Sviluppo di modelli QSAR predittivi mediante tecniche di Machine Learning: applicazione ad inibitori dell'HDAC1

Tesi Sperimentale
in
Chimica Farmaceutica (CHIM08)

SAPIENZA
UNIVERSITÀ DI ROMA

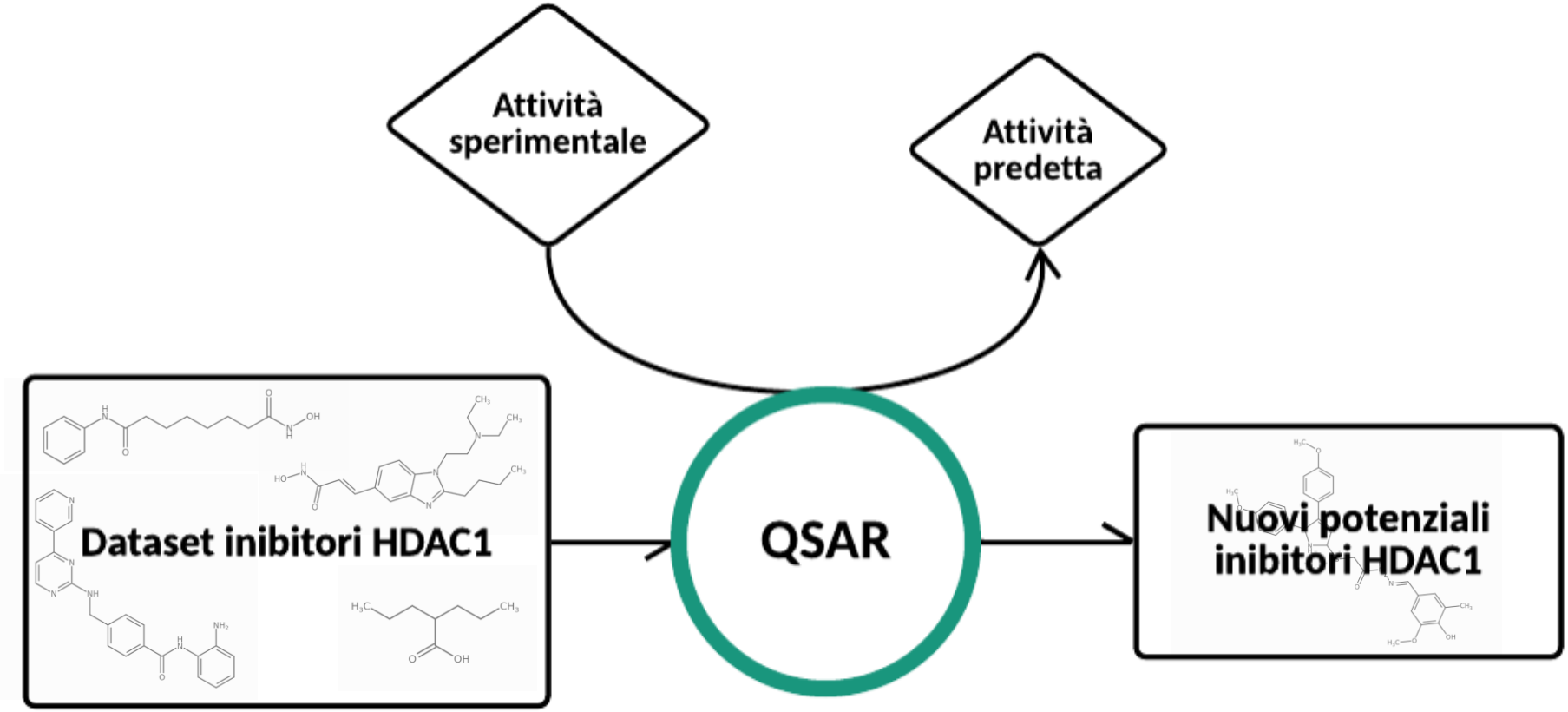
FACOLTÀ DI FARMACIA E MEDICINA
CORSO DI LAUREA IN
Chimica e Tecnologia
Farmaceutiche

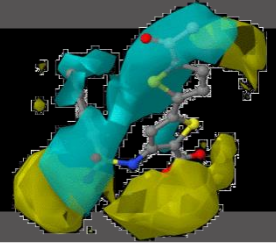
CANDIDATA:
Paola Caprioli
(matr. 1559309)

RELATORE:
Prof. Rino Ragno

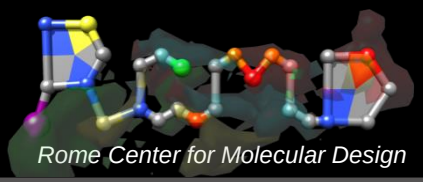
ANNO ACCADEMICO
2019/2020

Scopo del lavoro

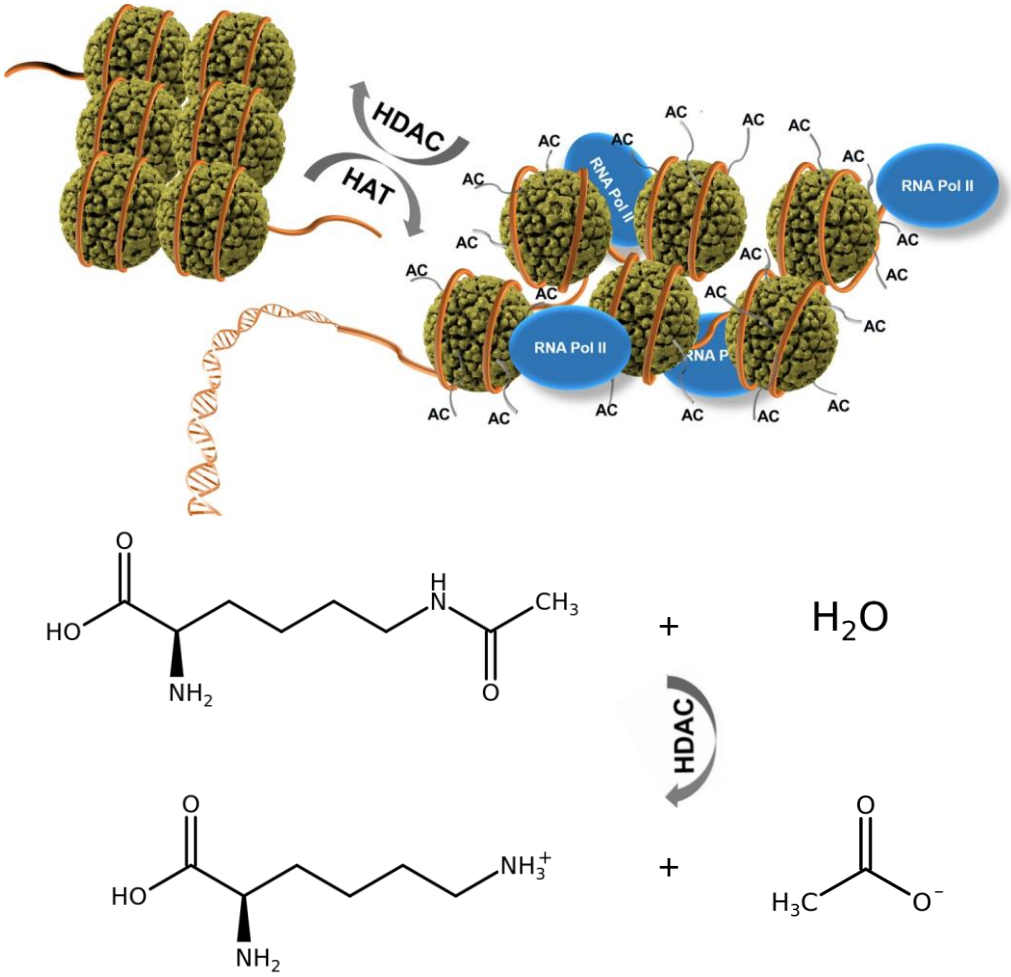




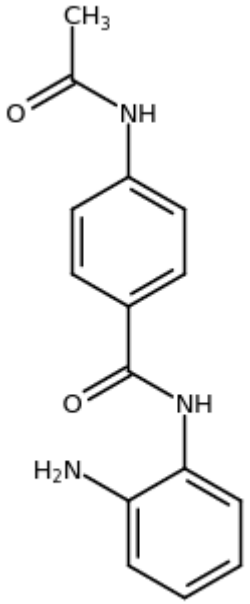
Target: HDAC 1



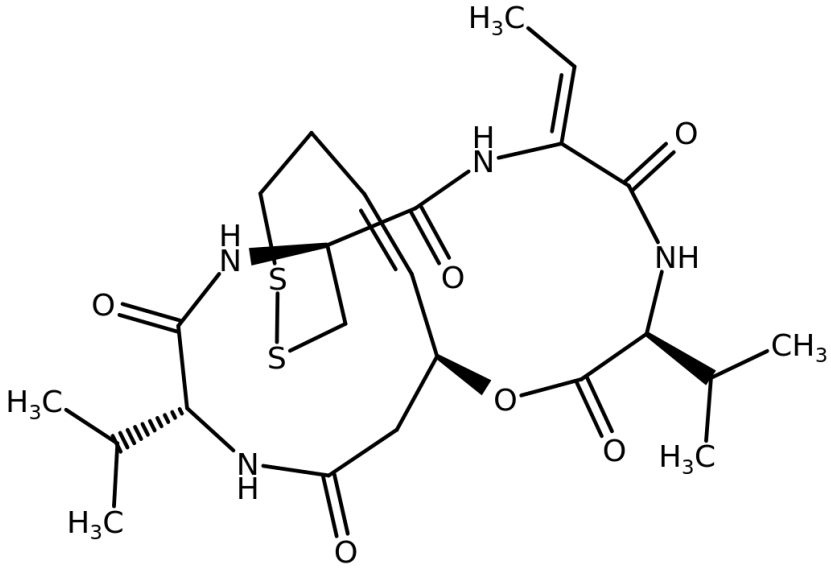
CLASSE	ENZIMA
I	HDAC1 HDAC2 HDAC3 HDAC8
IIa	HDAC4 HDAC5 HDAC7 HDAC9
IIb	HDAC6 HDAC10
III	SIRTUINE
IV	HDAC11



Inibitori HDAC1

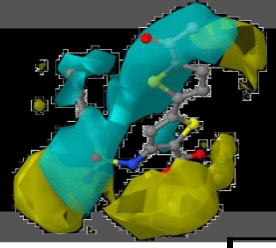


Tacedinaline

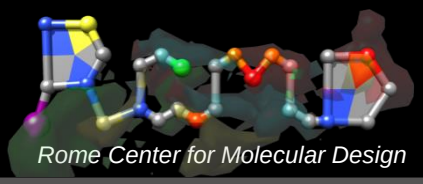


Romidepsina

Trials clinici di fase 2	**Approvato FDA (2009)**



Procedura sperimentale

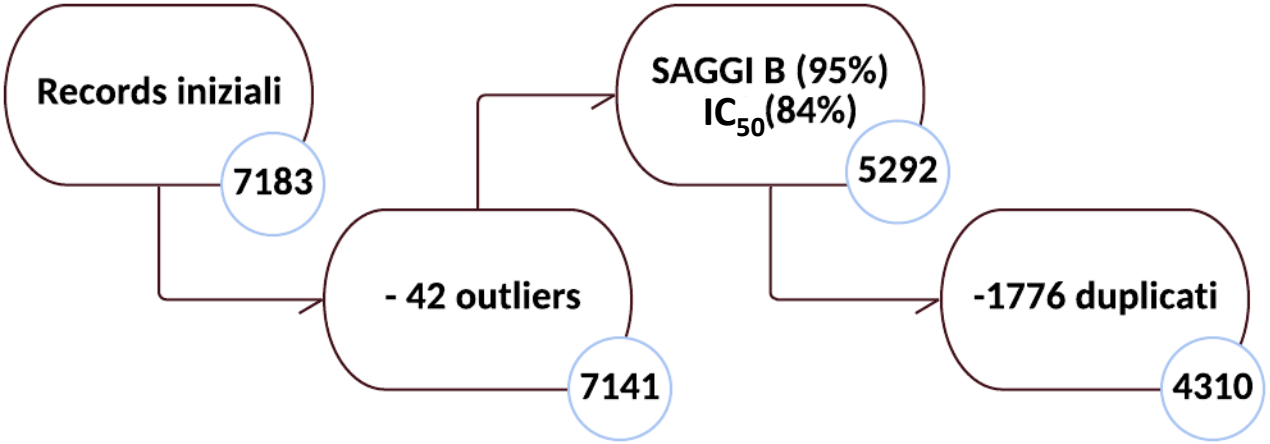


Steps

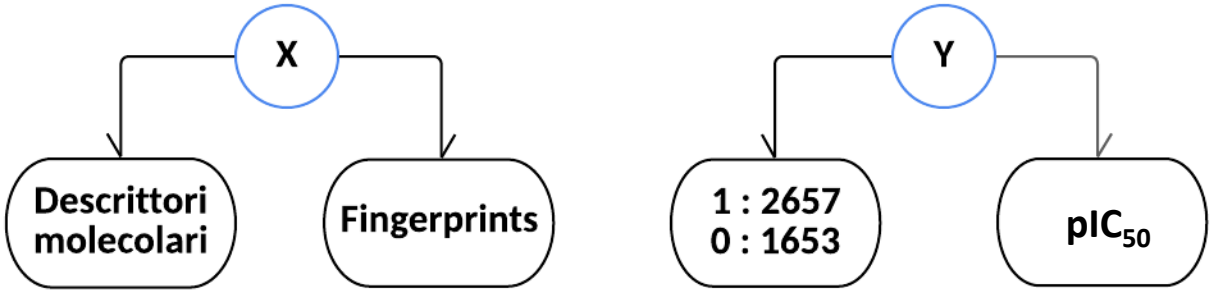
1	Generazione training set
2	Sviluppo modelli di classificazione binaria e regressione
3	Validazione esterna modelli
4	Analisi similarità molecolare test - training sets
5	Studio di selettività HDAC1-HDAC4
6	Virtual screening

1.Generazione Training set

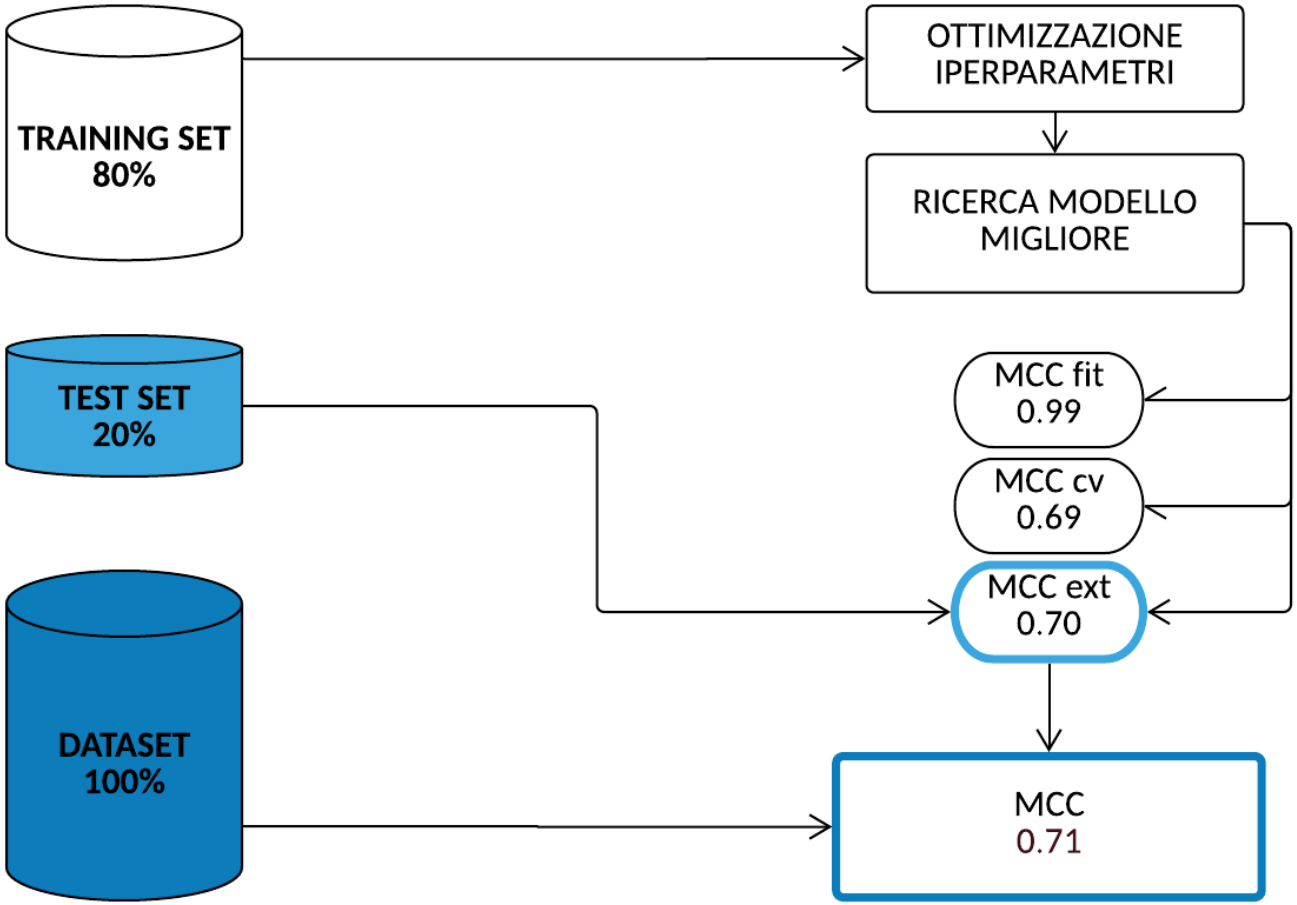
Generazione dataset



Generazione features(X) e labels(Y)



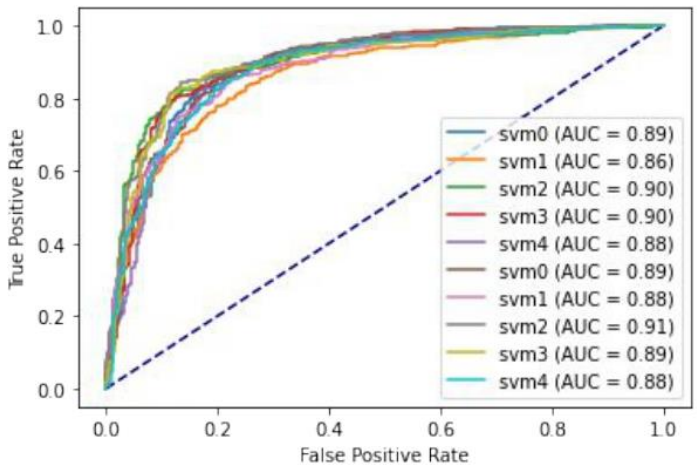
2.Sviluppo modelli classificazione e regressione



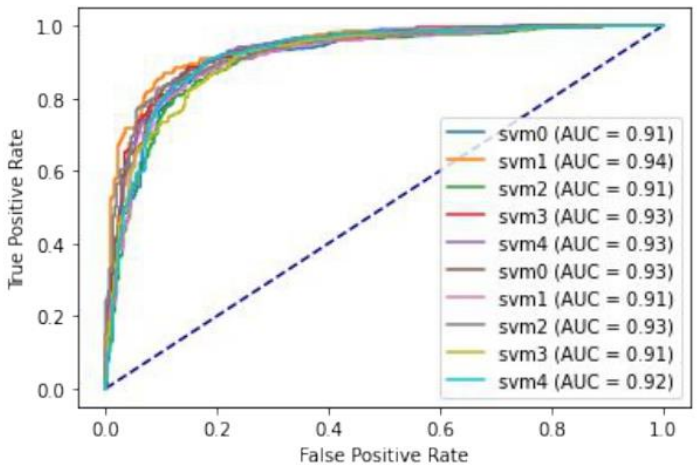
Metriche validazione classificazione

MCC	SVC	Randomforest	Gradientboosting
Desc. mol.	0.64 (A)	0.52	0.63
Fingerprints	0.70 (B)	0.68	0.69

A



B

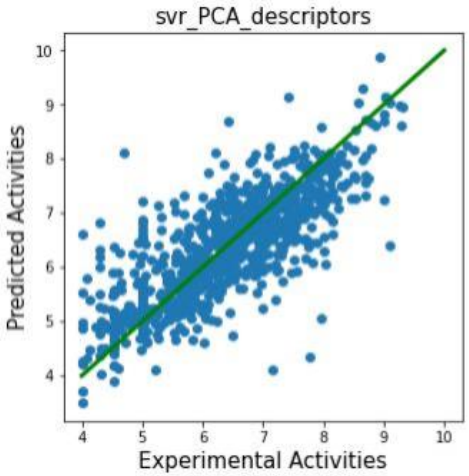


MCC	0.65 (A)	0.71 (B)

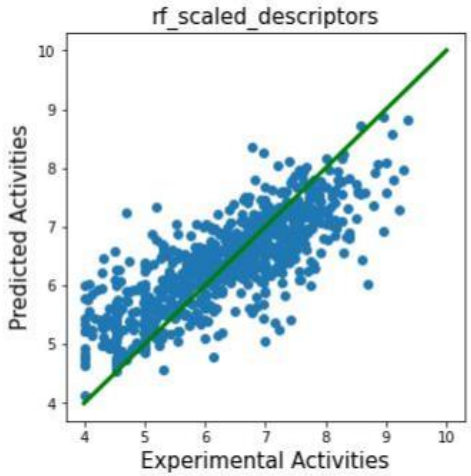
Metriche validazione regressione

R^2_{pred}	SVR	Randomforest
Desc. molecolari	0.59 (A)	0.56
Desc. mol. scalati	0.57	0.64 (B)
Fingerprints	0.73 (C)	0.57

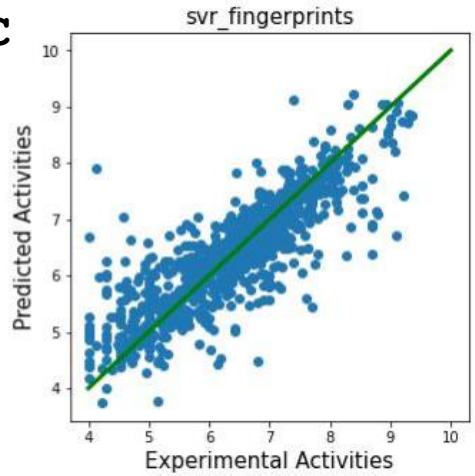
A



B

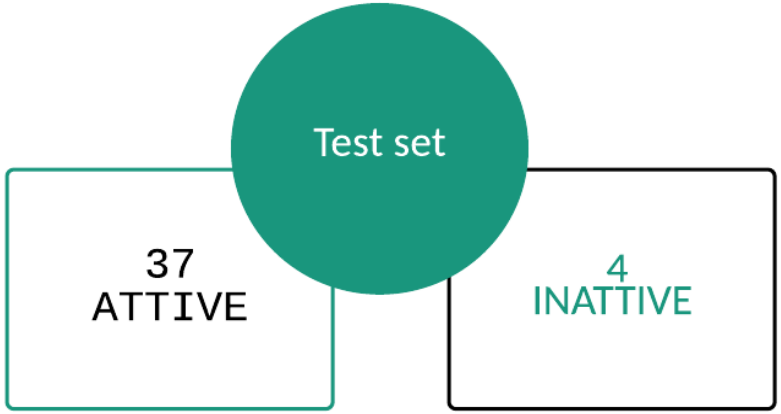


C



Q^2	0.64 (A)	0.65 (B)	0.75 (C)

3.Validazione esterna

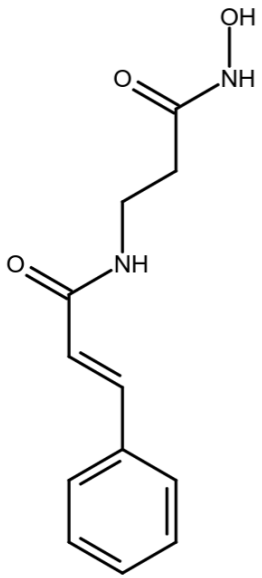


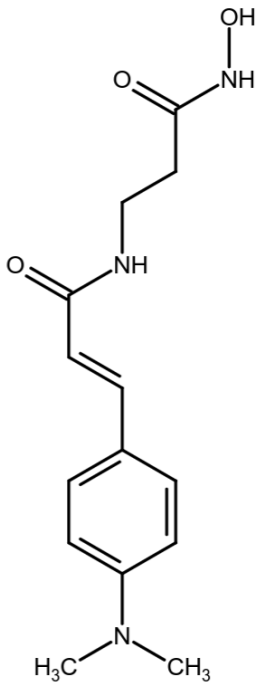
	MCC	R^2
Descrittori mol.	0.68	0.12
Fingerprints	0.88	0.28

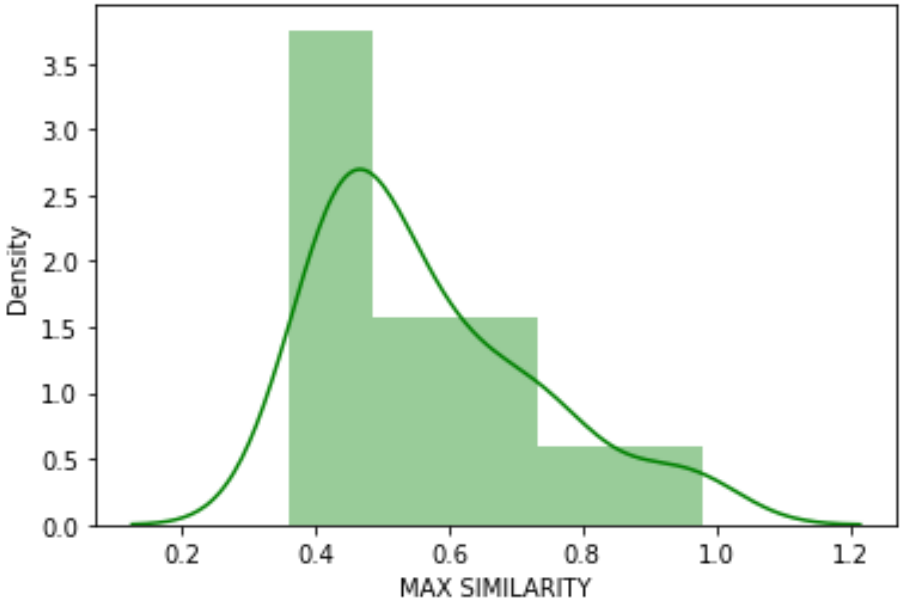
4.Analisi similarità molecolare test-training sets

Test

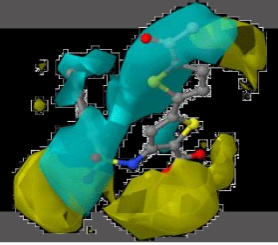
Training



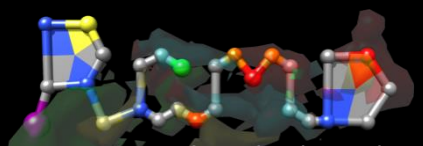




Max similarity = 0.63



5.Selettività HDAC1-HDAC4



GENERAZIONE DATASET E LABELS

pIC_{50}_HDAC1	pIC_{50}_HDAC4	Ratio	SELECTIVITY_HDAC1-4
5.24	6.47	-1.23	0
7.27	5.0	2.27	1
4.90	4.47	0.43	-1

545 molecole totali:

185 selettive HDAC4

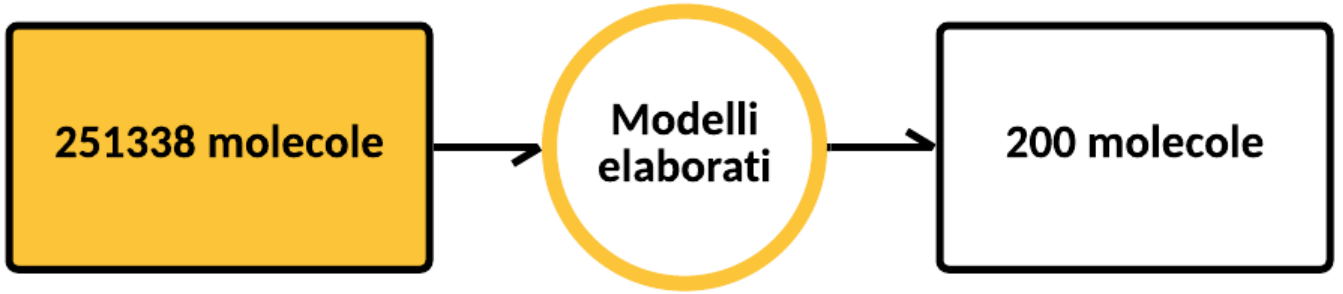
144 selettive HDAC1

216 non selettive

MODELLI DI CLASSIFICAZIONE E REGRESSIONE

	MCC	Q^2
Desc. molecolari	0.89	0.83
Desc. mol. scalati	-	0.83
Fingerprints	0.90	0.85

6.Virtual screening e sviluppi futuri

Virtual Screening	

Sviluppi futuri	Saggi biologici sulle 200 molecole
	Applicazione modelli elaborati a tutte le **HDAC**
	Studio di selettività per tutte le classi di **HDAC**