3° Πανελλήνιο Συνέδριο
Ιατρικής Φυσικής

27° Διαπανεπιστημιακό Συνέδριο
Ακτινολογίας

Τα Ορθά Σφάλματα: Ερμηνεύοντας τα σφάλματα ταξινόμησης ως βιοδείκτες ομοιότητας

Ι. Σταθόπουλος[1], Ν. Μπουζιάνης[2], Ε. Καραβασίλης[2], Γ. Βελονάκης[3], Ε. Ευσταθόπουλος[1]

[1]Εργαστήριο Εφαρμοσμένης Ιατρικής Φυσικής, Τμήμα Ιατρικής, ΕΚΠΑ ·

[2]Εργαστήριο Ιατρικής Φυσικής, Τμήμα Ιατρικής, ΔΠΘ ·

[3]Β΄ Εργαστήριο Ακτινολογίας, Τμήμα Ιατρικής, ΕΚΠΑ

30, 31 ΟΚΤ + 1 ΝΟΕ
2026

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΥΓΕΙΑΣ
- ΤΜΗΜΑ ΙΑΤΡΙΚΗΣ, ΛΑΡΙΣΑ

Οργάνωση

ΕΦΙΕ
Ένωση Φυσικών Ιατρικής Ελλάδος

ΕΛΛΗΝΙΚΟ ΚΟΛΛΕΓΙΟ ΑΚΤΙΝΟΛΟΓΙΑΣ

Εισαγωγή

- Η απόδοση των αλγορίθμων ΤΝ στην ιατρική απεικόνιση αξιολογείται κυρίως με ακρίβεια, ευαισθησία και ειδικότητα· οι λανθασμένες ταξινομήσεις θεωρούνται απλώς ανεπιθύμητες.
- Σε πολύπλοκες παθολογίες του εγκεφάλου, όμως, οι ταξινομητικές αστοχίες μπορεί να αντανακλούν πραγματικές απεικονιστικές και βιολογικές ομοιότητες μεταξύ νοσημάτων — όχι μόνο περιορισμούς του αλγορίθμου.
- Χρόνιες ισχαιμικές αλλοιώσεις, απομυελινωτικές πλάκες και υπερεντάσεις λευκής ουσίας (WMH) εμφανίζουν παρόμοια χαρακτηριστικά στις συμβατικές ακολουθίες MRI (FLAIR, T2).
- Τα radiomics (ποσοτικά χαρακτηριστικά έντασης, υφής και σχήματος) με gradient boosting (XGBoost) επιτρέπουν ταξινόμηση με ερμηνεύσιμα feature importances.

Οι τέσσερις κατηγορίες βλαβών

ΑΕΕ	**Αγγειακό εγκεφαλικό επεισόδιο** · 152 βλάβες
Όγκος	**Όγκοι εγκεφάλου** · 159 βλάβες
ΠΣ	**Πολλαπλή σκλήρυνση** · 1502 βλάβες
WMH	**Υπερεντάσεις λευκής ουσίας** · 2642 βλάβες

Σκοπός

Να διερευνηθεί κατά πόσο τα πρότυπα λανθασμένων ταξινομήσεων ενός μοντέλου μηχανικής μάθησης βασισμένου σε radiomics μπορούν να λειτουργήσουν ως δείκτες απεικονιστικής ομοιότητας μεταξύ εγκεφαλικών παθολογιών, παρέχοντας συμπληρωματική πληροφορία πέρα από τους συμβατικούς δείκτες απόδοσης και ενισχύοντας την ερμηνευσιμότητα των μοντέλων ΤΝ.

Η υπόθεση

Αν τα σφάλματα του ταξινομητή συγκεντρώνονται συστηματικά μεταξύ συγκεκριμένων κατηγοριών, τότε ο πίνακας σύγχυσης δεν μετρά μόνο αποτυχία — χαρτογραφεί τη «συνέχεια» των απεικονιστικών βιοδεικτών.

Μέθοδοι & Υλικά

Δεδομένα

- 124 εξετάσεις MRI εγκεφάλου από την κλινική ρουτίνα · 3 ακολουθίες ανά εξέταση: FLAIR, T2, T1 με σκιαγραφικό (T1ce)
- Κατάτμηση και εξαγωγή 4455 επιμέρους βλαβών: 152 ΑΕΕ, 159 όγκοι, 1502 ΠΣ, 2642 WMH
- Συνολικός όγκος βλαβών 4.34 εκατ. mm^3 (όγκοι 58%, WMH 22%, ΠΣ 14%, ΑΕΕ 6%)

Χαρακτηριστικά radiomics (320 ανά βλάβη)

54 έντασης: στατιστικά 1ης τάξης της κατανομής σήματος (18 × 3 ακολουθίες)

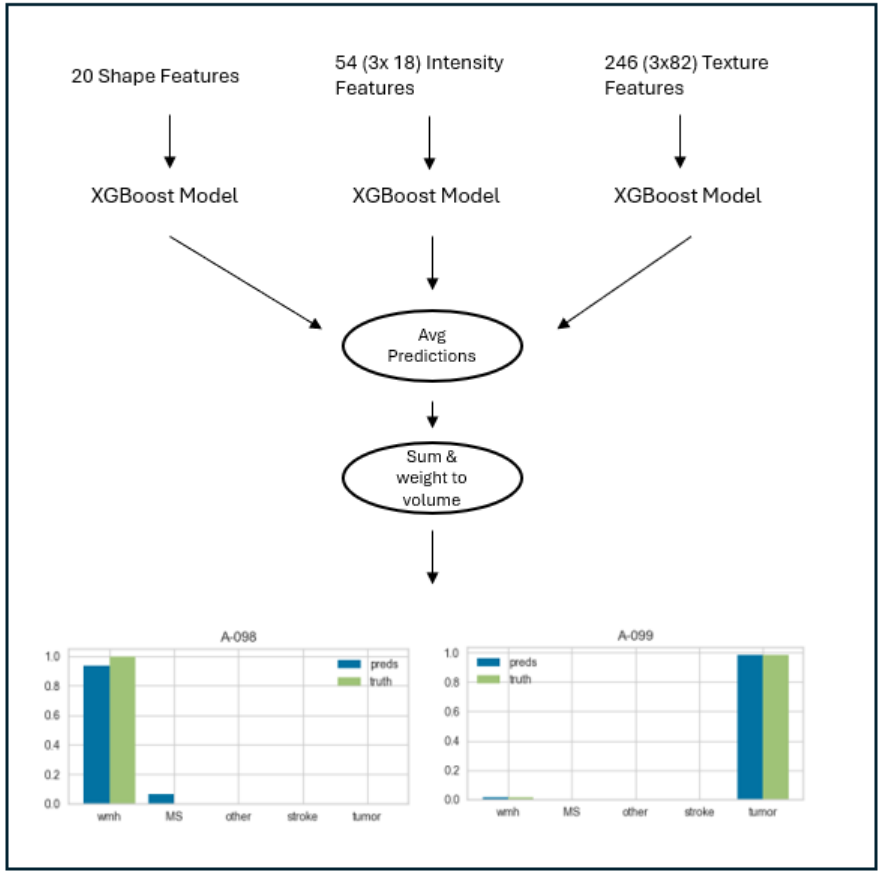
246 υφής: στατιστικά 2ης τάξης — GLCM, GLRLM κ.ά. (82 × 3 ακολουθίες): ετερογένεια και χωρική διάταξη εντάσεων

20 σχήματος: γεωμετρία/μορφολογία — διάκριση μαζών από διάχυτες ή συρρέουσες βλάβες

Μοντέλο & επικύρωση

- Ταξινομητής XGBoost (gradient-boosted δέντρα, cross-entropy loss) — ένα μοντέλο ανά ομάδα χαρακτηριστικών, μέσος όρος προβλέψεων
- Πενταπλή διασταυρούμενη επικύρωση σε επίπεδο εξέτασης: όλες οι βλάβες μιας εξέτασης στο ίδιο fold (χωρίς διαρροή πληροφορίας)
- Αξιολόγηση σε δύο επίπεδα: ανά βλάβη (per-sample) και ανά όγκο (mm^3, στάθμιση κατά όγκο)

Ροή εργασίας



Τρία μοντέλα XGBoost (σχήμα, ένταση, υφή) → μέσος όρος προβλέψεων → άθροιση και στάθμιση κατά όγκο. Κάτω: παράδειγμα προβλέψεων ανά εξέταση (μπλε) έναντι της αλήθειας (πράσινο).

Αποτελέσματα

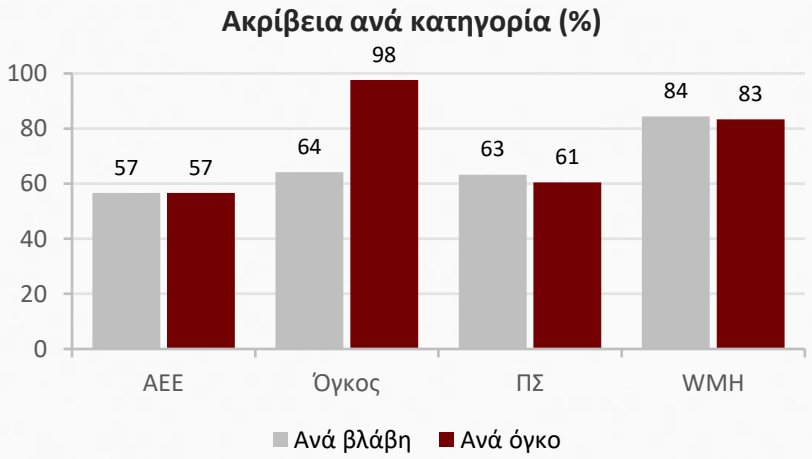
Πίνακας σύγχυσης ανά βλάβη (%) · ακρίβεια 75.6%

Πραγματική ↓ / Πρόβλεψη →	ΑΕΕ	Όγκος	ΠΣ	WMH	n (% συνόλου)
ΑΕΕ	**56.6**	5.3	7.9	30.3	152 (3%)
Όγκος	5.0	**64.2**	11.3	19.5	159 (4%)
ΠΣ	5.5	8.5	**63.2**	22.8	1502 (34%)
WMH	0.3	0.3	15.1	**84.4**	2642 (59%)

Πίνακας σύγχυσης ανά όγκο (%) · ακρίβεια 86.7%

Πραγματική ↓ / Πρόβλεψη →	ΑΕΕ	Όγκος	ΠΣ	WMH	n (% συνόλου)
ΑΕΕ	**56.6**	20.3	8.2	14.8	6%
Όγκος	0.7	**97.6**	1.5	0.2	58%
ΠΣ	0.4	25.5	**60.5**	13.5	14%
WMH	0.0	12.2	4.5	**83.3**	22%

Γραμμές: πραγματική κατηγορία · στήλες: πρόβλεψη του μοντέλου. Διαγώνιος = σωστή ταξινόμηση. Δεξιά στήλη: πλήθος βλαβών (% του συνόλου 4455) / ποσοστό συνολικού όγκου (% του συνόλου 4.338 cm³).



Τα σφάλματα δεν είναι τυχαία

ΑΕΕ → WMH 30.3% (και → ΠΣ 7.9%): οι χρόνιες ισχαιμικές αλλοιώσεις μιμούνται τις υπερεντάσεις λευκής ουσίας

ΠΣ → WMH 22.8%: απομυελινωτικές πλάκες και WMH μοιράζονται φαινότυπο και περικοιλιακή κατανομή

Όγκοι: σαφέστερο απεικονιστικό αποτύπωμα — 97.6% του όγκου τους ταξινομείται σωστά, ελάχιστη επικάλυψη με τις άλλες κατηγορίες

WMH: υψηλότερη ακρίβεια (84.4%) — μεγάλο φορτίο βλαβών, ομοιογενής υπογραφή σε FLAIR/T2

75.6%

συνολική ακρίβεια ανά βλάβη (4455 βλάβες)

86.7%

συνολική ακρίβεια ανά όγκο (4.338 cm^3)

Συμπεράσματα

- Οι λανθασμένες ταξινομήσεις δεν αποτελούν πάντοτε ένδειξη αποτυχίας του αλγορίθμου: ορισμένα σφάλματα αποτυπώνουν πραγματικές απεικονιστικές και βιολογικές ομοιότητες μεταξύ παθολογιών.
- **Η σύγχυση ΑΕΕ/ΠΣ ↔ WMH αναπαράγει γνωστές νευροακτινολογικές προκλήσεις — ο ταξινομητής «σφάλλει» εκεί όπου δυσκολεύεται και ο ακτινολόγος.**
- Οι όγκοι διαχωρίζονται καθαρά· οι ΑΕΕ και η ΠΣ παραμένουν οι δυσκολότερες κατηγορίες, ιδίως στις μικρές βλάβες.
- Η αξιολόγηση σε επίπεδο όγκου (86.7%) πλησιάζει περισσότερο την κλινική σημασία (φορτίο βλαβών) από την αξιολόγηση ανά βλάβη (75.6%).

Περιορισμοί & μελλοντική εργασία

- Μονοκεντρικά δεδομένα, ανισοκατανομή κατηγοριών (WMH 59% των βλαβών) · χωρίς εξωτερική επικύρωση
- Εναρμόνιση radiomics μεταξύ τομογράφων/πρωτοκόλλων · υβριδικά μοντέλα (CNN + radiomics) · ποσοτικοποίηση αβεβαιότητας · πολυκεντρική επικύρωση

Τα «σωστά λάθη» ως εργαλείο

30.3%
ΑΕΕ που ταξινομούνται ως WMH

22.8%
ΠΣ που ταξινομούνται ως WMH

97.6%
όγκος όγκων σωστά ταξινομημένος

< 1%
όγκοι που συγχέονται με ΑΕΕ/WMH (ανά όγκο)

- Επιπλέον δείκτης ερμηνευσιμότητας των μοντέλων ΤΝ
- Ανάδειξη περιοχών διαγνωστικής αβεβαιότητας που αξίζουν προσοχή από τον ακτινολόγο
- Νέα γνώση για τη συνέχεια των απεικονιστικών βιοδεικτών στις νευρολογικές παθήσεις

Βιβλιογραφία

1. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012;48:441–446.

2. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016;278:563–577.

3. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Res 2017;77:e104–e107.

4. Zwanenburg A, Vallières M, Abdalah MA, et al. The Image Biomarker Standardization Initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology 2020;295:328–338.

5. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. IEEE Trans Syst Man Cybern 1973;SMC-3:610–621.

6. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Proc 22nd ACM SIGKDD Int Conf Knowledge Discovery and Data Mining 2016:785–794.

7. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst 2017;30:4765–4774.

8. Wardlaw JM, Smith EE, Biessels GJ, et al. Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration (STRIVE). Lancet Neurol 2013;12:822–838.

9. Filippi M, Rocca MA, Ciccarelli O, et al. MRI criteria for the diagnosis of multiple sclerosis: MAGNIMS consensus guidelines. Lancet Neurol 2016;15:292–303.

10. Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol 2018;17:162–173.