



UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO



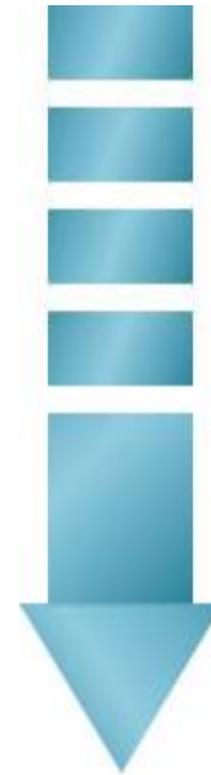
SVILUPPO DI UN SISTEMA DI RICONOSCIMENTO AUTOMATICO DI BIOFILM IN SCANSIONI DI RINOCITOLOGIA

Relatore:
Prof. Giovanni Dimauro

Candidato:
Enrico Ladisa
Mat. 649971

Sommario

- ▶ Cos'è il biofilm
- ▶ Obiettivo
- ▶ Analisi della texture
- ▶ Analisi colorimetrica
- ▶ Preprocessing dei dati
- ▶ Il classificatore
- ▶ Risultati ottenuti
- ▶ Sviluppi futuri



Cos'è il Biofilm

Il biofilm è una matrice protettiva prodotta da **colonie batteriche**, composta da polisaccaridi, proteine e DNA extracellulare.

Il biofilm protegge i batteri dagli agenti antimicrobici quali:

- ▶ Antibiotici
- ▶ Disinfettanti
- ▶ Detergenti

Cos'è il Biofilm

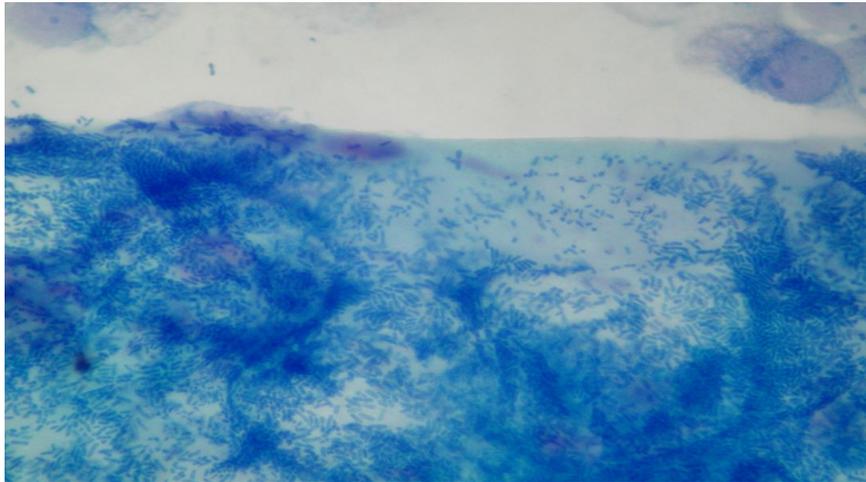
La produzione di biofilm è alla base di molte infezioni batteriche.

Presenza di biofilm è stata riscontrata anche nella **mucosa** di pazienti affetti da rinopatie infettive.

Il ritrovamento del biofilm negli strisci di mucosa è un punto chiave nella diagnosi di queste affezioni.

Cos'è il Biofilm

Le caratteristiche morfologiche e cromatiche del biofilm come:

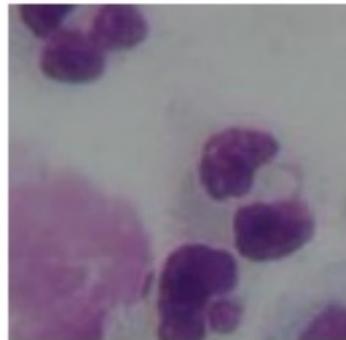


- ▶ L'aspetto di una **pellicola** dai contorni ben definiti;
- ▶ La pigmentazione del **ciano**;

lo rendono facilmente riconoscibile.

Obiettivo

Sviluppare un sistema software che sia in grado di distinguere il biofilm da altra materia in scansioni di rinocitologia.



ALTRO



BIOFILM

Obiettivo

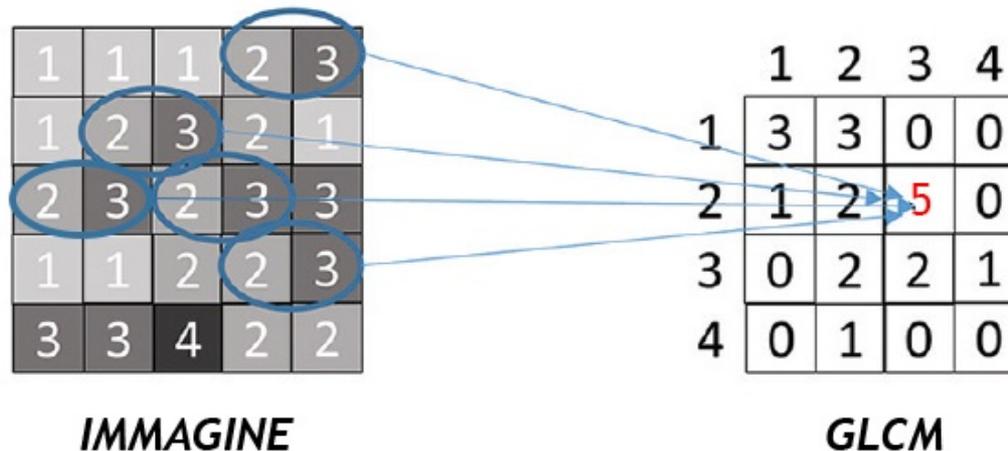
Tre passaggi fondamentali:

- ▶ Estrazione delle informazioni sulla **texture**;
- ▶ Estrazione delle informazioni sul **colore**;
- ▶ Addestramento di un **classificatore**.

Analisi della texture

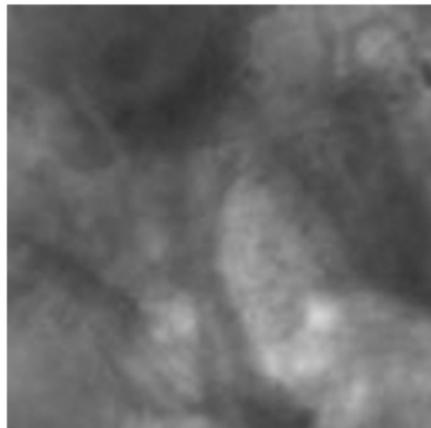
La texture è la disposizione geometrica dei livelli di grigio.

Questa disposizione può essere schematizzata tramite la *matrice di co-occorrenza dei livelli di grigio*.



Analisi della texture

Dalla matrice, si possono estrarre dei valori riassuntivi detti **descrittori di Haralick**.



$$\rightarrow \begin{bmatrix} a & b & \dots & n \\ a & b & \dots & n \\ \vdots & \vdots & \ddots & \vdots \\ a & b & \dots & n \end{bmatrix}$$

GLCM

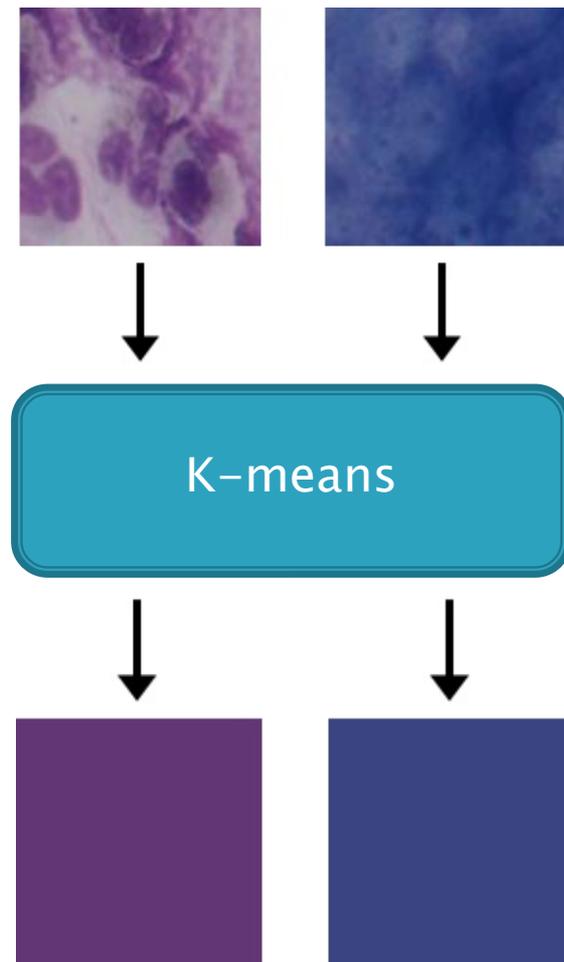
$$\rightarrow \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_{13} \end{bmatrix}$$

**HARALICK
FEATURES**

Analisi del colore

Ogni immagine è stata elaborata per estrarne il colore dominante.

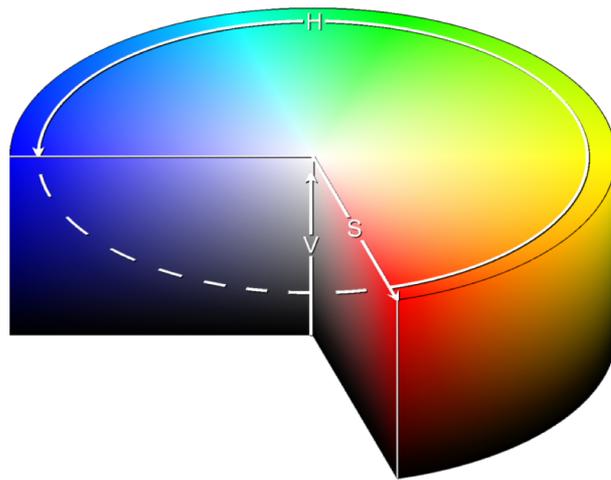
La ricerca del colore dominante è stata effettuata tramite l'algoritmo di clustering **K-means**.



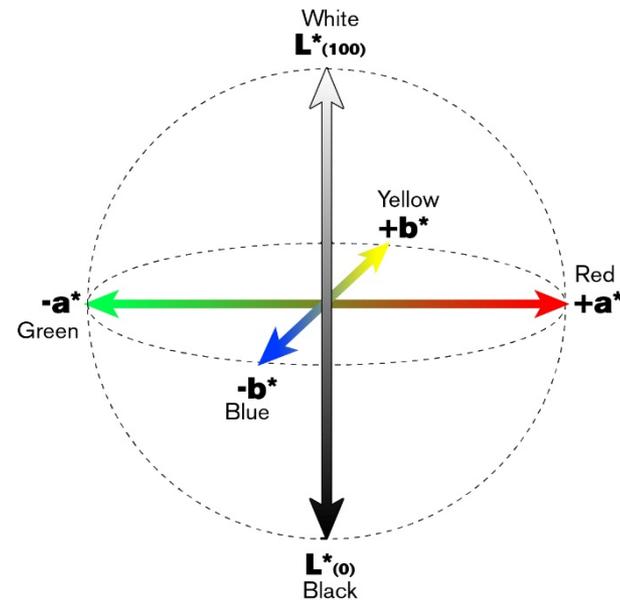
Analisi del colore

I modelli di colore utilizzati in questa sperimentazione sono il modello **HSV** e il modello **LAB**.

Modello HSV



Modello LAB

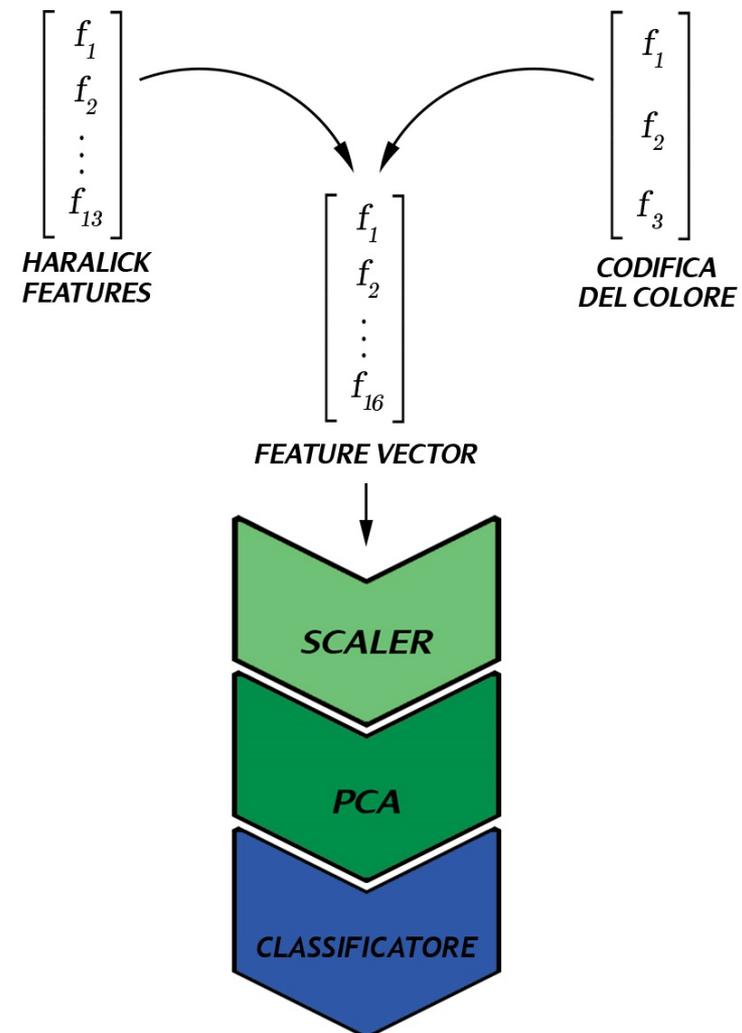


Preprocessing dei dati

I dati estratti dalle immagini sono stati rielaborati prima di essere dati in input al classificatore.

Sono stati effettuati:

- ▶ Scaling;
- ▶ Principal Components Analysis (PCA);



Il classificatore

I feature vector delle immagini contenenti biofilm sono utilizzati per addestrare il classificatore.

Algoritmo di
classificazione

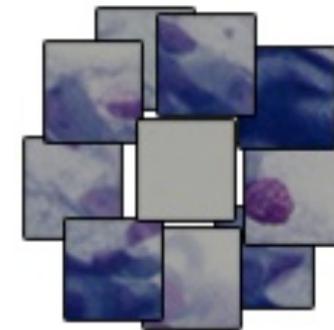
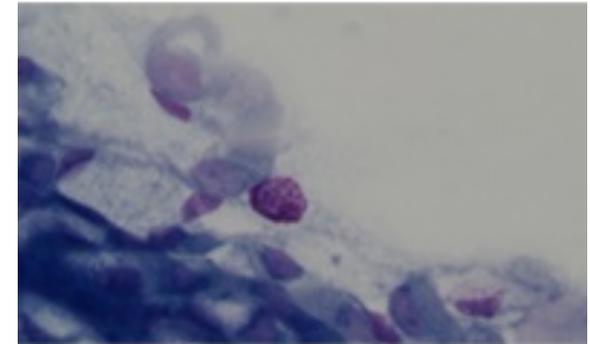


Isolation Forest

Il dataset

Per costruire il dataset, da scansioni di rinocitologia son state ritagliate tiles di dimensione 128 x 128 pixel.

- ▶ **Training set:** 50 tiles di biofilm;
- ▶ **Test set:** 100 tiles miste.



Ambiente di sviluppo

L'intero sistema è stato sviluppato utilizzando il linguaggio di programmazione **Python**.



Librerie utilizzate:

- ▶ OpenCv
- ▶ Mahotas
- ▶ Numpy
- ▶ Scikit-learn



Risultati ottenuti

Quattro diverse configurazioni:

▶ HSV, con PCA

	<i>TRUE</i>	<i>FALSE</i>
<i>POSITIVE</i>	41	0
<i>NEGATIVE</i>	50	9

▶ LAB, con PCA

	<i>TRUE</i>	<i>FALSE</i>
<i>POSITIVE</i>	44	6
<i>NEGATIVE</i>	44	6

▶ HSV, senza PCA

	<i>TRUE</i>	<i>FALSE</i>
<i>POSITIVE</i>	49	15
<i>NEGATIVE</i>	35	1

▶ LAB, senza PCA

	<i>TRUE</i>	<i>FALSE</i>
<i>POSITIVE</i>	48	19
<i>NEGATIVE</i>	31	2

Risultati ottenuti

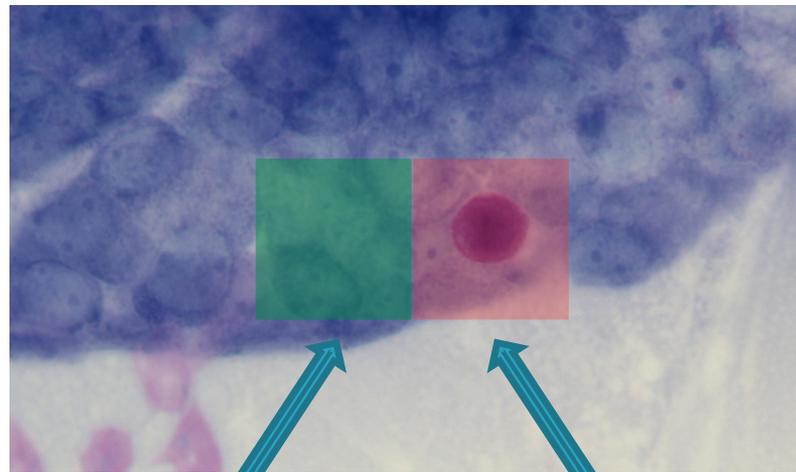
La configurazione più accurata è anche quella col maggior tasso di falsi positivi.

	<i>HSV</i>		<i>LAB</i>	
	<i>Con PCA</i>	<i>Senza PCA</i>	<i>Con PCA</i>	<i>Senza PCA</i>
<i>Accuratezza</i>	91%	84%	88%	79%
<i>Sensibilità</i>	0.82	0.98	0.88	0.96
<i>Miss Rate</i>	0.18	0.02	0.12	0.04

Un alto tasso di falsi negativi riduce la qualità dell'analisi.

Risultati ottenuti

La gravità di un numero ridotto di falsi negativi è attenuata dalle dimensioni delle tiles.



Tile classificata
correttamente

Falso negativo

Configurazione ottimale: HSV, senza PCA.

Sviluppi Futuri

- ▶ Aumentare l'accuratezza del sistema.
- ▶ Generalizzare la dimensione delle immagini.
- ▶ Integrazione con sistemi diagnostici moderni.



UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO



Grazie per l'attenzione!