



Apprendimento automatico e screening ad alto contenuto per lo sviluppo di farmaci

Sašo Džeroski

Istituto Jozef Stefan, Ljubljana, Slovenia

Team: Dragi Kocev,

Panče Panov, Tomaž

Stepišnik, Martin Breskvar

Interreg

ITALIA-SLOVENIJA



TRAIN

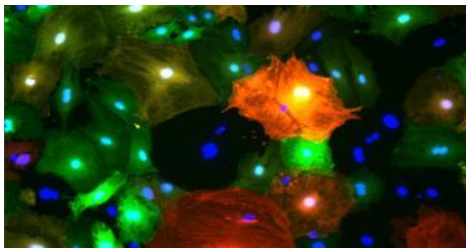


UNIONE EUROPEA
EVROPSKA UNIJA

Progetto standard co-finanziato dal Fondo europeo di sviluppo regionale
Standardni projekt sofinancira Evropski sklad za regionalni razvoj

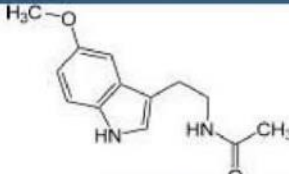
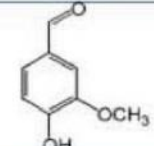
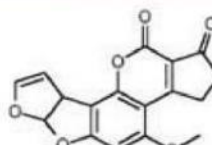
Screening ad alto contenuto

- Schermi composti
- Schermi genomici (schermi miRNA)
- Immagini prese al microscopio
- Usa l'apprendimento automatico per apprendere modelli di screening composto virtuale, identificare nuovi candidati



Apprendimento da schermate composte

Struttura-
attività di apprendimento
Le relazioni

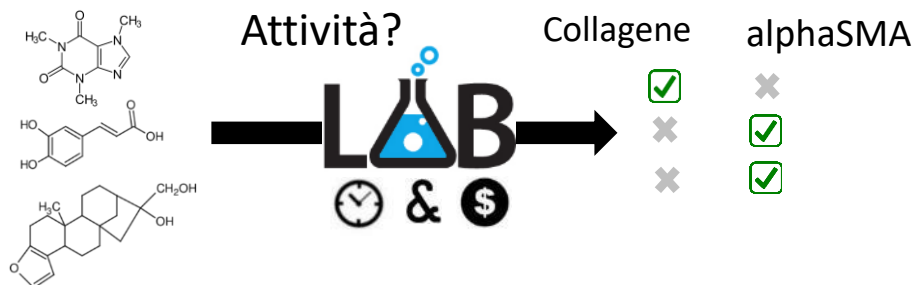
		input datatype specification	output datatype specification
		input: molecule datatype	output: real datatype
		compound	activity
data example			0.25
			0.28
			0.37



Imparare da schermi composti

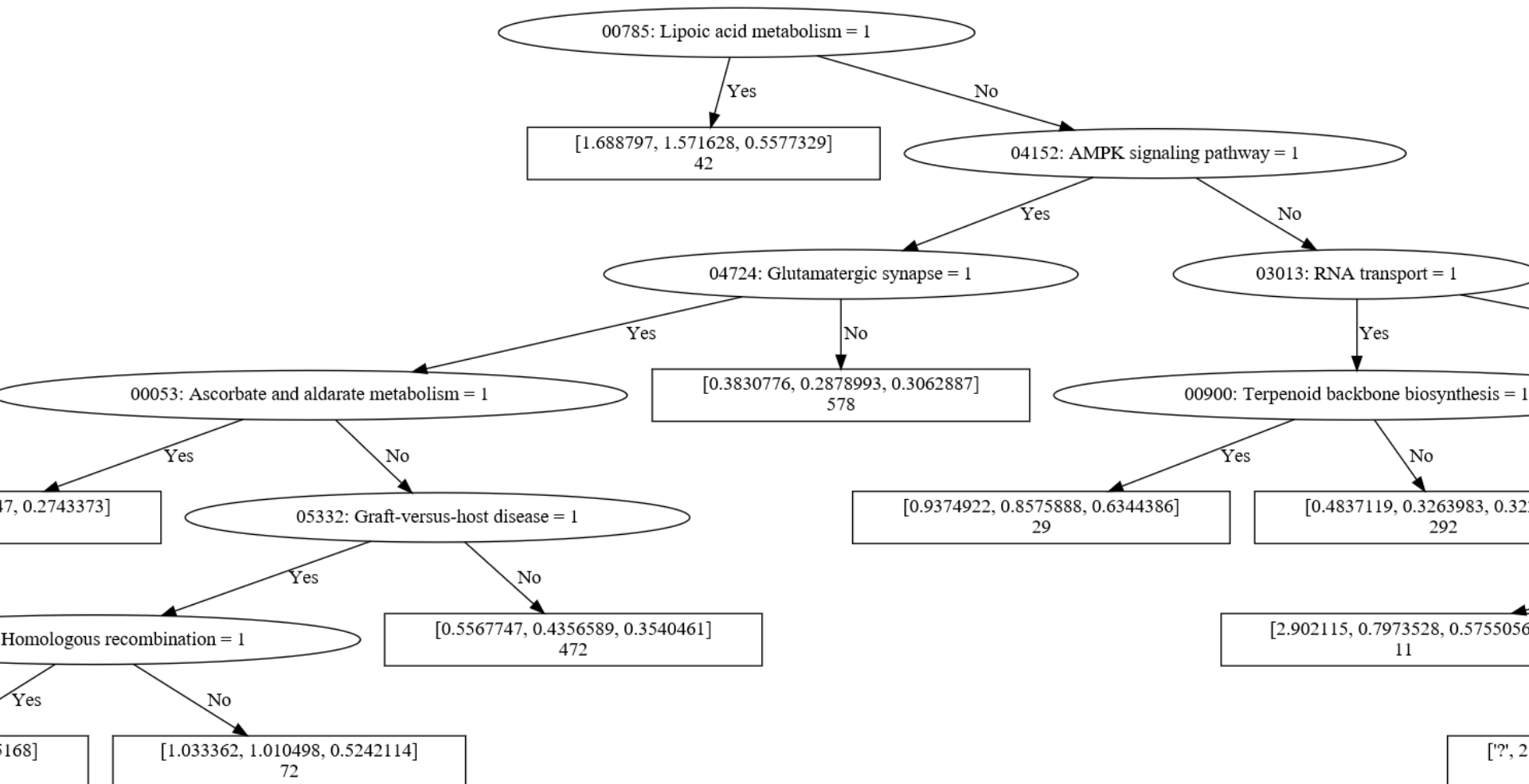
Dati sullo schermo composto

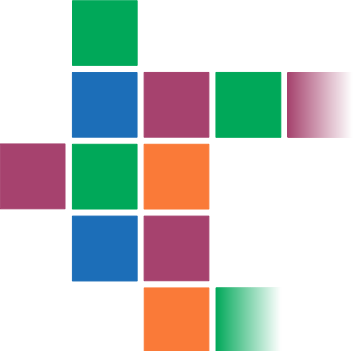
	Spazio descrittivo				Spazio target	
Esempio 1	1	VERO	0.49	0.69	0.68	3.91
Esempio 2	2	FALSO	0.08	0.07	0.56	7.59
Esempio 3	1	FALSO	0.08	0.07	0.10	7.57
Esempio 4	2	VERO	0.49	0.69	0.08	8.86
...





La tecnologia ML: alberi di regressione multi-target





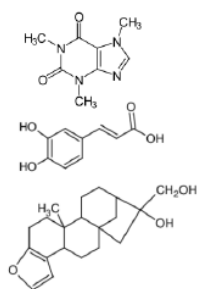
Screening virtuale

Labeled data

	Descriptive space				Target space	
Example 1	1	TRUE	0.49	0.69	0.68	3.91
Example 2	2	FALSE	0.08	0.07	0.56	7.59
Example 3	1	FALSE	0.08	0.07	0.10	7.57
Example 4	2	TRUE	0.49	0.69	0.08	8.86
...

Unlabeled data

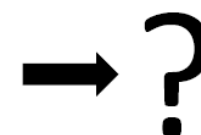
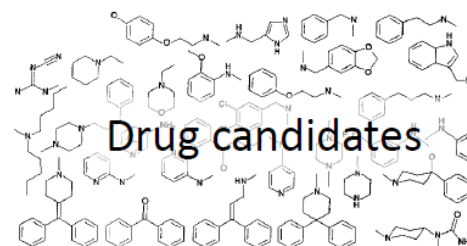
Example N+1	1	TRUE	0.86	0.35	?	?
Example N+2	2	FALSE	0.09	0.05	?	?
Example N+3	4	FALSE	0.07	0.01	?	?
Example N+4	2	TRUE	0.91	0.78	?	?
Example N+5	2	TRUE	0.42	0.69	?	?
...



Learn

MODELLO DI
SCREENING VIRTUALE

Predict





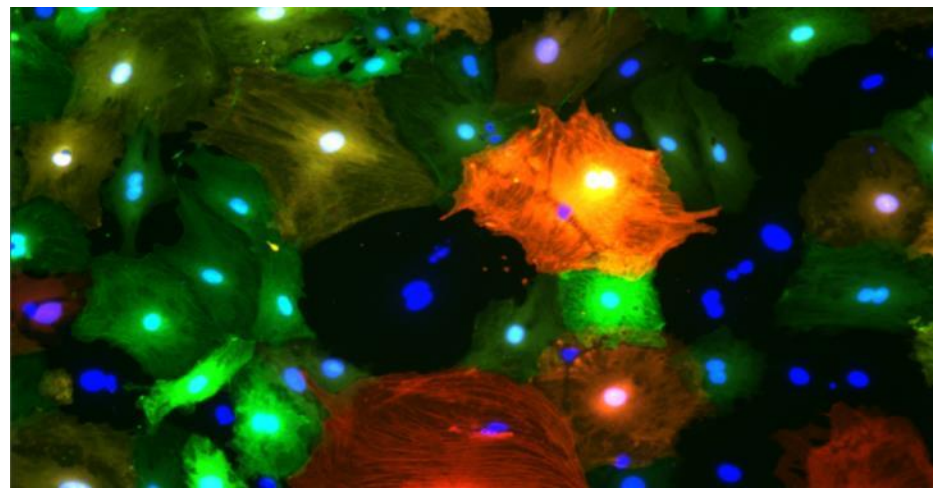
Analisi dei dati da schermate ad alto contenuto

- Composti descritti da impronte digitali
- Generato dalla libreria SW RDkit di chemioinformatica open source
- Sono state utilizzate le impronte digitali FCFP2 (1024 funzioni)
- Considerati anche i profili di proteine mirate
- Questi sono gli attributi
- Saggi fotografati al microscopio
- Funzionalità estratte dalle immagini
- Questi sono quindi gli obiettivi



Riduzione della fibrosi nell'infarto del miocardio

- Schermo ad alto contenuto usando una libreria di 640
- Farmaci approvati dalla FDA (ENZO)
- Identificare i farmaci per ridurre la fibrosi nell'infarto del miocardio
- Lo schermo utilizzava fibroblasti cardiaci murini che si differenziano in miofibroblasti in coltura, esprimendo un aumento di SMA-RFP alfa e collagene-alfa1-EGFP
- Obiettivi: intensità di
 - alphaSMA
 - collagene
- attributi
 - Le impronte digitali



Test delle previsioni

- Alcune conoscenze / vincoli specifici del dominio applicati: composti previsti filtrati per farmaci approvati dalla FDA che non sono corticosteroidi
- Stringhe SMILE utilizzate in Chemmine per identificare sostanze con somiglianza strutturale con composti non commerciali con valori previsti elevati
- Sono stati identificati tre composti correlati che sono descritti in letteratura per avere un effetto anti-fibrotico
- Sono stati identificati quattro composti correlati che non erano stati precedentemente descritti per avere un effetto anti-fibrotico
- Testato in laboratorio umido e uno funziona davvero bene 😊



Esperienze all'interno del progetto

- Analizzati diversi schermi differenti
 - Schermi composti
 - Schermi miRNA
- Abbiamo scoperto che possiamo usare i modelli appresi dagli schermi miRNA per eseguire lo screening composto virtuale
- Eccellente collaborazione con l'ICGEB
- Interessante rappresentazione della realtà virtuale