

# ANALISI DI CLUSTERING CON DATI DI TOMOGRAFIA A COERENZA OTTICA IN PAZIENTI AFFETTI DA NEUROPATIA OTTICA EREDITARIA DI LEBER NELL'AMBITO ML NON-SUPERVISIONATO TRAMITE TECNICA DI NON-NEGATIVE MATRIX FACTORIZATION

Romagnoli Martina<sup>1,2</sup>, Carbonelli Michele<sup>3</sup>, Amore Giulia<sup>3</sup>, D'Agati Pietro<sup>4</sup>, Barboni Piero<sup>5</sup>, Vestri Annarita<sup>2</sup>, Carelli Valerio<sup>1, 3#</sup>, La Morgia Chiara<sup>3,4#</sup>

# Questi autori hanno contribuito in eguale misura.

<sup>1</sup> IRCCS Istituto delle Scienze Neurologiche di Bologna, Programma di Neurogenetica - Bologna (Italia)

<sup>2</sup> Scuola di Specializzazione in Statistica Sanitaria e Biometria, Dipartimento di Sanità Pubblica e Malattie Infettive, Sapienza Università di Roma, (Italia)

<sup>3</sup> Dipartimento di Scienze Biomediche e Neuromotorie, Università di Bologna - Bologna (Italia)

<sup>4</sup> IRCCS Istituto delle Scienze Neurologiche di Bologna, UOC Clinica Neurologica - Bologna (Italia)

<sup>5</sup> Dipartimento di Oftalmologia, Università Vita-Salute, IRCCS Ospedale San Raffaele - Milano (Italia)

## Introduzione

La neuropatia ottica ereditaria di Leber (LHON) è una malattia genetica, neurodegenerativa del nervo ottico, dovuta a mutazioni del DNA mitocondriale, caratterizzata da un'improvvisa e severa perdita della visione centrale, nella maggior parte dei casi bilaterale, che colpisce soprattutto i giovani maschi adulti (età d'esordio tipica: 18-35 anni) anche se si riportano esordi di malattia tra i 2 e gli 87 anni d'età <sup>[1]</sup>.

La LHON è una malattia rara a penetranza incompleta <sup>[1]</sup>, ovvero, tutti i soggetti di una famiglia sono portatori del difetto genetico causativo, alcuni rimangono asintomatici e solo un sottogruppo sviluppa la malattia. Al momento, non esistono elementi predittivi certi per dire chi si ammalerà ad un certo punto della sua vita, ma possono esserci segni subclinici documentabili se il paziente viene esaminato mediante tomografia a coerenza ottica (OCT), differenti a seconda che questo sia asintomatico o sintomatico-acuto <sup>[2,3,4]</sup>.

L'OCT è uno strumento non invasivo grazie al quale è possibile misurare lo spessore dei diversi strati della retina e delle fibre del nervo ottico. In particolare, lo strumento OCT a nostra disposizione è il *DRI OCT Triton* (Topcon) che si avvale di una tecnologia Swept-Source di imaging multi-modale del fondo oculare.

Tramite l'OCT possono essere rilevate alcune alterazioni nei portatori di mutazione LHON asintomatici, mentre coloro che sono nella fase attiva di malattia (entro i primi 6 mesi dall'esordio di malattia) presentano ulteriori patterns di alterazione dei parametri OCT <sup>[2,3,4]</sup>.

Un quesito clinico molto rilevante e ancora privo di risposta è dato dall'identificazione di parametri OCT putativi che caratterizzano la transizione da portatore di mutazione LHON asintomatico all'essere LHON sintomatico con perdita irreversibile della vista.

## Obiettivi

Obiettivo di questa analisi di clustering mediante metodica di ML Non-negative Matrix Factorization (NMF) <sup>[5]</sup> è stato suddividere i campioni (occhi) in clusters con profili OCT simili in modo non-supervisionato ed individuare le variabili OCT che maggiormente influenzavano la migliore clusterizzazione risultante.

## Metodi

La popolazione oggetto di questa analisi esplorativa è costituita da pazienti con diagnosi genetica di LHON (sia portatori asintomatici che sintomatici in fase acuta di malattia) e individui sani (né portatori di mutazione genetica LHON né affetti da patologie oculari del nervo ottico).

Il dataset oggetto dell'analisi di clustering include dati OCT relativamente a 30 occhi di pazienti classificati come "affected\_acute" (sintomatici in fase acuta di malattia); 99 occhi di pazienti classificati come "carrier" (portatori asintomatici) e 44 occhi di individui sani. Per ogni individuo, il dataset contiene i dati relativi a 2 predittori demografici (età; sesso) e a 41 predittori OCT, per un totale di 43 predittori ( $p = 43$ ), relativi a 173 unità osservate (corrispondenti ai 173 occhi dei soggetti valutati tramite OCT). Nel dataset in questione, per ciascuna unità osservata (occhio) era poi disponibile l'outcome clinico (Y), specificamente una variabile categorica a 3 livelli (categorie) come segue:

\_ "affected\_acute" = portatore della mutazione causativa LHON che presenta calo dell'acuità visiva e scotoma centrale, ovvero, sintomatico in fase acuta di malattia;

\_ carrier = paziente portatore della mutazione LHON che non presenta calo dell'acuità visiva, ovvero, asintomatico;

\_ control = individuo sano.

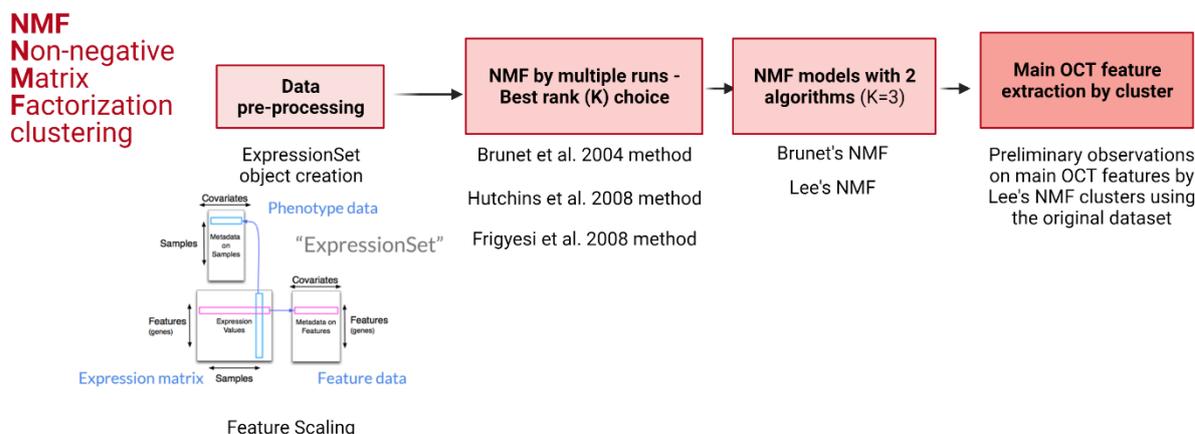
Sulla base delle suddette premesse scientifiche, questo studio si è concentrato sull'implementazione di un algoritmo di clustering e ha rappresentato la prima analisi esplorativa di questo tipo applicata a una tipica fonte dati OCT (Fig.1a). In particolare, questa analisi di clustering è stata effettuata utilizzando R (versione software 4.0.0) ed i relativi pacchetti NMF, factoextra, FactoMineR e cluster, e dati OCT raccolti retrospettivamente dallo strumento *Swept-Source DRI OCT Triton* (Topcon, Tokyo, Giappone), attualmente utilizzato come da pratica clinica di routine presso gli Ambulatori Neuro-oftalmologici dell'IRCCS Istituto delle Scienze Neurologiche di Bologna (Bologna, Italia).

## Risultati

L'algoritmo di clustering implementato ha confermato la presenza di tre clusters, di cui uno formato esclusivamente da campioni clinicamente e precedentemente classificati come "control" (sani) o "LHON carrier" (portatori asintomatici) (Fig.1b). Inoltre, le variabili OCT maggiormente rilevanti per questa clusterizzazione, sono risultate riconducibili alla nota degenerazione LHON-maculare delle cellule ganglionari, osservabile in fase acuta di malattia (Fig.1b).

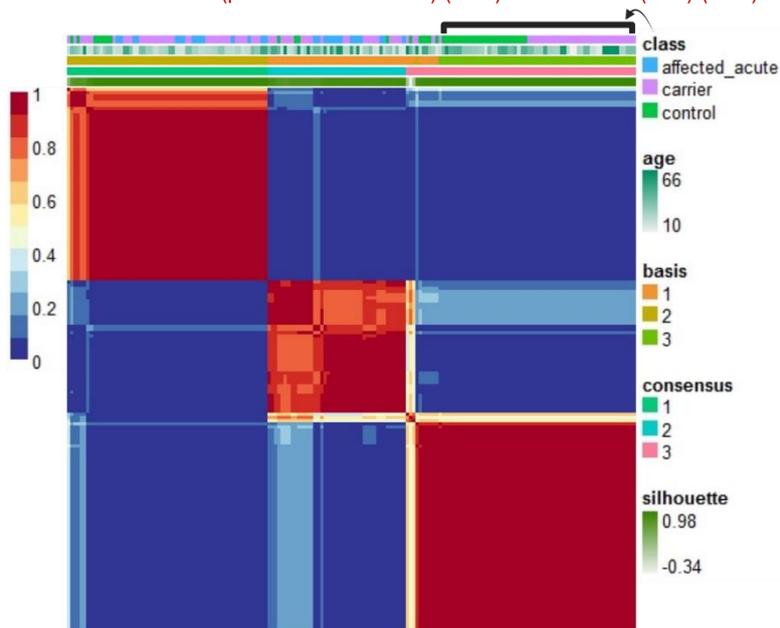
Figura 1.

A. Diagramma di flusso dell'analisi esplorativa del clustering tramite tecnica NMF.



B. Consensus matrix e risultati preliminari ottenuti implementando l'algoritmo NMF di Lee (50 runs, K = 3).

Il clustering NMF ha confermato la presenza di una struttura dati nel dataset OCT originale scegliendo di fattorizzare per K = 3; un aspetto degno di nota è che uno dei cluster così risultanti (cluster\_3) includeva solo occhi precedentemente e clinicamente classificati come "carrier" (portatori asintomatici) (59%) e di controllo (sani) (41%).



## Conclusioni

Questi risultati preliminari suggeriscono l'utilità di questo tipo di algoritmi di NMF-clustering nell'estrazione automatica di informazioni da dataset OCT massivi per la comprensione di ulteriori aspetti chiave di questa malattia.

## **Bibliografia**

- [1] Yu-Wai-Man P., Votruba M., Burte F. et al., A neurodegenerative perspective on mitochondrial optic neuropathies. *Acta Neuropathol.*, 2016 Dec;132(6):789-806.
- [2] Barboni P., Savini G., Valentino M.L., et al., Retinal nerve fiber layer evaluation by optical coherence tomography in Leber's hereditary optic neuropathy. *Ophthalmology*, 2005 Jan;112(1):120-6.
- [3] Balducci N., Savini G., Cascavilla M.L., et al., Macular nerve fibre and ganglion cell layer changes in acute Leber's hereditary optic neuropathy. *Br J Ophthalmol.*, 2016 Sep;100(9):1232-7.
- [4] Barboni P., Savini G., Feuer W.J., et al., Retinal nerve fiber layer thickness variability in Leber hereditary optic neuropathy carriers. *Eur J Ophthalmol.*, 2012 Nov-Dec;22(6):985-91.
- [6] Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics.*, 2010 Jul 2;11:367.