

Unsupervised Clustering of Optical Coherence Tomography Data in Patients with Leber Hereditary Optic Neuropathy using Non-Negative Matrix Factorization and K-Means: A Comparison

Romagnoli Martina⁽¹⁾, Carbonelli Michele⁽²⁾, Amore Giulia⁽²⁾, Caporali Leonardo⁽¹⁾, Fiorini Claudio⁽¹⁾, Barboni Piero⁽³⁾, Vestri Annarita⁽⁴⁾, Luigi Palla⁽⁴⁾, Carelli Valerio^(1,2), La Morgia Chiara^(2,5)

(1) IRCCS Institute of Neurological Sciences of Bologna, Neurogenetics Program – Bologna (Italy)

(2) Department of Biomedical and Neuromotor Sciences, University of Bologna – Bologna (Italy)

(3) Department of Ophthalmology, Vita-Salute University, IRCCS San Raffaele Hospital – Milan (Italy)

(4) Department Public Health and Infectious Disease, University of Rome La Sapienza – Rome (Italy)

(5) IRCCS Institute of Neurological Sciences of Bologna, Neurology Unit – Bologna (Italy)

CORRESPONDING AUTHOR: Romagnoli Martina, martina.romagnoli87@gmail.com

INTRODUCTION

Leber Hereditary Optic Neuropathy (LHON) is a rare genetic neurodegenerative disorder of the optic nerve, caused by mitochondrial DNA (mtDNA) pathogenic variants. It leads to sudden and severe central vision loss, mostly bilateral, typically in young adult males (onset age 18–35), though cases from 2 to 87 year of disease onset have been reported [1]. LHON has incomplete penetrance: all family members may carry the causative mtDNA pathogenic variant, but only some develop the disease phenotype. No definite predictors of disease conversion exist. However, subclinical signs can be detected through Optical Coherence Tomography (OCT), which vary between LHON asymptomatic carriers and symptomatic patients [2,3,4]. OCT is a non-invasive imaging technique that measures the thickness of retinal layers and optic nerve fibers. We used the DRI OCT Triton (Topcon), a swept-source multimodal imaging OCT device. LHON asymptomatic carriers may show early retinal alterations, while symptomatic individuals in the acute phase (within 6 months from onset) present distinct OCT phenotypes. Identifying putative OCT parameters predicting clinical conversion is an urgent unmet clinical need.

OBJECTIVES

To apply unsupervised clustering techniques to OCT data to identify latent subgroups of eyes with similar structural patterns, and assess their coherence with known clinical classes.

METHODS

We analyzed 173 eyes from symptomatic LHON patients (acute phase), asymptomatic LHON carriers, and healthy controls, based on 41 OCT parameters related to Ganglion Cell Layer (GCL), Retinal Nerve Fiber Layer (RNFL), and choroidal thickness. Data were normalized and clustered using: (1) Non-negative Matrix Factorization (NMF) via Brunet and Lee methods, running 50 iterations with cluster number (k) optimization based on internal quality indices [5]; (2) K-means clustering with optimal k selected using Elbow and Gap statistics. We also constructed a complete ExpressionSet object including phenotypic and clinical metadata to facilitate integration and visualization [5].

RESULTS

All methods identified an optimal partition into 3 clusters, broadly consistent with the clinical classification. Brunet-based NMF outperformed Lee-NMF in capturing the clinical structure (purity 0.601 vs 0.572; entropy 0.744 vs 0.784), likely due to its ability to model sparse data, such as OCT matrices where a few variables dominate the individual profiles. Then, the extracted metagenes (partitions) showed localized structural patterns in RNFL and GCL sectors. K-means also separated groups meaningfully, although with more overlap, especially among symptomatic eyes.

CONCLUSIONS

Among the clustering methods tested, Brunet-based NMF emerged as the most suitable for unsupervised stratification of LHON patients, carriers, and controls based on OCT data. Its advantage lies in the ability to highlight sparse but informative features — i.e., those OCT parameters that best discriminate between clinical groups — allowing for more distinct phenotypic clustering. These findings support the use of data-driven approaches for structural profiling and future development of predictive tools for LHON conversion.

REFERENCES

1. Carelli V, Ross-Cisneros F. N., Sadun A. A., Mitochondrial dysfunction as a cause of optic neuropathies. *Prog Retin Eye Res.* 2004 Jul; 23(1):53–89
2. Carbonelli M, La Morgia C, Romagnoli M et al. Capturing the Pattern of Transition From Carrier to Affected in Leber Hereditary Optic Neuropathy. *Am J Ophthalmol.* 2022 Sep;241:71-79.
3. Barboni P, Carbonelli M., Savini G. et al. Natural history of Leber's hereditary optic neuropathy: longitudinal analysis of the retinal nerve fiber layer by optical coherence tomography. *Ophthalmology.* 2010 Mar; 117(3): 623–627
4. Borrelli E., Triolo G., Cascavilla M. L. et al. Changes in choroidal thickness follow the RNFL changes in Leber's hereditary optic neuropathy. *Sci Rep.* 2019; 9(1):10728
5. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics.*, 2010 Jul 2;11:367.