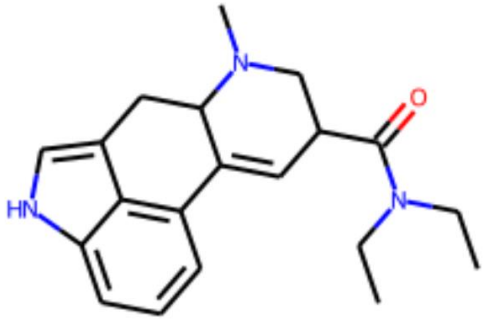
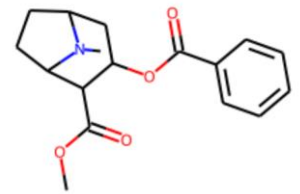
SEMINARIO SU APPLICAZIONI DI RETI CONVOLUZIONALI IN CAMPO MEDICO-FARMACEUTICO



ING. ISABELLA MENDOLIA, PHD
DOTT. SALVATORE CONTINO, PHD
ING. LUCA CRUCIATA

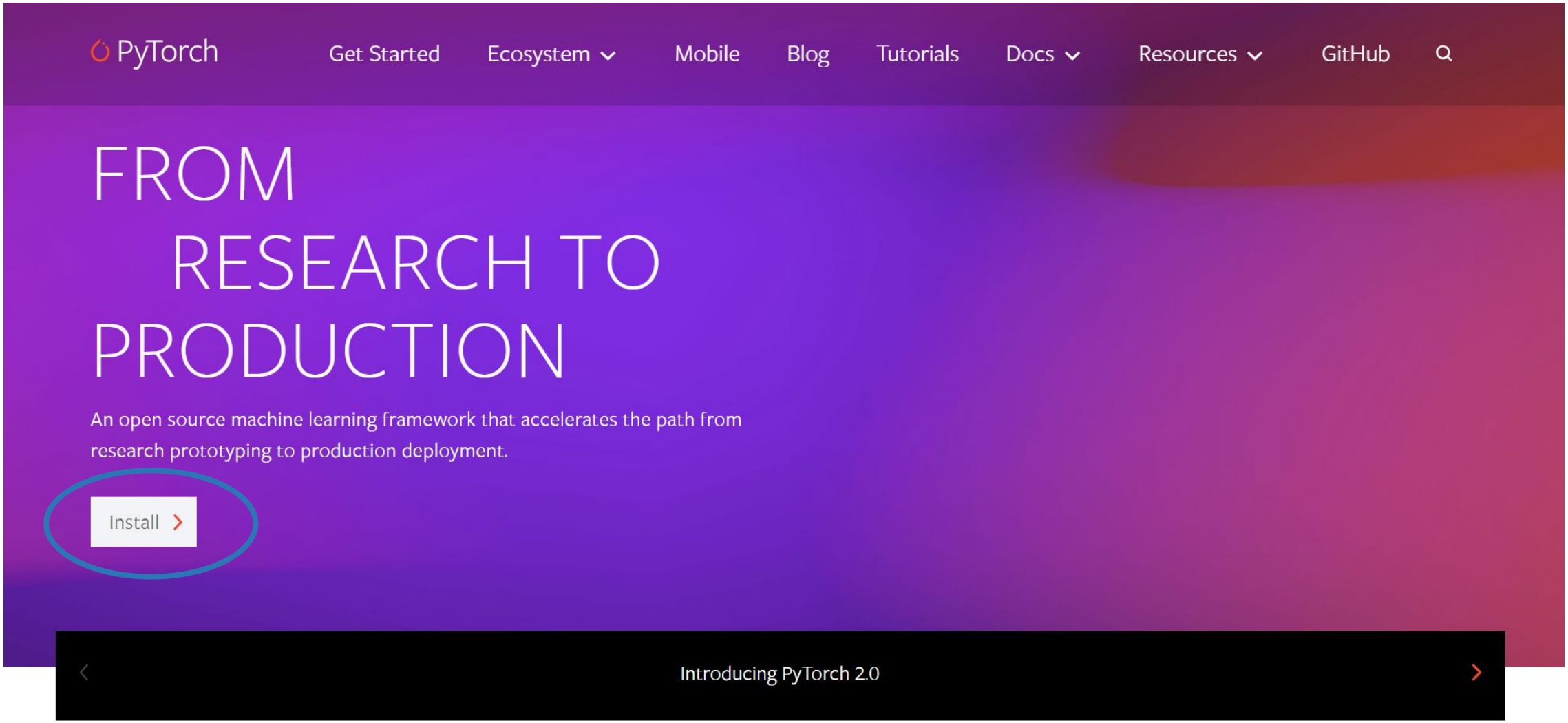


Università degli Studi di Palermo
dipartimento di ingegneria unipa
LABORATORIO DI INTERAZIONE UOMO-MACCHINA
CHILAB

Sommario

- Breve presentazione di PyTorch ed installazione
- Classificazione di molecole con l'utilizzo di fingerprint
- Classificazione di molecole con l'utilizzo di spettri NMR
- Image Segmentation
- Esercitazione guidata su un task

PYTORCH 2.X: FASTER, MORE PYTHONIC AND AS DYNAMIC AS EVER



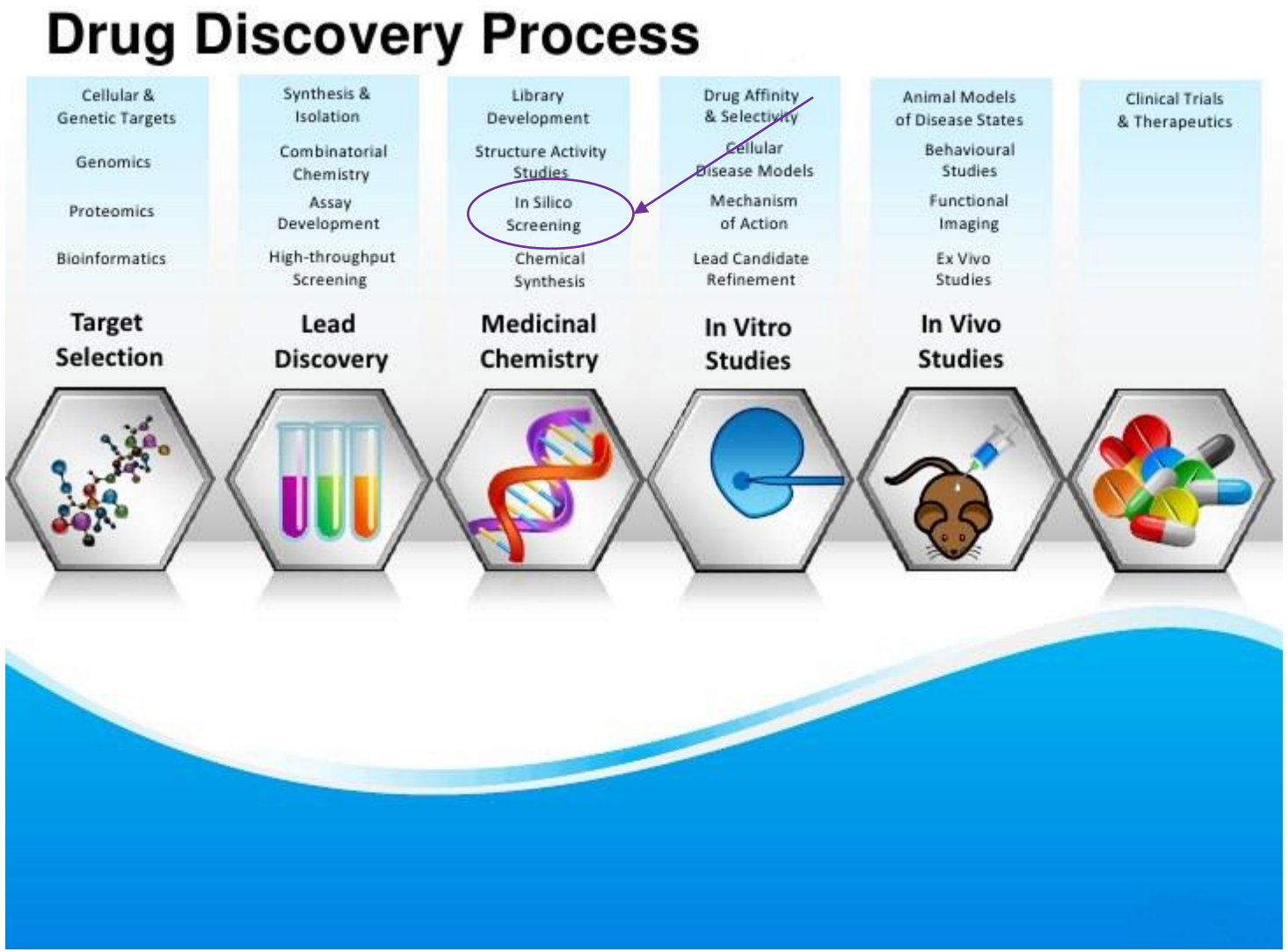
START LOCALLY

Select your preferences and run the install command. Stable represents the most currently tested and supported version of PyTorch. This should be suitable for many users. Preview is available if you want the latest, not fully tested and supported, builds that are generated nightly. Please ensure that you have **met the prerequisites below (e.g., numpy)**, depending on your package manager. Anaconda is our recommended package manager since it installs all dependencies. You can also install previous versions of PyTorch. Note that LibTorch is only available for C++.

PyTorch Build	**Stable (1.13.0)**	Preview (Nightly)		
Your OS	Linux	Mac	**Windows**	
Package	**Conda**	Pip	LibTorch	Source
Language	**Python**	C++ / Java		
Compute Platform	**CUDA 11.6**	CUDA 11.7	~~ROCm 5.2~~	CPU
Run this Command:	`conda install pytorch torchvision torchaudio pytorch-cuda=11.6 -c pytorch -c nvidia`			

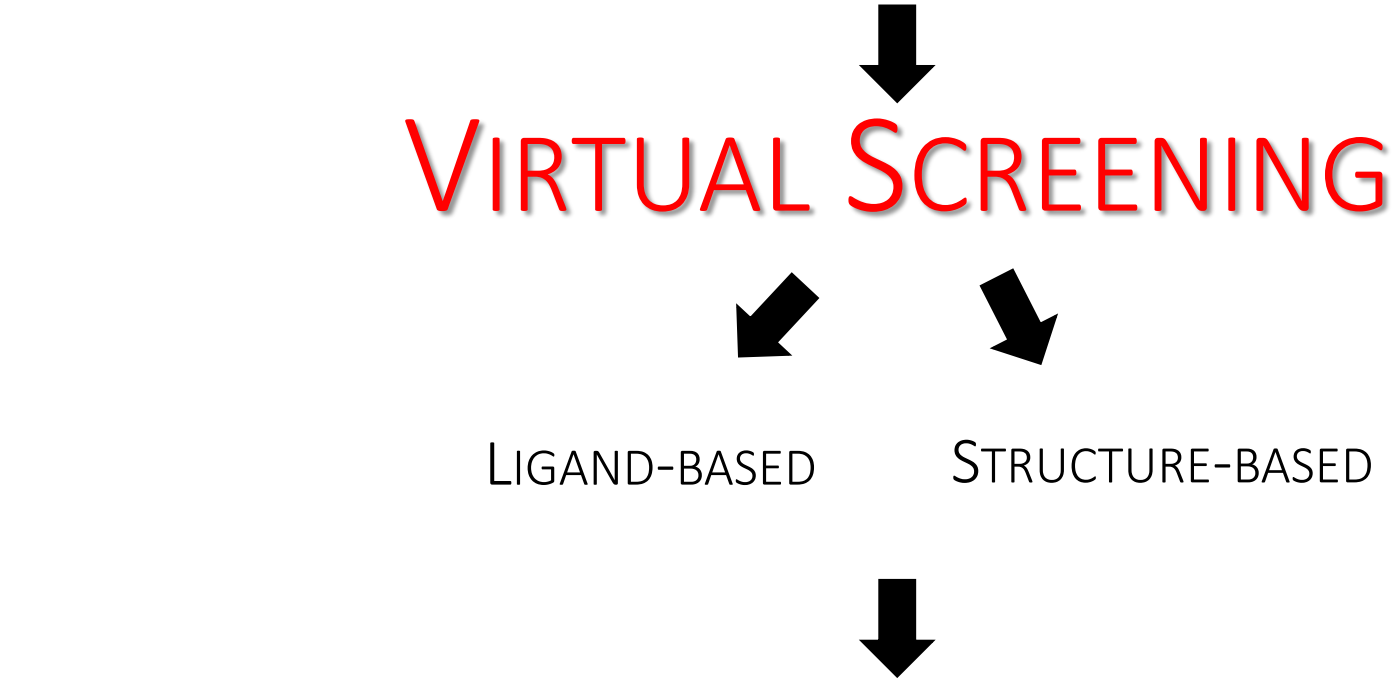
NEL FRATTEMPO CHE PYTORCH SI INSTALLA……

APPLICAZIONI DI RETI CONVOLUZIONALI IN AMBITO FARMACEUTICO



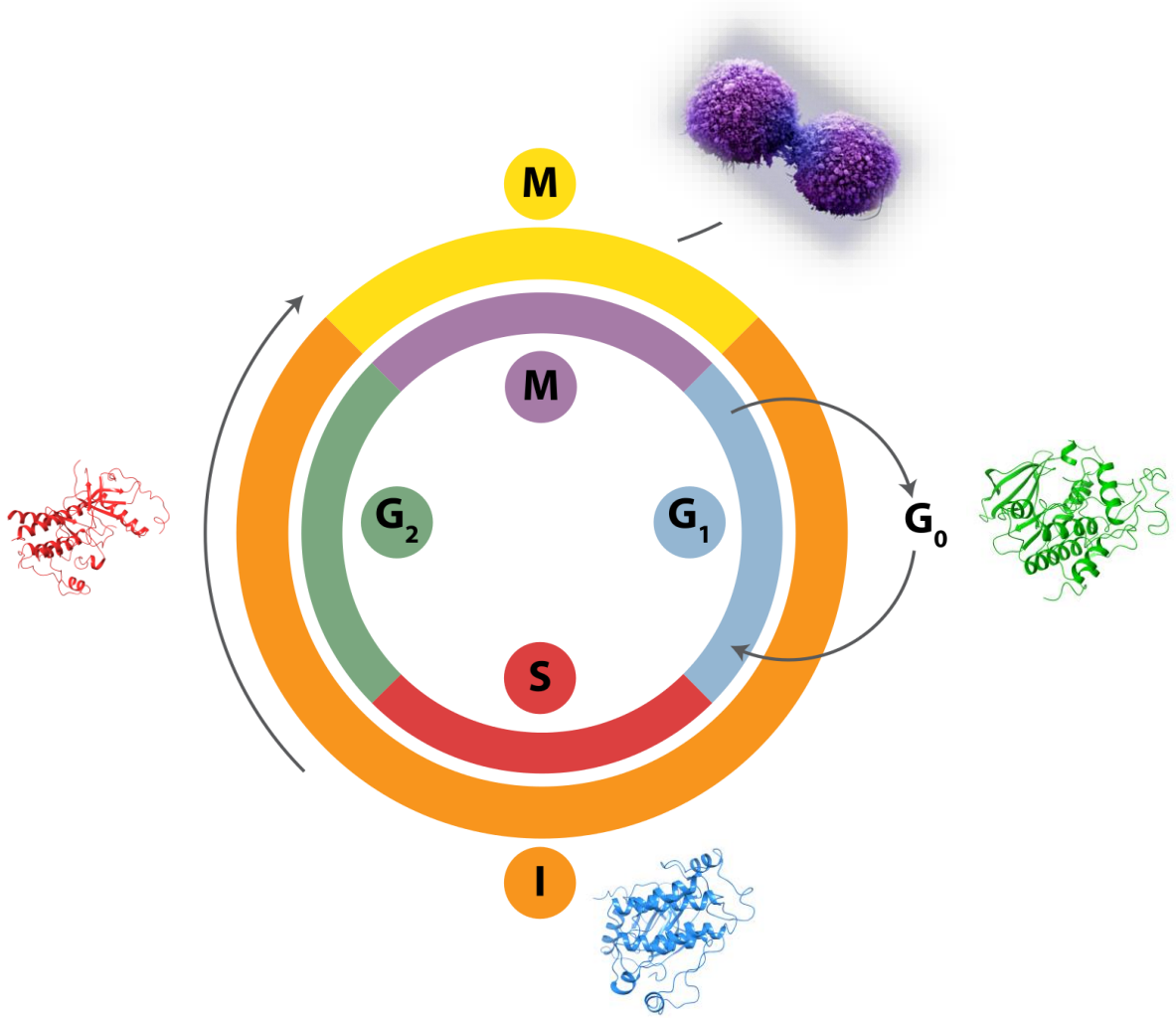
Architetture Deep Per Drug Design

Drug Design Glossario Del Processo



Applicazioni di Deep Learning in Drug Discovery

Selezione Del Target



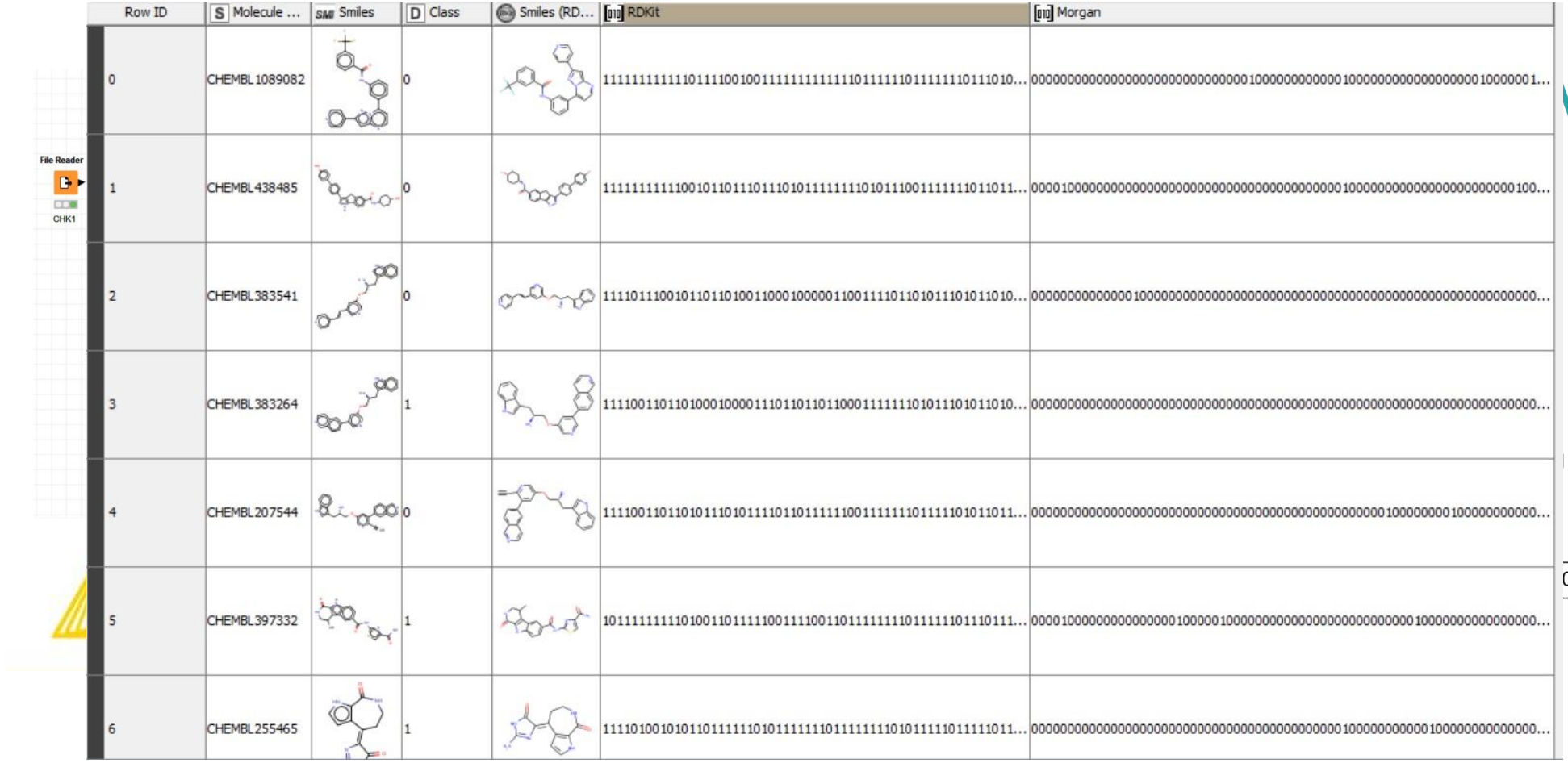
Ciclin-Dipendent Kinase 1 (CDK1)

È il principale regolatore della fase M del ciclo cellulare.

Questa fase se non correttamente regolata induce l'insorgenza di fenotipi cancerosi.

Applicazioni di Deep Learning in Drug Discovery

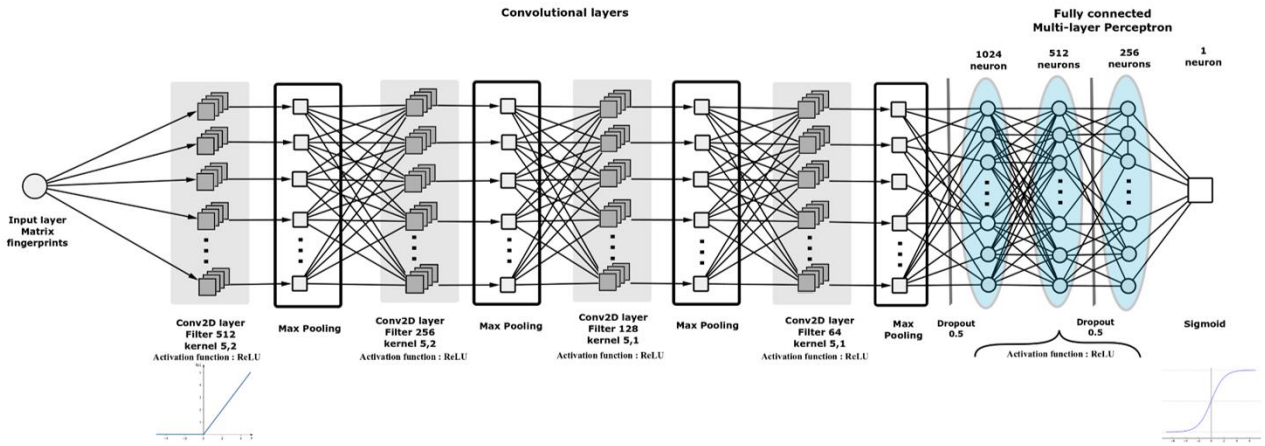
Implementazione Del Dataset



Row ID	Molecule ...	Smiles	Class	Smiles (RD...	RDKit	Morgan
0	CHEMBL1089082		0		1111111111110111100100111111111111101111110111111101110...	0000000000000000000000000000001000000000001000000000000000001000000...
1	CHEMBL438485		0		1111111111100101101110111010111111110101110011111110110...	0000100000000000000000000000000000000000001000000000000000000100...
2	CHEMBL383541		0		1111011100101101101001100010000011001111011010111010110...	0000000000000010000000000000000000000000000000000000000000000000...
3	CHEMBL383264		1		1111001101101000100001110110110110001111111010111010110...	0000000000000000000000000000000000000000000000000000000000000000...
4	CHEMBL207544		0		1111001101101011101011110110111111001111111011111010110...	0000000000000000000000000000000000000000000000000000000010000000...
5	CHEMBL397332		1		1011111111101001101111100111100110111111110111111011101...	0000100000000000000010000010000000000000000000000000000010000000...
6	CHEMBL255465		1		1111010010101101111110101111111011111111010111101111101...	0000000000000000000000000000000000000000001000000000001000000000...

Applicazioni di Deep Learning in Drug Discovery

Esperimenti con CNN 2D



Fingerprints	Accuracy	Loss	F1-score	AUC
M,L	0.92	0.56	0.88	0.9563
R,M,A	0.900	0.6800	0.8600	0.9527
M,A,L,F	0.9200	0.6000	0.8877	0.9444
R,M,A,L,F	0.9163	0.6082	0.8820	0.9513
R,M,A,T,L,F	0.8945	0.6280	0.8557	0.9494

Lenght 1024 bit

Fingerprints	Accuracy	Loss	F1-score	AUC
M,F	0.8981	0.4463	0.8679	0.9555
M,T,L	0.9345	0.3900	0.9117	0.9685
R,M,T,F	0.9418	0.4268	0.9001	0.94
R,A,T,L,F	0.9127	0.4052	0.8867	0.963
R,M,A,T,L,F	0.9236	0.3950	0.9004	0.9774

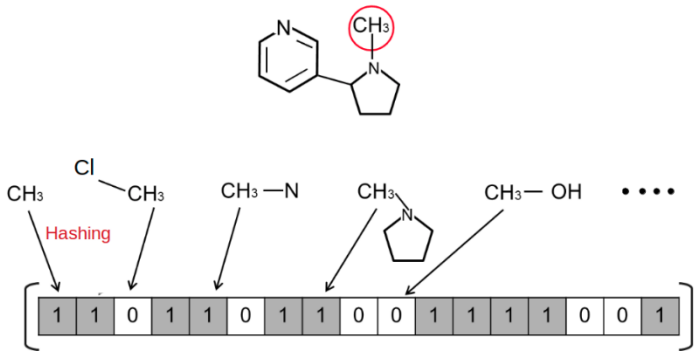
Lenght 512 bit

Fingerprints	Accuracy	Loss	F1-score	AUC
L,F	0.909	0.4087	0.8792	0.9655
R,L,F	0.9127	0.4734	0.8846	0.9606
R,A,L,F	0.9054	0.4914	0.8749	0.9572
R,M,T,L,F	0.8909	0.538	0.8623	0.9624
R,M,A,T,L,F	0.8981	0.5982	0.8679	0.9537

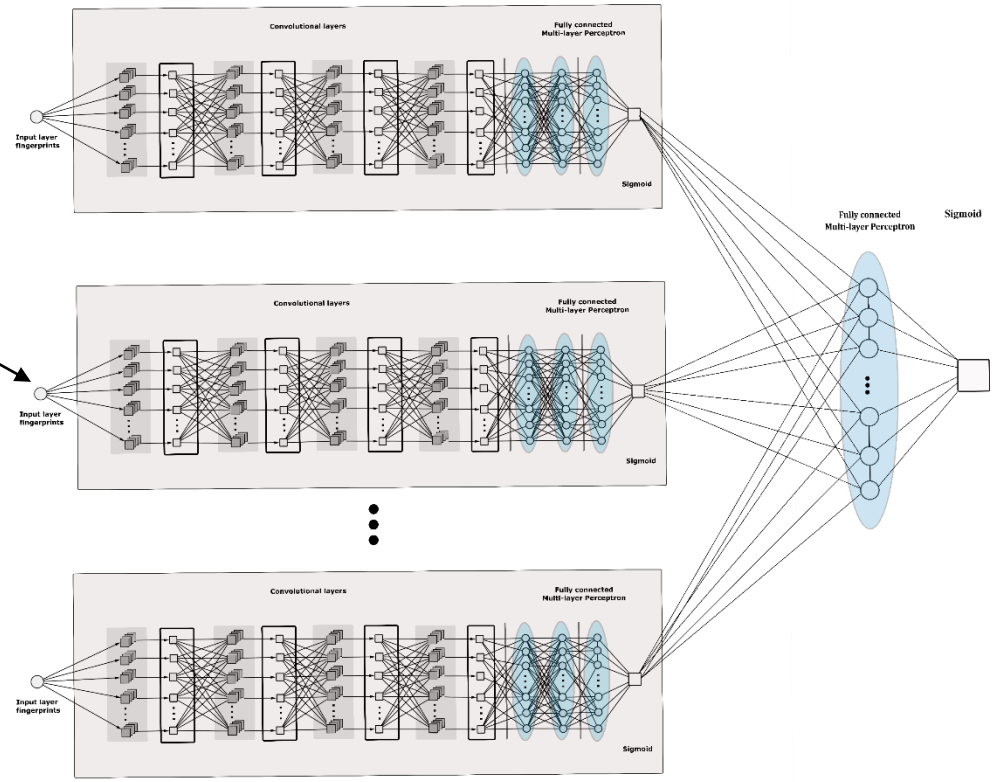
Lenght 256 bit

Applicazioni di Deep Learning in Drug Discovery

Definizione Dell'Architettura Tuned-MLP-Out



1. RDKit
2. Morgan
3. AtomPair
4. Torsion
5. FeatMorgan
6. Layered
7. ECFP4



Applicazioni di Deep Learning in Drug Discovery

Risultati Ottenuti

Table 1 Results for the active/inactive discrimination task, and Training scheme 1

From: Convolutional architectures for virtual screening

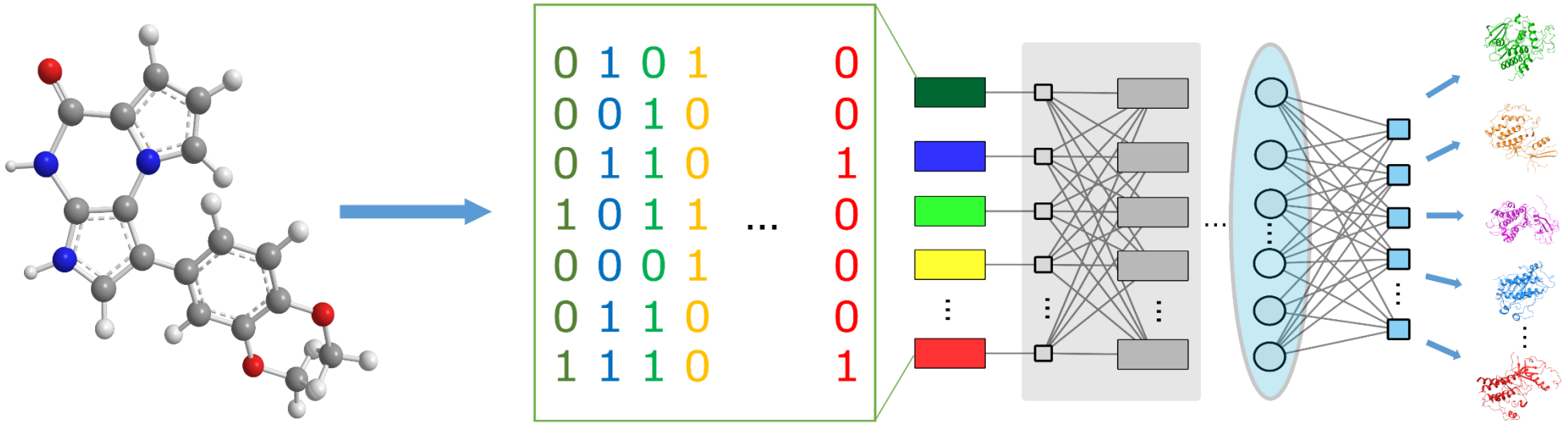
Architecture	Bal. accuracy	Sensitivity	Loss	AUC	F1-score	MCC
Tuned-MLP-Out	0.9880	0.9855	0.0405	0.9979	0.9510	0.9462
Voting	0.9768	0.9710	0.2093	0.9920	0.8965	0.9033
CNN 1D (F)	0.9687	0.9710	0.0688	0.9904	0.8979	0.8813
CNN 2D (R-M-F)	0.9679	0.9565	0.0770	0.9912	0.8918	0.8817
Random Forest (F)	0.9510	0.8985	0.6405	0.9837	0.6065	0.8962
SVM (F)	0.9421	0.8985	0.7883	0.9868	0.8857	0.8731

Fingerprint types: *(R)DKIT,(M) organ, (F) eatMorgan, (L)ayered*

Mendolia, I., Contino, S., Perricone, U., Ardizzone E., Pirrone, R. Convolutional architectures for virtual screening. BMC Bioinformatics 21, 310 (2020). https://doi.org/10.1186/s12859-020-03645-9
Mendolia, I., Contino, S., Perricone, U., Ardizzone E., Pirrone, R. A Convolutional Neural Network for Virtual Screening of Molecular Fingerprints. Conference paper. ICIAP Pages 399-409 (2019).

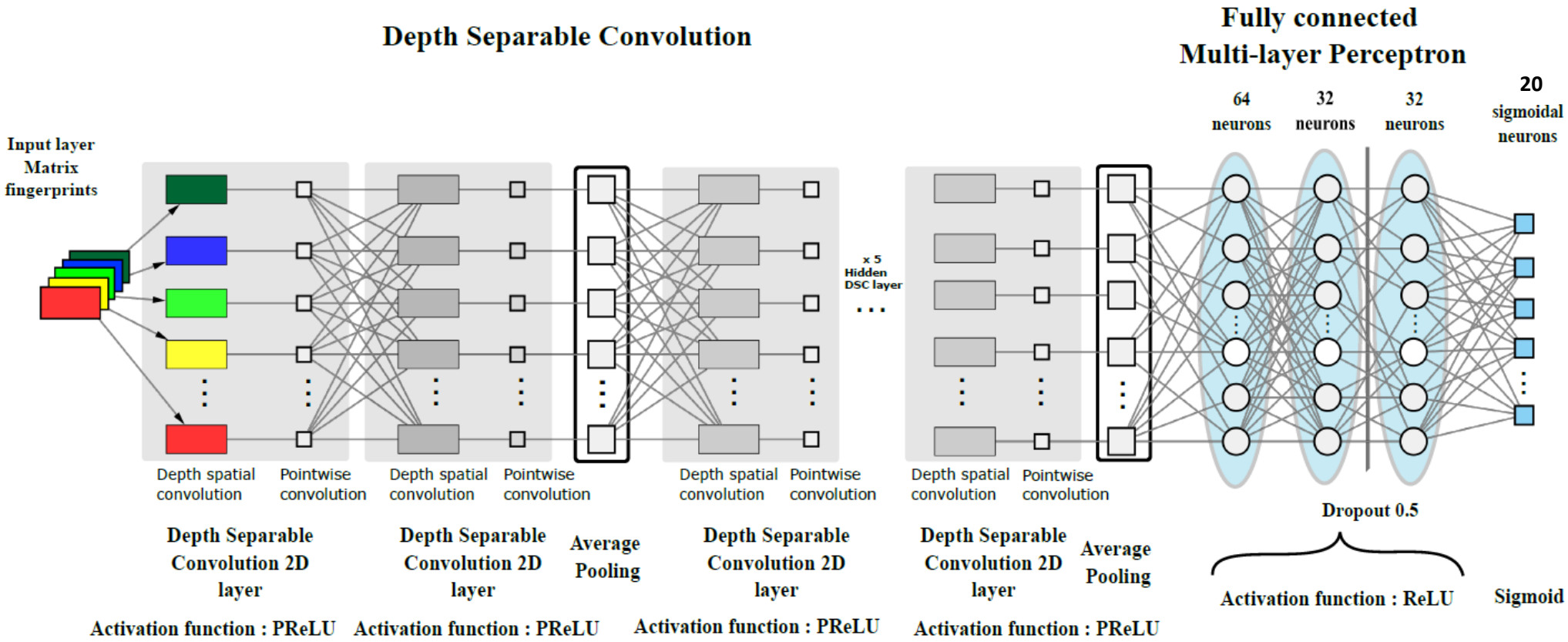
Applicazioni di Deep Learning in Drug Discovery

EMBER-EMBedding multiplE molecular fingErprint



Applicazioni di Deep Learning in Drug Discovery

Architettura proposta



Applicazioni di Deep Learning in Drug Discovery

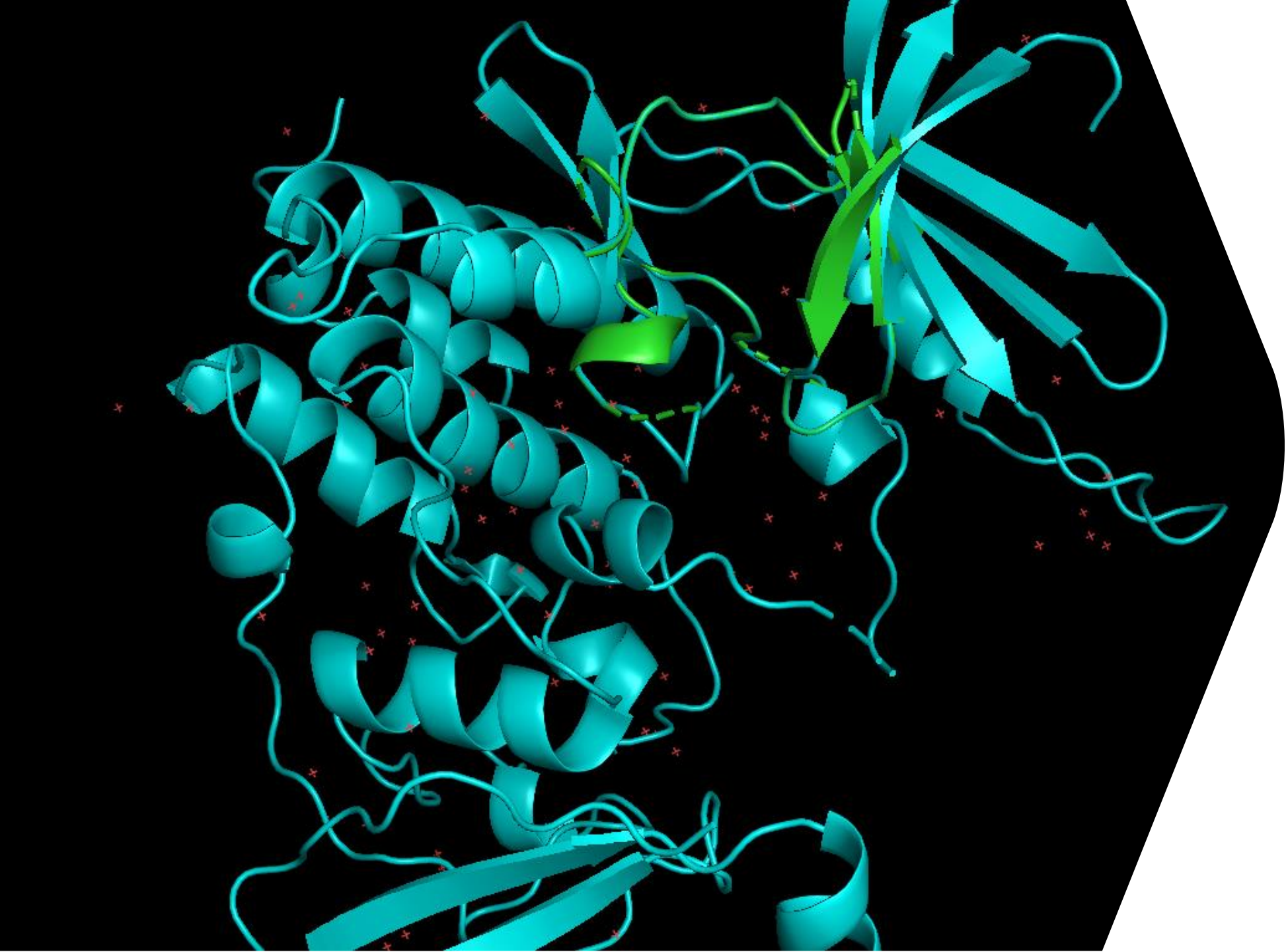
Risultati

Target	Acc.	Loss	Sensitivity	MCC	AUC	F1-score
ACK	0.9957	0.0226	0.5000	0.6742	0.9834	0.6463
ALK	0.9930	0.0402	0.6575	0.7913	0.9904	0.7804
CDK1	0.9910	0.0314	0.4537	0.6397	0.9850	0.6059
CDK2	0.9859	0.0431	0.5281	0.6338	0.9845	0.6287
CDK6	0.9966	0.0210	0.5865	0.7523	0.9895	0.7305
INSR	0.9893	0.0329	0.3779	0.5830	0.9858	0.5342
ITK	0.9945	0.0232	0.5886	0.7302	0.9905	0.7154
JAK2	0.9898	0.0472	0.8474	0.9090	0.9950	0.9114
JNK3	0.9967	0.0154	0.5905	0.7610	0.9901	0.7381
MELK	0.9957	0.0229	0.7081	0.8270	0.9897	0.8188
CHK1	0.9895	0.0512	0.6385	0.7650	0.9846	0.7565
CK2A1	0.9942	0.0253	0.5166	0.6944	0.9857	0.6667
CLK2	0.9936	0.0259	0.2255	0.4137	0.9771	0.3485
DYRK1A	0.9916	0.0321	0.4080	0.5987	0.9776	0.5591
EGFR	0.9845	0.0604	0.7536	0.8331	0.9874	0.8357
ERK2	0.9881	0.0563	0.7295	0.8292	0.9886	0.8272
GSK3	0.9843	0.0554	0.5827	0.6892	0.9762	0.6856
IRAK4	0.9936	0.0287	0.7611	0.8611	0.9938	0.8571
MAP2K1	0.9931	0.0319	0.5497	0.7184	0.9795	0.6954
PDK1	0.9945	0.0271	0.6310	0.7757	0.9875	0.7613

Target	TP/P 1%*	TP/P 2%*	TP/P 5%*	TP/P 10%*	EF 1%	EF 2%	EF 5%	EF 10%
ACK	72/106	84/106	95/106	101/106	68	40	18	10
ALK	131/254	202/254	229/254	247/254	52	40	18	10
CDK1	111/205	150/205	189/205	196/205	54	37	18	10
CDK2	118/303	194/303	264/303	289/303	39	32	17	10
CDK6	79/104	90/104	98/104	101/104	76	43	19	10
INSR	110/217	145/217	195/217	206/217	51	33	18	9
ITK	107/158	125/158	148/158	155/158	68	40	19	10
JAK2	134/832	268/832	669/832	804/832	16	16	16	10
JNK3	81/105	88/105	95/105	102/105	77	42	18	10
MELK	130/185	157/185	178/185	181/185	70	42	19	10
CHK1	134/343	233/343	300/343	324/343	39	34	17	9
CK2A1	100/151	117/151	141/151	146/151	66	39	19	10
CLK2	59/102	73/102	87/102	96/102	58	36	17	9
DYRK1A	97/174	126/174	152/174	162/174	56	36	17	9
EGFR	134/702	268/702	586/702	664/702	19	19	17	9
ERK2	133/525	267/525	471/525	505/525	25	25	18	10
GSK3	132/393	226/393	327/393	353/393	34	29	17	9
IRAK4	134/339	263/339	320/339	333/339	40	39	19	10
MAP2K1	118/191	142/191	167/191	178/191	62	37	17	9
PDK1	123/187	149/187	170/187	181/187	66	40	18	10

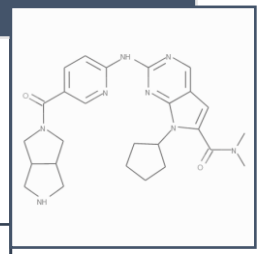
$$EF = \frac{N_{\text{experimental}}{}^{x\%}}{N_{\text{expected}}{}^{x\%}} = \frac{N_{\text{experimental}}{}^{x\%}}{N_{\text{active}} \cdot x\%}$$

Mendolia, I., Contino, S.,De Simone, G., Perricone, U., Pirrone, R. EMBER - embedding multiple molecular fingerprints for virtual screening. Under review on International Journal of Molecular Sciences



Descrittori Molecolari

Grafi molecolari



Canonical Smiles

CN(C)C(=O)c1cc2cnc(Nc3ccc(C(=O)N4CC5CNCC5C4)cn3)nc2n1C1CCCC1

InChI

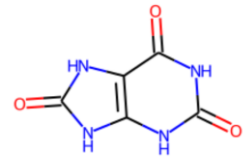
InChI=1S/C26H32N8O2/c1-32(2)25(36)21-9-17-13-29-26(31-23(17)34(21)20-5-3-4-6-20)30-22-8-7-16(12-28-22)24(35)33-14-18-10-27-11-19(18)15-33/h7-9,12-13,18-20,27H,3-6,10-11,14-15H2,1-2H3,(H,28,29,30,31)

Molecular Fingeprints

10101110111011011111011111…1101110111111111111111011111

InChI Key

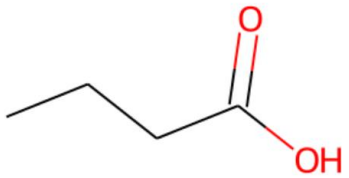
GOBAXMJYGDOAFA-UHFFFAOYSA-N

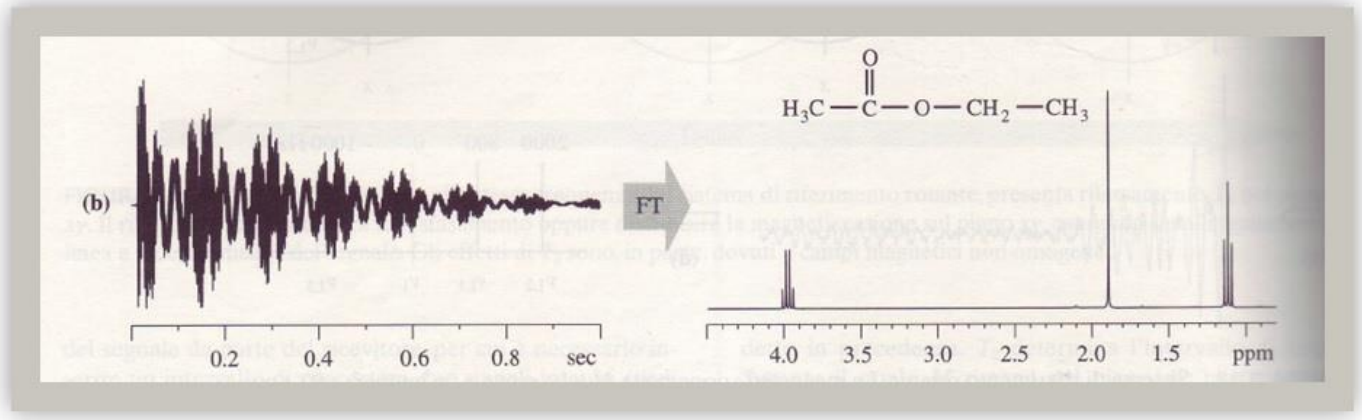


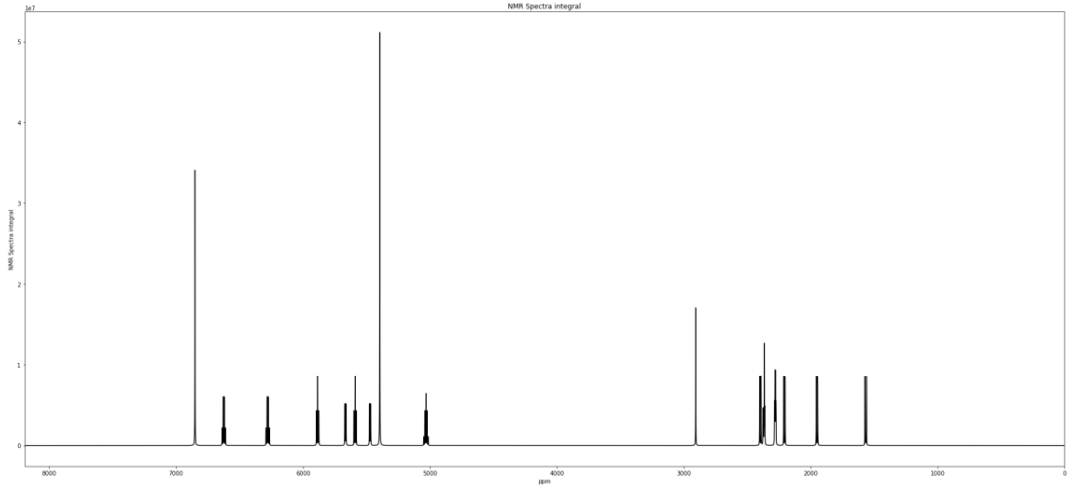
NMR-LIKE

Risonanza Magnetica Nucleare degli Atomi di Idrogeno

H-NMR





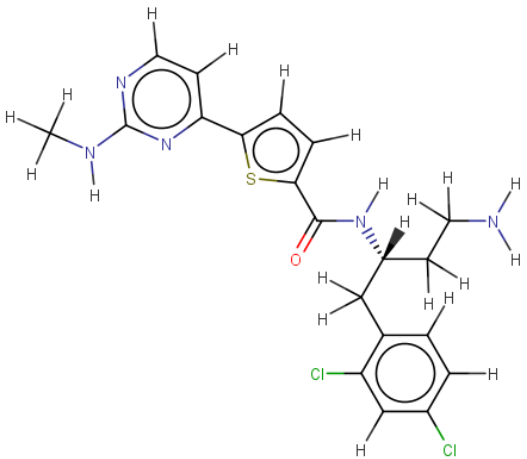


DESCRITTORE NMR-LIKE

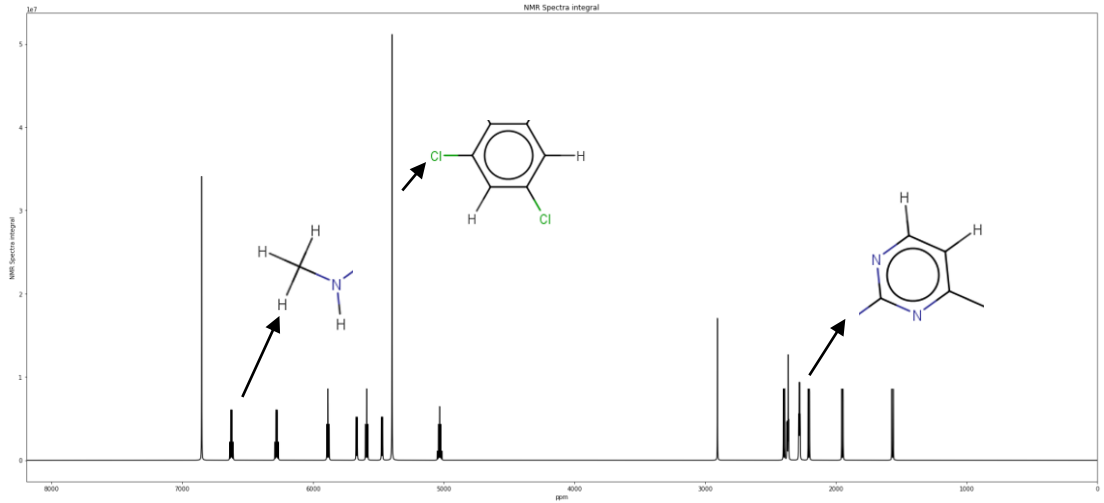
Canonical SMILES

Cc1nc2nc(c3ccc(CN4CCC(CC4)c5nc6cccnc6[nH]5)cc3)c(cn2n1)c7ccccc7

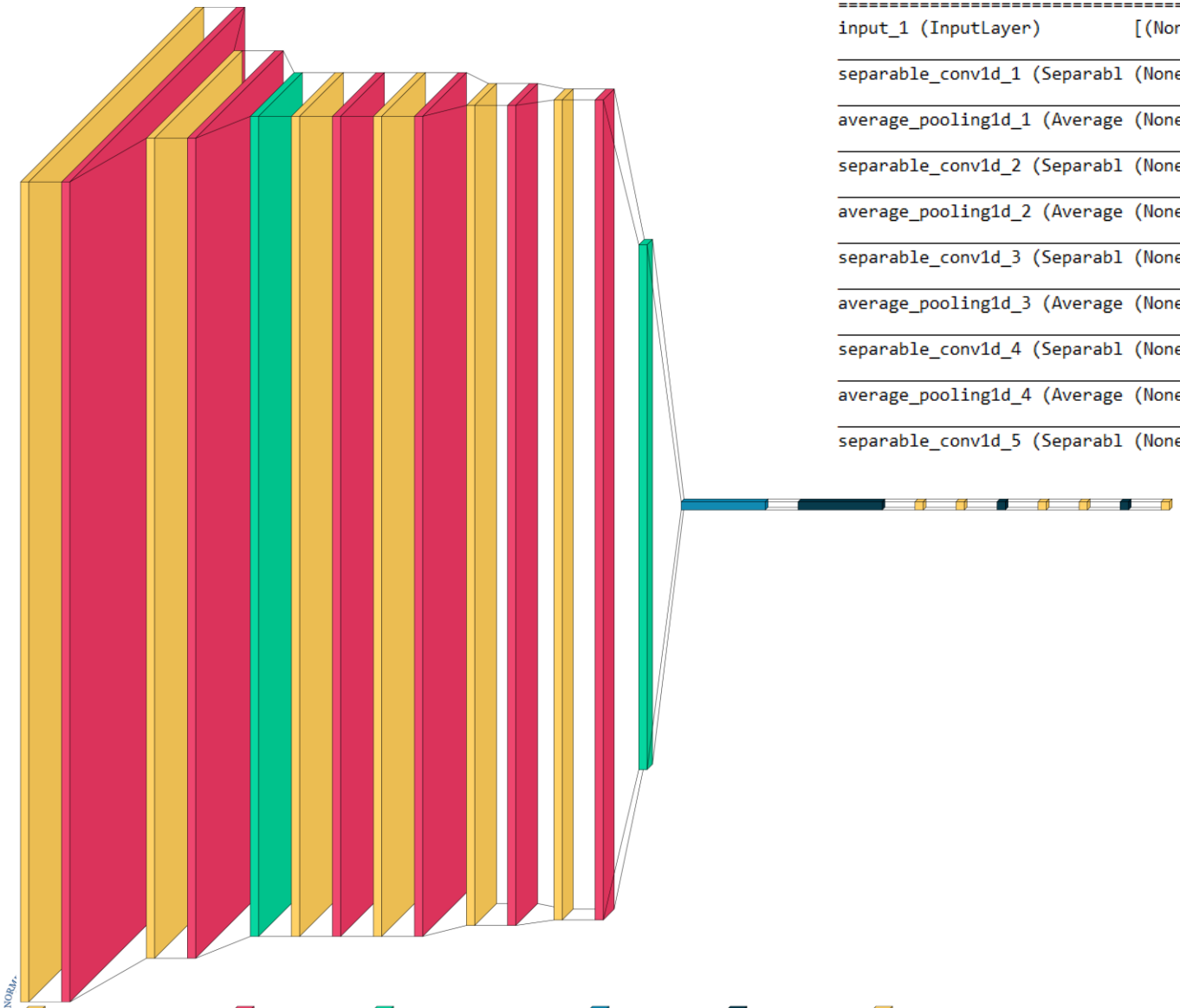
Molecular Graph



H-NMR Spectrum. (JDX Camp, Vettore da 8192)



DESCRITTORE NMR-LIKE

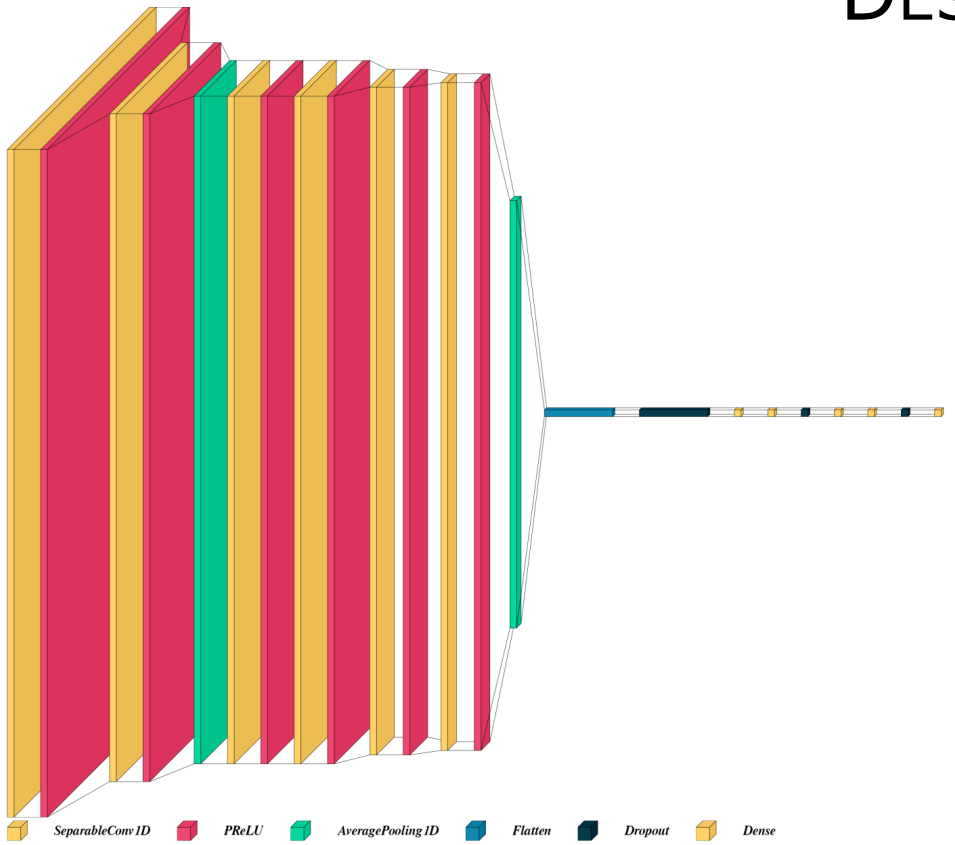


Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 1, 1024)]	0
separable_conv1d_1 (Separabl	(None, 256, 1022)	515
average_pooling1d_1 (Average	(None, 128, 1022)	0
separable_conv1d_2 (Separabl	(None, 128, 1020)	16896
average_pooling1d_2 (Average	(None, 64, 1020)	0
separable_conv1d_3 (Separabl	(None, 64, 1018)	4352
average_pooling1d_3 (Average	(None, 32, 1018)	0
separable_conv1d_4 (Separabl	(None, 32, 1016)	1152
average_pooling1d_4 (Average	(None, 10, 1016)	0
separable_conv1d_5 (Separabl	(None, 16, 1014)	206
average_pooling1d_5 (Average	(None, 5, 1014)	0
flatten_1 (Flatten)	(None, 5070)	0
dense_1 (Dense)	(None, 64)	324544
dropout_1 (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 128)	8320
dense_3 (Dense)	(None, 128)	16512
dropout_2 (Dropout)	(None, 128)	0
dense_4 (Dense)	(None, 64)	8256
dense_5 (Dense)	(None, 1)	65

Total params: 380,818
Trainable params: 380,818
Non-trainable params: 0

SeparableConv1D *PReLU* *AveragePooling1D* *Flatten* *Dropout* *Dense*

DESCRITTORE NMR-LIKE



Protein	Accuracy	Loss	Sensitivity	MCC	AUC	F1 score
CDK1	0.7846	0.5831	0.8795	0.4570	*0.9058*	0.4916
CLK2	0.8031	0.6842	**0.9254**	0.4542	0.9223	*0.4493*
INSR	0.9284	0.2068	*0.7324*	0.6155	0.9268	0.6500
JAK2	**0.9447**	**0.1833**	0.7848	**0.6933**	**0.9473**	**0.7209**

Protein	Tanimoto similarity	Active	Inactive
CDK1	1	869	6278
JAK2	0.87	5526	55260
INSR	0.86	1423	14230
CLK2	0.82	671	7040

Protein	TP/P 1%	TP/P 2%	TP/P 5%	TP/P 10%	EF 1%	EF 2%	EF 5%	EF 10%
	13/83	27/83	47/83	8	8	7	6	
	10/67	23/67	41/67	9	7	7	6	
	30/142	61/142	98/142	11	11	9	7	
	20/553	277/553	417/553	11	11	10	8	

molecole, [2] 1% = 8 molecole, [3] 1% = 16molecole, [4] 1% = 61 molecole

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 1, 1024)]	0
separable_conv1d_1 (Separabl	(None, 256, 1022)	515
average_pooling1d_1 (Average	(None, 128, 1022)	0
separable_conv1d_2 (Separabl	(None, 128, 1020)	16896
average_pooling1d_2 (Average	(None, 64, 1020)	0
separable_conv1d_3 (Separabl	(None, 64, 1018)	4352
average_pooling1d_3 (Average	(None, 32, 1018)	0
separable_conv1d_4 (Separabl	(None, 32, 1016)	1152
average_pooling1d_4 (Average	(None, 10, 1016)	0
separable_conv1d_5 (Separabl	(None, 16, 1014)	206
average_pooling1d_5 (Average	(None, 5, 1014)	0
flatten_1 (Flatten)	(None, 5070)	0
dense_1 (Dense)	(None, 64)	324544
dropout_1 (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 128)	8320
dense_3 (Dense)	(None, 128)	16512
dropout_2 (Dropout)	(None, 128)	0
dense_4 (Dense)	(None, 64)	8256
dense_5 (Dense)	(None, 1)	65

Total params: 380,818
Trainable params: 380,818
Non-trainable params: 0

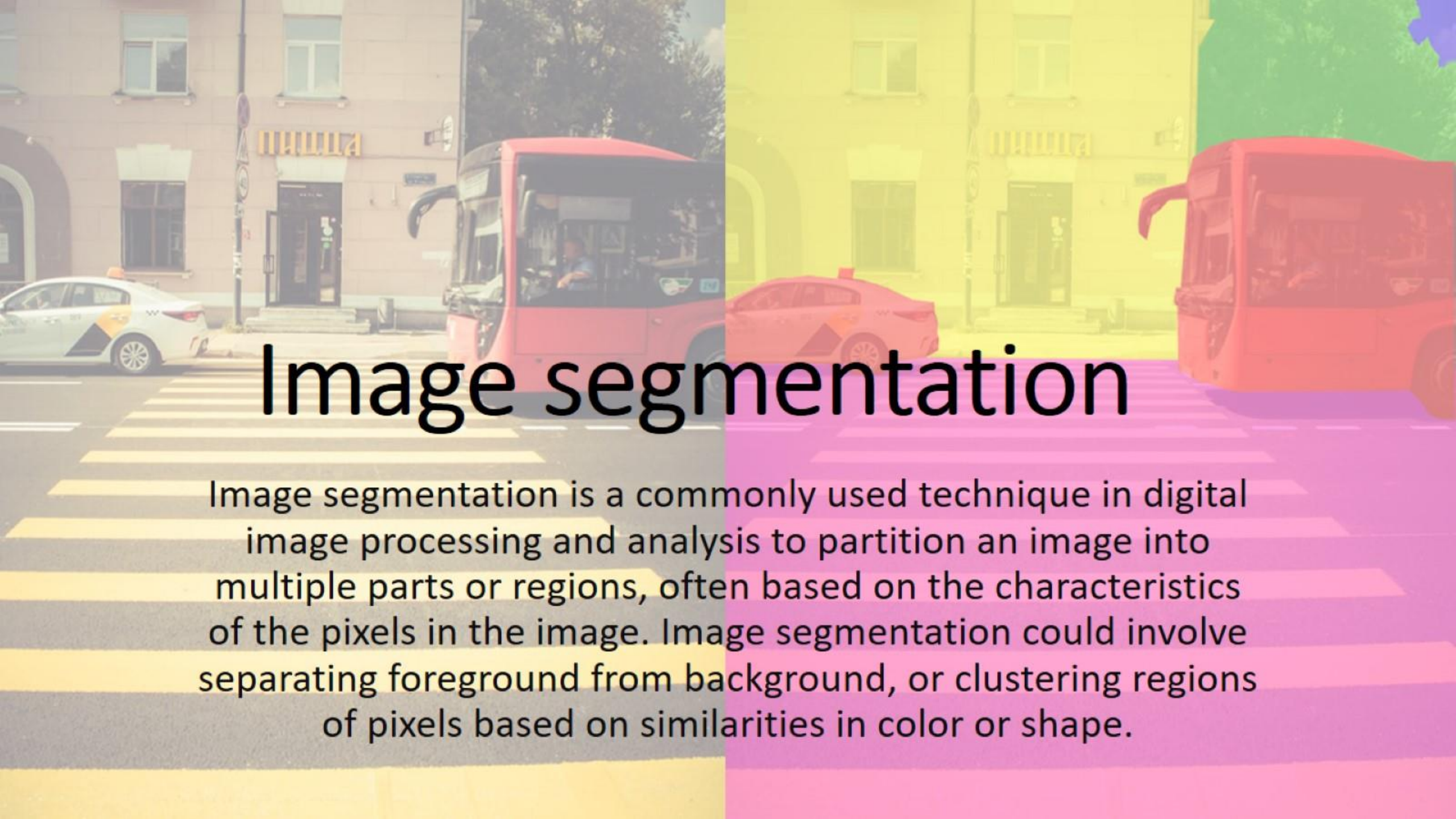
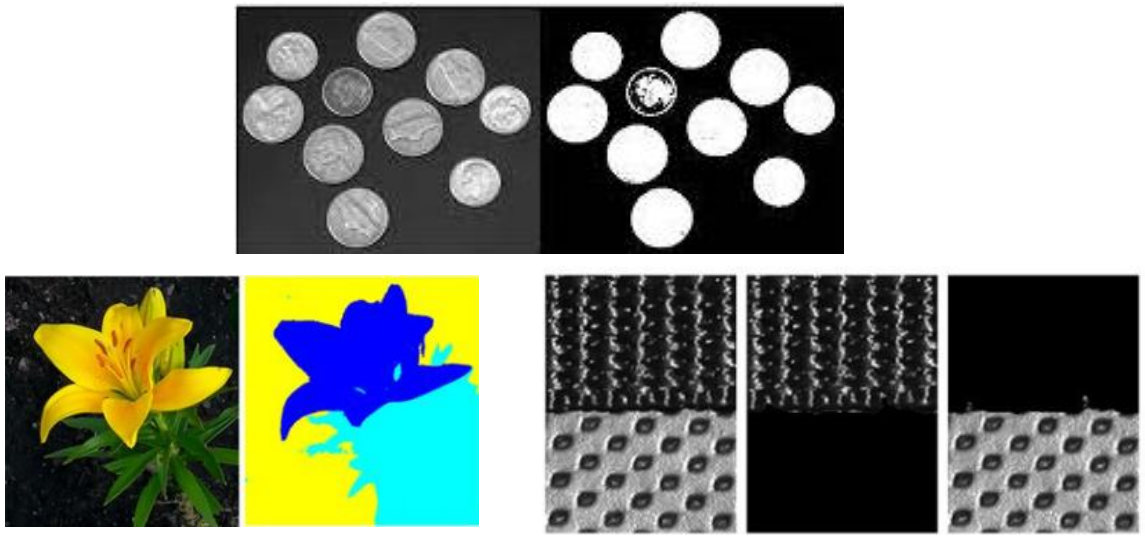


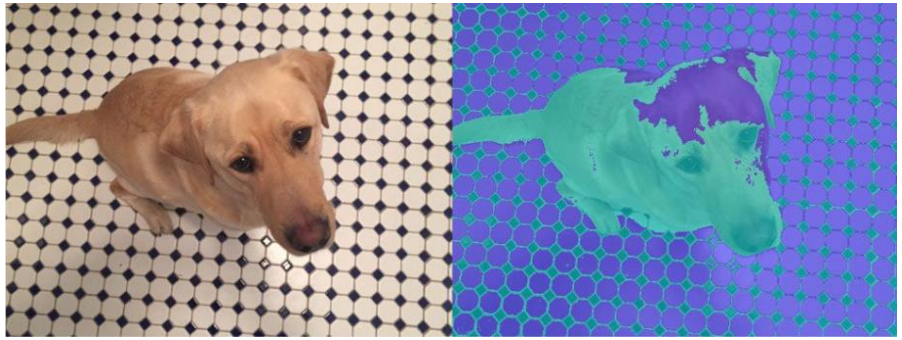
Image segmentation

Image segmentation is a commonly used technique in digital image processing and analysis to partition an image into multiple parts or regions, often based on the characteristics of the pixels in the image. Image segmentation could involve separating foreground from background, or clustering regions of pixels based on similarities in color or shape.

SEGMENTAZIONE IN IMAGE PROCESSING



THRESHOLDING

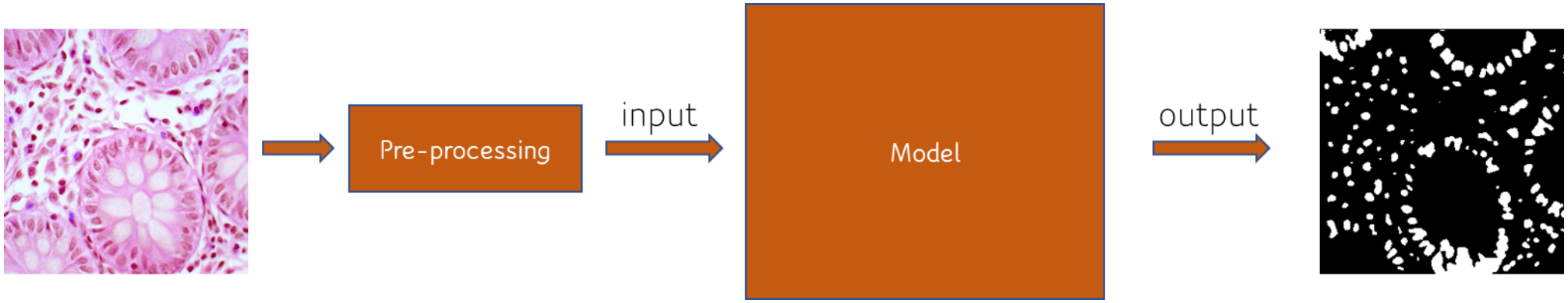


CLUSTERING

STRUTTURA DELLE MASCHERE TARGET

Struttura one-hot encoding
[1,0,0,0,0], [0,1,0,0,0], [0,0,1,0,0], [0,0,0,1,0], [0,0,0,0,1]

COME FUNZIONA LA SEGMENTAZIONE?



MODELLI USATI IN SEGMENTAZIONE

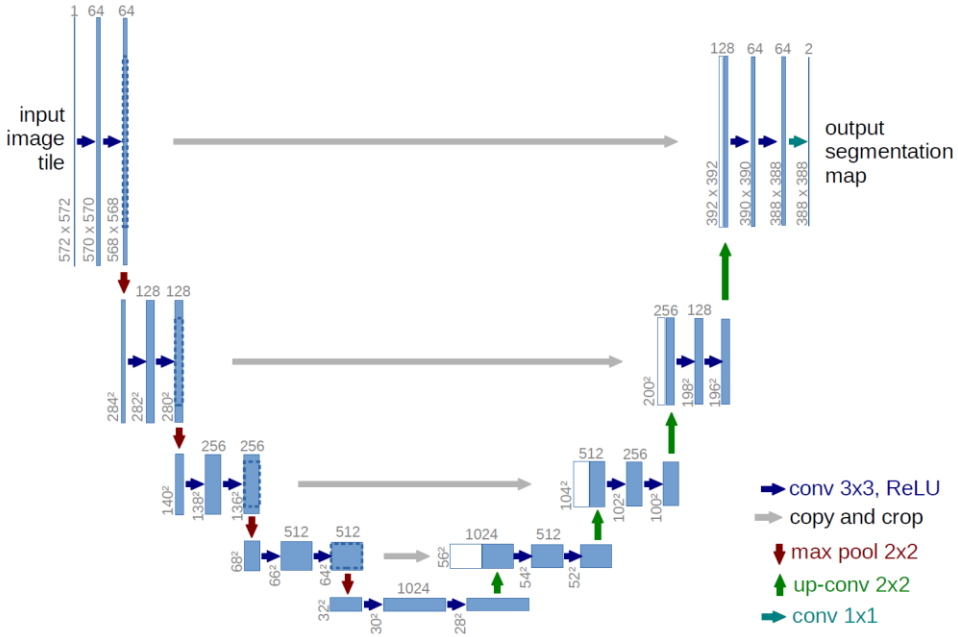
Approcci convoluzionali:

- U-Net
- RCNN (Region Based Convolution Neural Networks)

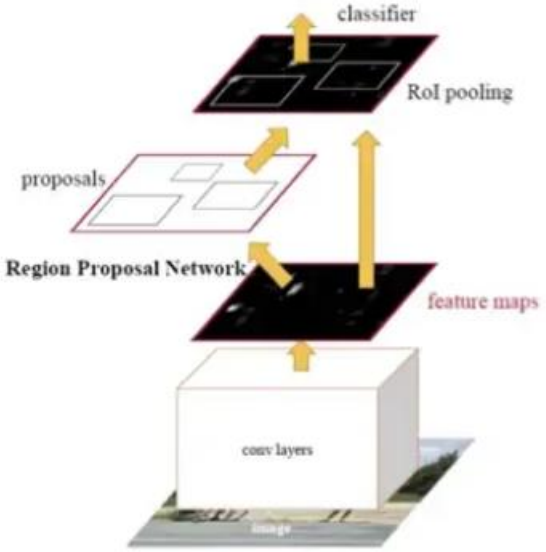
Altri approcci:

- RNN (Recurrent Neural Network)
- GAN (Generative Adversarial Network)
- ViT (Vision Trasformers)

Approcci Convoluzionali

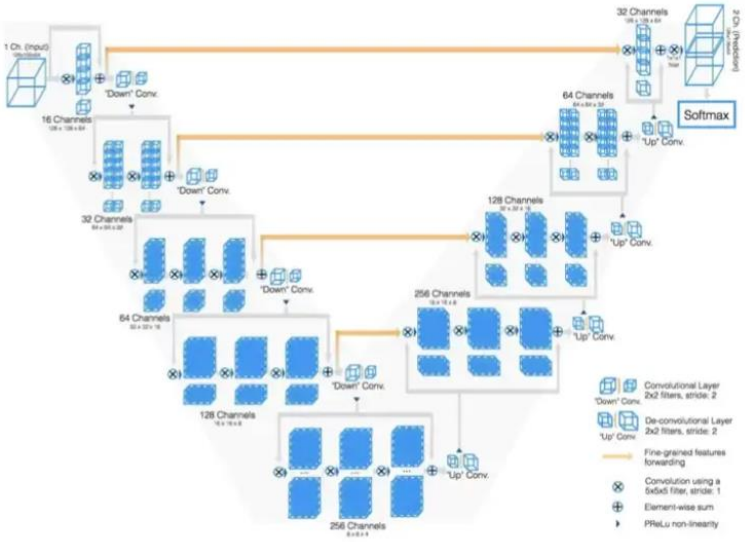


U-Net

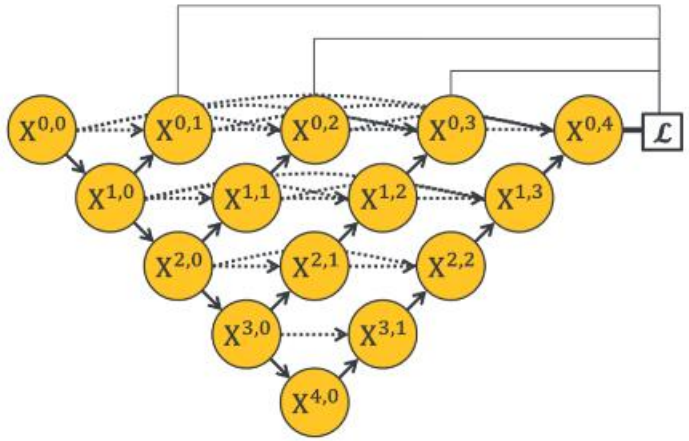


R-CNN

U-NET E ALCUNE VARIANTI

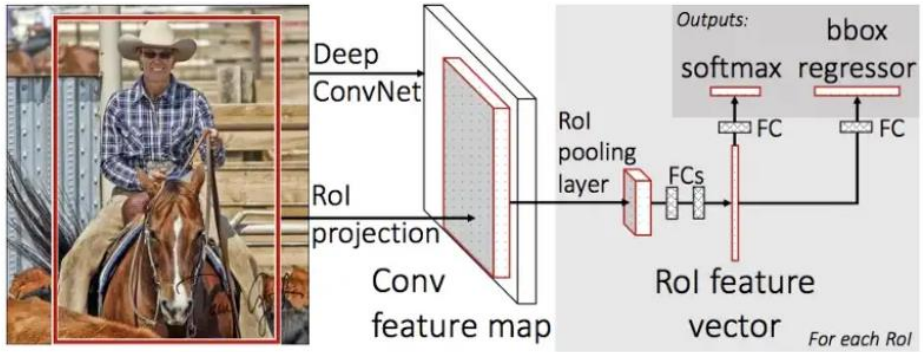


V-Net



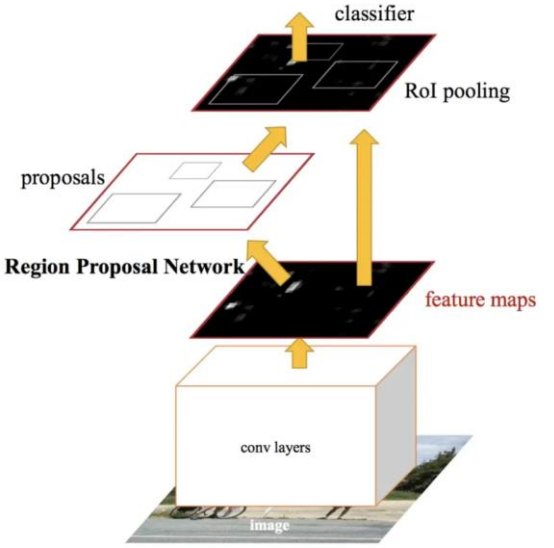
U-Net++

R-CNN E DIVERSE IMPLEMENTAZIONI

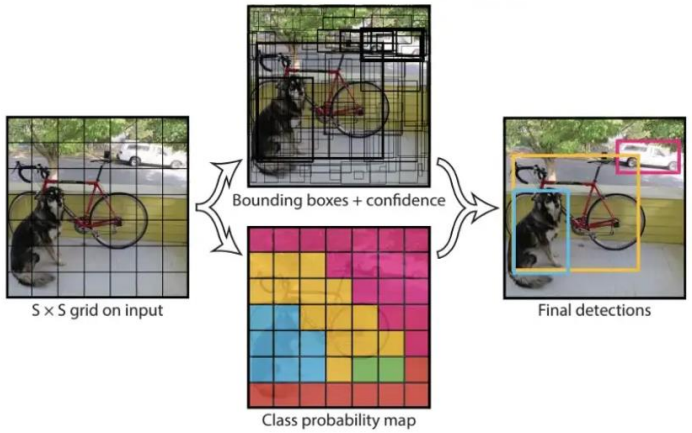


Fast-RCNN

Selective Search

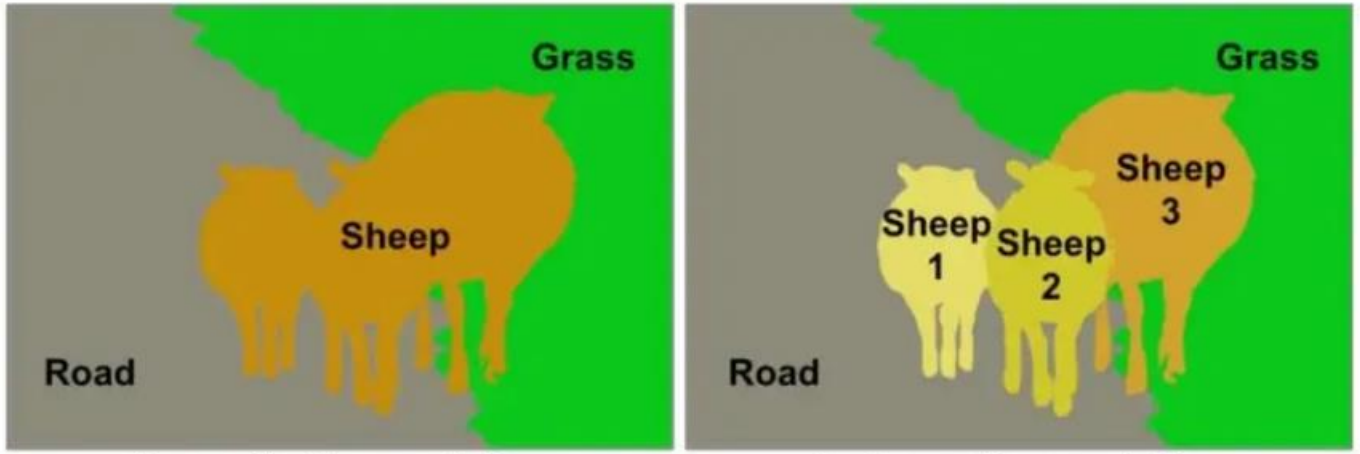


Faster-RCNN

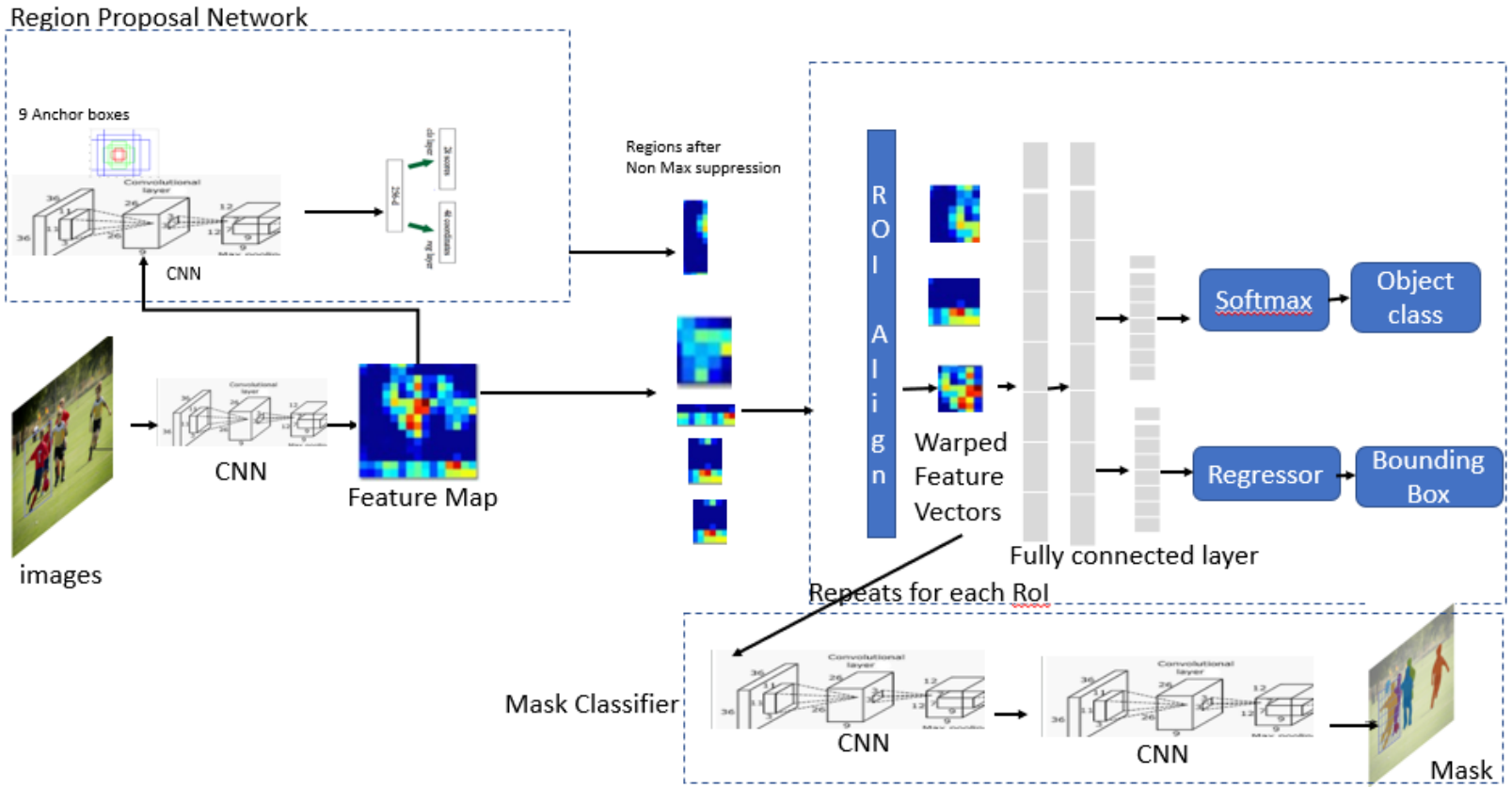


YOLO

SEMANTIC VS INSTANCE SEGMENTATION



MASK R-CNN

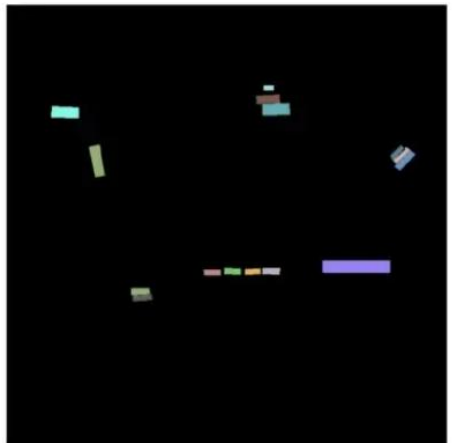

Region Proposal Network
9 Anchor boxes
CNN
Regions after
Non Max suppression
images
CNN
Feature Map
ROI Align
Warped
Feature
Vectors
Softmax
Object
class
Regressor
Bounding
Box
Fully connected layer
Repeats for each RoI
Mask Classifier
CNN
CNN
Mask

METRICHE

- Pixel accuracy
- IOU (Intersection Over Union) → $$J(A,B) = \frac{|A \cap B|}{|A \uplus B|} = \frac{|A \cap B|}{|A| + |B|}.$$
- Dice Coefficient → $$DSC = \frac{2|X \cap Y|}{|X| + |Y|}$$

ESEMPIO DEL CALCOLO DELLE METRICHE

INPUT



TARGET

PREDICTION

- PA = 95%
- IOU = 47,5%
- DICE= 47,5%