	0.7655	0.6061	0.6111	1	0.7586	12 7 1 11	0.7666	XgBoost/mr mr	mean	57

32 -> 15	0.4485	0. 3228	0.3077	0.5	0.3809	8 9 4 4	0.48	MLP/p_val ue	majority voting	20
64 -> 2	0.3676	0.3777	0.25	0.625	0.3571	2 15 3 5	0.28	XgBoost/p_ value	mean	27
128 -> 7	0.1985	0.2328	0.1333	0.25	0.1739	4 13 6 2	0.24	XgBoost/mr mr	mean	57

N.features prima e dopo	Precision	Recall	f1	confusion matrix	Accuracy	classificatore/ Selector	mode
32 -> 15	0.2308	0.375	0.2857	7 10 5 3	0.4	MLP/p_valu e	majority voting
64 ->2	0.32	1	0.4848	0 17 0 8	0.32	XgBoost/p_v aluet	mean
128 -> 7	0.1818	0.25	0.2105	8 9 6 2	0.4	XgBoost/mrm r	mean

### Val sopra, Test sotto nel caso shuffle di train e aree 0.7

N.features prima e dopo	Auc	Pr Auc	Precision	Recall	f1	confusion matrix	Accuracy	classificator e/Selector	mode	N-features selected by p_value (0.1)
32 -> 4	0.786	0.733	0.722	0.929	0.8125	11 5 1 13	0.8	XGBoost/L ogiatic	mean	12
64 -> 14	0.7321	0.7	0.722	0.9286	0.8125	11 5 1 13	0.8	XgBoost/p_ value	majority voting	14
128 -> 48	0.933	0.9121	0.8235	1	0.9032	13 3 0 14	0.9	MLP/Logis tic	mean	55

32 -> 4	0.471	0.337	0.357	0.625	0.4545	8 9 3 5	0.52	XGBoost/L ogiatic	mean	12
64 -> 14	0.5662	0.3716	0.333	0.75	0.4615	5 12 2 6	0.44	XgBoost/p_ value	majority voting	14
128 -> 48	0.5	0.3244	0.3077	0.5	0.3809	8 9 4 4	0.48	MLP/Logis tic	mean	55

N.features prima e dopo	Precision	Recall	f1	confusion matrix	Accuracy	classificatore/ Selector	mode
32 ->4	0.3125	0.625	0.417	6 11 3 5	0.44	XGBoost/Log iatic	mean
64 ->	0.277	0.625	0.3846	4 13 3 5	0.36	XgBoost/p_v alue	majority voting
128 -> 48	0.222	0.25	0.2353	10 7 6 2	0.48	MLP/Logisti c	mean

Val sopra, Test sotto nel caso NON shuffle di train e aree0.7

N.features prima e dopo	Auc	Pr Auc	Precision	Recall	f1	confusion matrix	Accuracy	classificator e/Selector	mode	N-features selected by p_value (0.1)
32 -> 9	0.775	0.721	0.555	0.909	0.689	11 8 1 10	0.7	MLP/mrmr	majority voting	18
64 -> 3	0.6316	0.514	0.4583	1	0.6286	6 13 0 11	0.566	RandomFor est/mrmr	mean	19
128 -> 6	0.6363	0.4888	0.55	1	0.7097	10 9 0 11	0.7	SVM/mrmr	majority voting	53

32 -> 9	0.4338	0.2996	0.3333	0.5	0.4	9 8 4 4	0.52	MLP/mrmr	majority voting	18
64 -> 3	0.3897	0.2813	0.375	0.75	0.5	7 10 2 6	0.52	RandomFor est/mrmr	mean	19
128 -> 6	0.2647	0.2447	0.2	0.375	0.2609	5 12 5 3	0.32	SVM/mrmr	majority voting	53

N.features prima e dopo	Precision	Recall	f1	confusion matrix	Accuracy	classificatore/ Selector	mode
32 -> 9	0.25	0.375	0.3	8 9 5 3	0.44	MLP/mrmr	majority voting
64 -> 3	0.3888	0.875	0.5384	6 11 1 7	0.52	RandomFore st/mrmr	mean
128 -> 6	0.3888	0.875	0.5385	6 11 1 7	0.52	SVM/mrmr	majority voting

# Risultati immagini originali (complete)

### Val sopra, Test sotto nel caso shuffle di train e aree 0.3

N.features prima e dopo	Auc	Pr Auc	Precision	Recall	f1	confusion matrix	Accuracy	classificator e/Selector	mode	N-features selected by p_value (0.1)
32 -> 9	0.709	0.646	0.65	0.929	0.765	9 7 1 13	0.7	RF/rf	mean	25
64 -> 3	0.835	0.826	0.765	0.929	0.839	12 4 1 13	0.833	SVM/mrmr	majority voting	45
128 -> 2	0.9152	0.907	0.7368	1	0.8484	11 5 0 14	0.8333	RandomF orest/p_va lue	majority voting	74
32 -> 9	0.603	0.489	0.455	0.625	0.526	11 8 3 5	0.64	RF/rf	mean	25
64 -> 3	0.39	0.402	0.25	0.5	0.333	5 12 4 4	0.36	SVM/mrmr	majority voting	45
128 -> 2	0.4853	0.4094	0.3333	0.625	0.4348	7 10 3 5	0.48	RandomF orest/p_va lue	majority voting	74

N.features prima e dopo	Precision	Recall	f1	confusion matrix	Accuracy	classificatore/ Selector	mode
32 -> 9	0.273	0.375	0.316	9 8 5 3	0.48	RF/rf	mean
64 -> 3	0.353	0.75	0.48	6 11 2 6	0.48	SVM/mrmr	majority voting
128 -> 2	0.4286	0.75	0.5454	9 8 2 6	0.6	RandomFor est/p_value	majority voting

### Val sopra, Test sotto nel caso SENZA shuffle di train e aree 0.3

N.features prima e dopo	Auc	Pr Auc	Precision	Recall	f1	confusion matrix	Accuracy	classificator e/Selector	mode	N-features selected by p_value (0.1)
32 -> 7	0.742	0.637	0.666	0.909	0.769	14 5 1 10	0.8	XGBoost/lo gistic	mean	28
64 -> 22	0.799	0.666	0.647	1	0.786	13 63 0 11	0.8	XGBoost/L ogistic	mean	42
128 -> 32	0.818	0.698	0.647	1	0.786	13 6 0 11	0.8	XGBoostt/ mrmr	majority voting	100

32 -> 7	0.544	0.456	0.333	0.625	0.435	7 10 3 5	0.48	XGBoost/lo gistic	mean	28
64 -> 22	0.397	0.383	0.333	0.5	0.4	9 8 4 4	0.52	XGBoost/L ogistic	mean	42
128 -> 32	0.426	0.297	0.266	0.5	0.348	6 11 4 4	0.4	XGBoostt/ mrmr	majority voting	100

N.features prima e dopo	Precision	Recall	f1	confusion matrix	Accuracy	classificatore/ Selector	mode
32 -> 7	0.222	0.25	0.235	10 7 6 2	0.48	XGBoost/logi stic	mean
64 -> 22	0.286	0.5	0.364	7 10 4 4	0.44	XGBoost/Log istic	mean
128 -> 32	0.385	0.625	0.476	9 8 3 5	0.56	XGBoostt/m rmr	majority voting

# Confronto fra i migliori risultati basato sull'allenamento con dataset intero

N.features prima e dopo	Precision	Recall	f1	confusion matrix	Accuracy	classificator e/Selector	mode	Immagine	Shuffle
32 -> 9	0.3125	0.625	0.4166	6 11 3 5	0.44	RandomFor est/mrmr	mean	Sfondo nero, area 0.5	Si
64 -> 21	0.4	1	0.571	5 12 0 8	0.52	MLP/Rand omForest	majority voting	Sfondo nero, area 0.7	No
128 -> 2	0.4286	0.75	0.5454	9 8 2 6	0.6	RandomF orest/p_va lue	majority voting	Originale, area 0.3	Si

### F

# CONCLUSIONI

- Avere 128 features non corrisponde a un indice di similarità migliore.
- Anche partendo da 128 features, i migliori selector ne selezionano sempre poche.
- il selector Lasso non viene mai preso come migliore
- il classifier logistic non viene mai scelto come migliore
- potremmo provare a creare una rete neurale anche per classificare per vedere se otteniamo dei miglioramenti
- Abbiamo svolto diversi test sull'autoencoder per provare ad ottenere delle ricostruzioni accettabili ma senza cambiamenti, potremmo magari provare un autoencoder diverso, per esempio una rete pretrainata come encoder o utilizzare un transformer...
- Utilizzare le immagini originali ha portato a risultati migliori (in termini di confusion matrix) rispetto a 32 e 64, nonostante il basso valore di ssim
- La riduzione dellé are sembra aiutare la classificazione



# Vecchi risultati per confronto

### Val sopra, Test sotto nel caso senza shuffle di train e val

N.features prima e dopo	Auc	Pr Auc	Precision	Recall	f1	confusion matrix	Accuracy	classificatore/ Selector
16 -> 11	0.684	0.5085	0.5	1	0.666	8 11 0 11	0.633	SVM/MRMR
32 -> 15	0.565	0.622	0.526	0.909	0.666	10 9 1 10	0.666	SVM/LOGIST IC
64 -> 4	0.789	0.721	0.625	0.909	0.741	13 6 1 10	0.767	MLP/LOGIST IC
128 -> 6	0.713	0.676	0.5	1	0.6	8 11 0 11	0.633	XGBOOST/R F
16 -> 11	0.5	0.34	0.285	0.5	0.3636	7 10 4 4	0.44	SVM/MRMR
32 -> 15	0.463	0.398	0.308	0.5	0.3809	8 9 4 4	0.48	SVM/LOGIST IC
64 -> 4	0.434	0.337	0.273	0.375	0.316	9 8 5 3	0.48	MLP/LOGIST IC
128 -> 6	0.353	0.314	0.25	0.625	0.357	2 15 3 5	0.28	XGBOOST/R F

### Val sopra, Test sotto nel caso shuffle di train e val

		•	•					
N.features prima e dopo	Auc	Pr Auc	Precision	Recall	f1	confusion matrix	Accuracy	classificatore/ Selector
16 -> 13	0.69	0.589	0.75	0.857	0.8	12 4 2 12	0.8	MLP/LOGIST IC
32 -> 5	0.857	0.846	0.8125	0.9286	0.866	13 3 1 13	0.866	SVM/LASSO
64 ->3	0.8214	0.778	0.7	1	0.8235	10 6 0 14	0.8	SVM/LOGIST IC
128 -> 10	0.835	0.744	0.737	1	0.8484	11 5 0 14	0.833	MLP/LOGIST IC
16 -> 13	0.330	0.262	0.222	0.25	0.2353	10 7 6 2	0.48	MLP/LOGIST IC.
32 -> 5	0.375	0.2775	0.3	0.375	0.333	10 7 5 3	0.52	SVM/LASSO
64 -> 3	0.5588	0.4076	0.25	0.375	0.3	8 9 5 3	0.44	SVM/LOGIST IC
128 -> 10	0.375	0.282	0.294	0.625	0.4	5 12 3 5	0.4	MLP/LOGIST IC