Five-year dementia prediction and decision support system based on real-world data

Δημοσιεύτηκε στο περιοδικό **Frontiers in Aging Neuroscience** (Sec. Alzheimer's Disease and Related Dementias, Volume 17 – 2025)

Εθελοντικές Συμμετοχές → **Πραγματικά Δεδομένα Ηλεκτρονικών Ιατρικών Φακέλων** (EHR)

Στόχος

- Ανάπτυξη μεθοδολογίας για την πενταετή **πρόβλεψη παραγόντων εμφάνισης άνοιας** και την **υποστήριξη αποφάσεων** από το ιατρικό προσωπικό με ακρίβεια και ερμηνευσιμότητα των αποτελεσμάτων

Καινοτομία

- Η μελέτη υπερβαίνει την απλή ανάλυση μεμονωμένων παραγόντων κινδύνου, εντοπίζοντας **συνδυασμούς βιοδεικτών** που διαμορφώνουν **μονοπάτια** ή **σύνολα κανόνων** που σχετίζονται με την εμφάνιση της νόσου. Η μέθοδος παρέχει ερμηνεύσιμες πληροφορίες στους κλινικούς για την πιθανή έναρξη της νόσου και υποστηρίζει αποτελεσματικά τη διαβάθμιση του κινδύνου

Δεδομένα

- Τα συμβατικά μοντέλα εκτίμησης κινδύνου για την άνοια, συχνά βασίζονται σε περιορισμένες και επιλεκτικές ομάδες εθελοντών, οι οποίες δεν αντιπροσωπεύουν απαραιτήτως το δημογραφικό εύρος ούτε το μέγεθος του δείγματος των ασθενών μιας κλινικής

JOHNS HOPKINS MEDICINE SCHOOL OF MEDICINE

research-day ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

BiHELab

Five-year dementia prediction and decision support system based on real-world data

Κριτήρια Επιλογής και Διαχωρισμός Δείνματος

Στο αρχικό δείγμα από το σύστημα Ιατρικού Φακέλου του **John's Hopkins** είχαμε **685.765 ασθενείς**

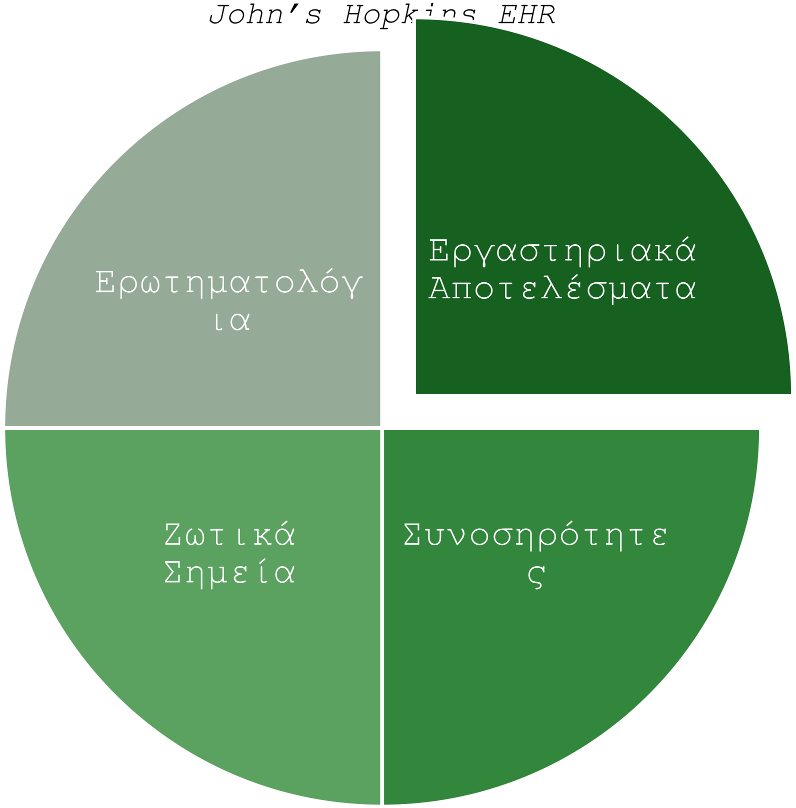
Αφού εφαρμόσουμε κριτήρια επιλογής ασθενών, όπως για παράδειγμα συγκεκριμένα τμήματα της κλινικής (π.χ. Memory Care) και καταστάσεις επίσκεψης (π.χ. ολοκληρωμένη), καταλήγουμε με **197.481 ασθενείς**

Στο δείγμα των 197 χιλιάδων ασθενών αναζητούμε διαγνώσεις άνοιας εντός πενταετούς χρονικού παραθύρου μεταξύ της πρώτης επίσκεψης και της διάγνωσης

Με τη παραπάνω μεθοδολογία εντοπίσαμε **139.477 ασθενείς** που δεν έλαβαν ποτέ κάποια διάγνωση άνοιας και **2.738 ασθενείς (1,9%)** με άνοια

Σύνολο Κλινικών Δεδομένων

John's Hopkins EHR



JOHNS HOPKINS MEDICINE SCHOOL OF MEDICINE

research-day ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

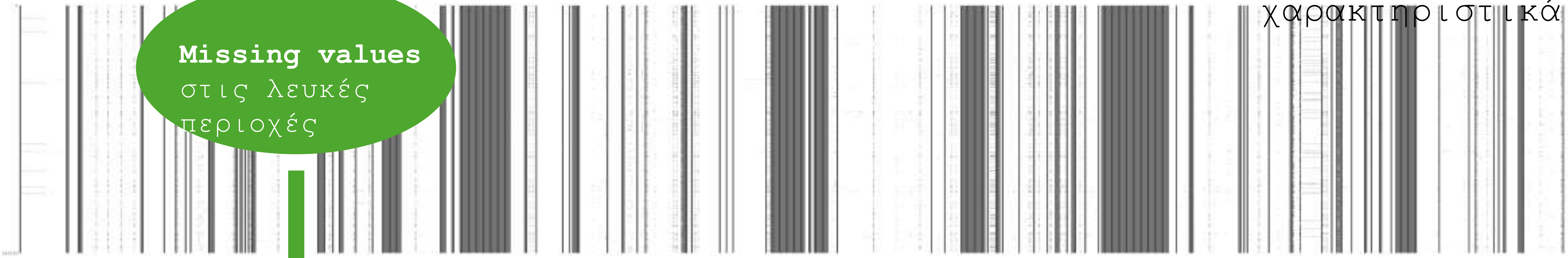
HELLENIC REPUBLIC IONIAN UNIVERSITY 1984

BiHELab

Five-year dementia prediction and decision support system based on real-world data

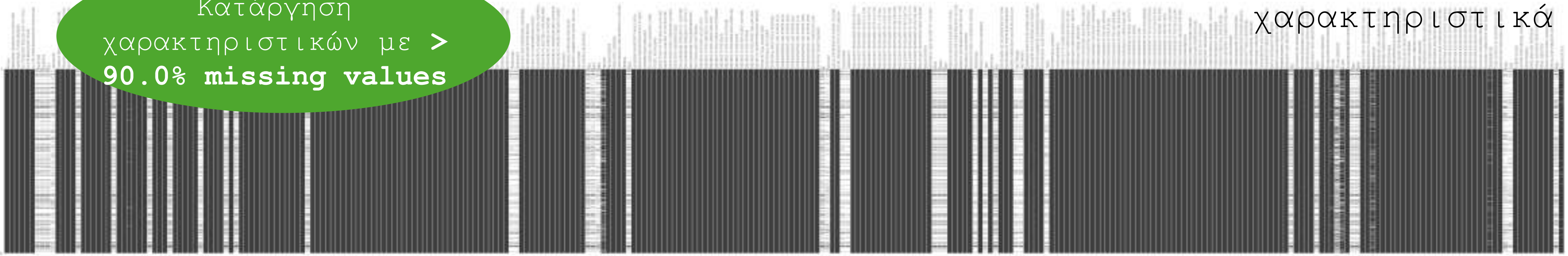
Προεπεξεργασία χαρακτηριστικών (features)

Αρχικά **1,193** χαρακτηριστικά



Missing values στις λευκές περιοχές

Μετά την προεπεξεργασία **166** χαρακτηριστικά



Κατάργηση χαρακτηριστικών με > **90.0% missing values**

JOHNS HOPKINS MEDICINE SCHOOL OF MEDICINE

research-day ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

BiHE Lab

Five-year dementia prediction and decision support system based on real-world data

Πολλαπλές Μονάδες Μέτρησης στα Εργαστηριακά Αποτελέσματα

Μέθοδος Μετατροπής Μονάδων Μέτρησης (UoM Conversions)

- Απαιτούνται πολυάριθμες παραδοχές
- Η μετατροπή σε μία μονάδα αναφοράς δεν είναι πάντοτε εφικτή

cohort_id	pat_enc_csn_id	reference_low	reference_high	ord_num_value	normalized_value
0q9dt6zgdu	122213834722712	150.0	450.0	189	0.13
0q9dt6zgdu	122213834722712	80.0	100.0	80.3	0.015
57t6zg76uy	377192104124341	60.0	99.0	118	1.487
57t6zg76uy	4093850910115426	24.0	44.0	21.3	-0.135
57t6zg76uy	3337238575382784	9.2	12.7	9.1	-0.029
57t6zg76uy	377192104124341	3.5	5.3	3.7	0.111
57t6zg76uy	2247100467764054	1.0	4.0	6.5	1.833
0q9dt6zgdu	122213834722712	0.6	1.3	0.54	-0.086
0q9dt6zgdu	3862318516563560	0.1	1.2	0.37	0.245
0q9dt6zgdu	122213834722712	0.0	0.01	0	0

$$x_{scaled} = \frac{x - \mathrm{NR}_{min}}{\mathrm{NR}_{max} - \mathrm{NR}_{min}}$$

normalized_value :

< 0	Measurement is less than the lower bound of NR
[0..1]	Measurement is within NR
>1	Measurement is greater than the upper bound of NR

(where **measurement** : ord_num_value)

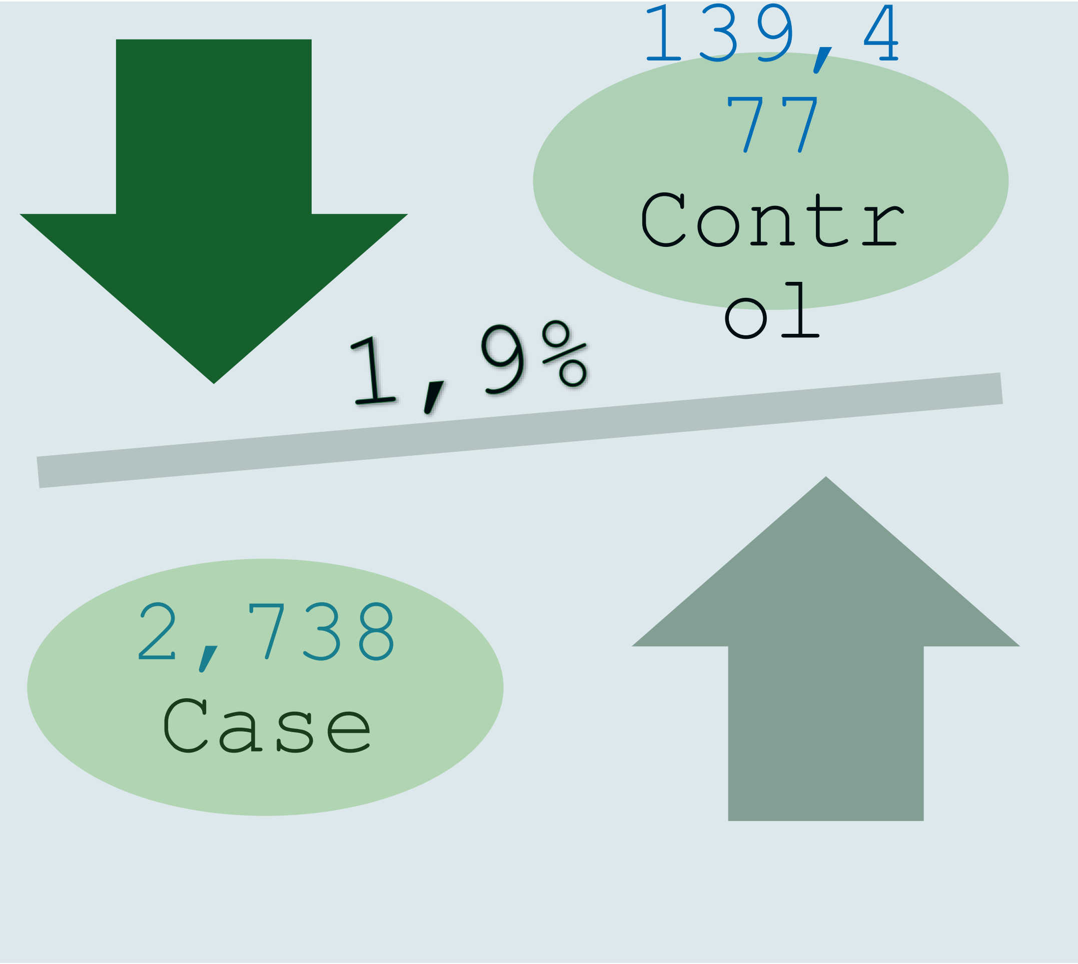
Μέθοδος Κανονικοποίησης Κλινικών Αποτελεσμάτων

- Κανονικοποίηση των αποτελεσμάτων εντός του φυσιολογικού εύρους τιμών τους

Five-year dementia prediction and decision support system based on real-world data

Διαχείριση της **Ανισορροπίας** μεταξύ κλάσεων

- Η Control κλάση χωρίστηκε σε 51 ισομεγέθη υποσύνολα, καθένα από τα οποία είχε το ίδιο μέγεθος με τη Case (μειοψηφική) κλάση
- Κάθε υποσύνολο συνδυάστηκε με την ίδια μειοψηφική κλάση, ώστε να δημιουργηθεί ένα ισορροπημένο σύνολο εκπαίδευσης
- Το μοντέλο εκπαιδεύτηκε ανεξάρτητα σε κάθε υποσύνολο
- Οι μετρικές απόδοσης συγκεντρώθηκαν και συνδυάστηκαν



JOHNS HOPKINS MEDICINE SCHOOL OF MEDICINE

research-day ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

BiHE Lab

Five-year dementia prediction and decision support system based on real-world data

Διαδικασία **Ανάπτυξης** και **Εκπαίδευσης** Μοντέλου

- Αξιολογήθηκαν δύο αλγόριθμοι ensemble μάθησης: **Random Forest (RF)** και **XGBoost (XGB)**, κατάλληλοι για υψηλής διαστασιμότητας και ετερογενή δεδομένα (όπως τα EHRs)

- Και οι δύο μέθοδοι βασίζονται σε δέντρα αποφάσεων και μειώνουν τον κίνδυνο overfitting σε σχέση με ένα μεμονωμένο δέντρο

- Το **Random Forest** μειώνει τη διακύμανση μέσω bootstrap δειγματοληψίας και τυχαίας επιλογής χαρακτηριστικών.

- Το **XGBoost** χρησιμοποιεί boosting και shrinkage για αποφυγή προσαρμογής σε θόρυβο

Και τα δύο μοντέλα παρουσίασαν παρόμοια απόδοση ως προς την ακρίβεια και τη γενίκευση

Η αξιολόγηση έγινε με 5-repeats, 5-fold Repeated Stratified K-Fold cross-validation σε 51 υποσύνολα δεδομένων (συνολικά **1,275 επαναλήψεις**)

Η τελική απόδοση προέκυψε από τον μέσο όρο των αποτελεσμάτων σε μη χρησιμοποιημένα (test) δεδομένα

Η βελτιστοποίηση υπερπαραμέτρων έγινε ξεχωριστά για κάθε υποσύνολο μέσω Bayesian optimization με cross-validation (BayesSearchCV)

Το **Random Forest** θεωρείται πιο ερμηνεύσιμο σε κλινικά πλαίσια, καθώς κάθε δέντρο μπορεί να αναλυθεί ξεχωριστά

JOHNS HOPKINS MEDICINE SCHOOL OF MEDICINE

research-day ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

BiHE Lab

Five-year dementia prediction and decision support system based on real-world data

Ικανότητες **Πρόβλεψης** και **Γενίκευσης** του Μοντέλου μετά από 1,275 Επαναλήψεις Εκπαίδευσης

Metric	Random Forest[a]		XGBoost[a]	
	Value (Std)	95% CI[b]	Value (Std)	95% CI[b]
Mean AUROC	0.776 (0.014)	0.775–0.777	0.795 (0.014)	0.794–0.795
Mean precision control	0.718 (0.015)	0.718–0.719	0.727 (0.025)	0.726–0.729
Mean precision case	0.698 (0.015)	0.697–0.698	0.718 (0.021)	0.717–0.719
Mean test accuracy	0.707 (0.014)	0.706–0.708	0.722 (0.014)	0.722–0.723
Mean F1-score control	0.699 (0.016)	0.698–0.700	0.719 (0.016)	0.718–0.720
Mean F1-score case	0.715 (0.013)	0.714–0.715	0.725 (0.014)	0.725–0.726

[a]Library: scikit-learn 1.6.1.
[b]95% confidence intervals for mean estimates were calculated based on the t-distribution.

Και τα δύο μοντέλα παρουσιάζουν τιμές **AUROC** που κυμαίνονται από **0,77 έως 0,79.** Τα Διαστήματα Εμπιστοσύνης 95% (CI) και για τα δυο μοντέλα υπολογίστηκαν με κατανομή t-Student, παρέχοντας εκτίμηση της ακρίβειας γύρω από τις μετρικές απόδοσης

JOHNS HOPKINS MEDICINE SCHOOL OF MEDICINE

research-day ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

BiHE Lab

Five-year dementia prediction and decision support system based on real-world data

Αναγνώριση της **Άνοιας** μέσω διαγνωστικών μονοπατιών που προκύπτουν από πρότυπα παραγόντων κινδύνου

Decision path
IF
CREATININE[a] <0.94 AND
VLDLCALC[a] <0.10 AND
WEIGHT/SCALE <3167.60 AND
CHOLHDL[a] <0.39 AND
TSH[a] ≥0.19 AND
PROT[a] <0.47 AND
WBC[a] ≥0.21
THEN DEMENTIA: 75% Total Samples: 252

[a] Laboratory markers with measurement normalized within the **normal reference range** (NR) Section 2.3.

Διαγνωστικά μονοπάτια που προσφέρουν ένα πλαίσιο που υποστηρίζει **ταχύτερη** και **ακριβέστερη** διάγνωση της άνοιας

JOHNS HOPKINS MEDICINE SCHOOL OF MEDICINE

research-day ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

BiHELab

Five-year dementia prediction and decision support system based on real-world data

Ενδεικτικοί κανόνες απόφασης

Αν σε δείγμα **252** ασθενών ισχύει

Υψηλή **Κρεατινίνη**

Χαμηλή **VLDLCALC**

Βάρος κάτω από 89,8 Kg

Φυσιολογική **CHOL/HDL**

Χαμηλή **TSH**

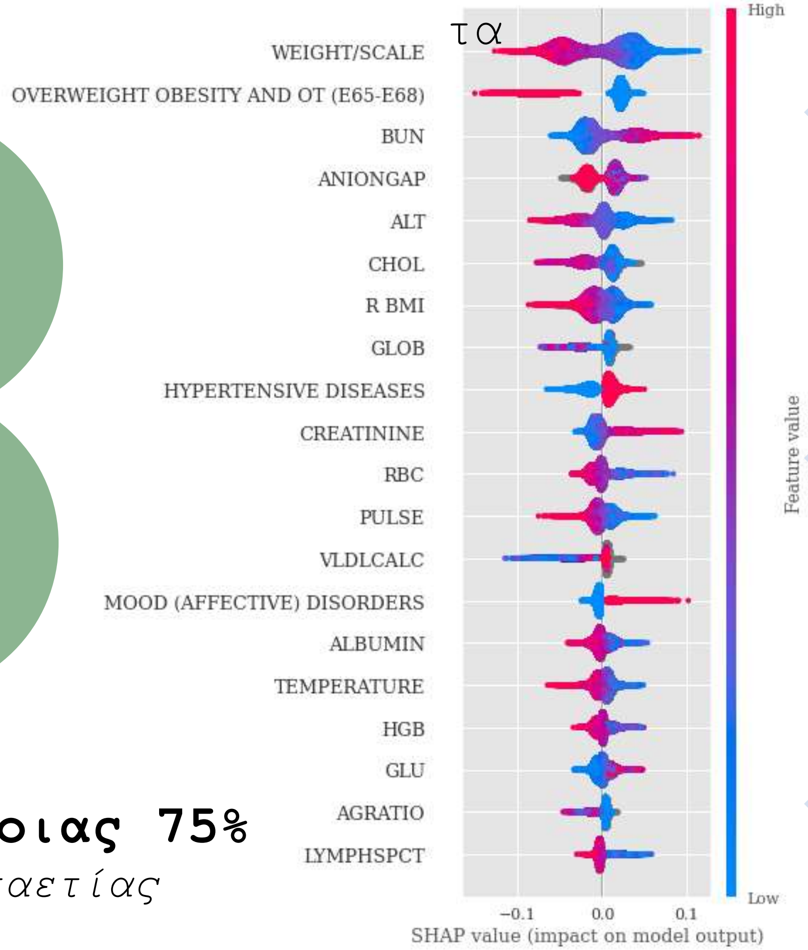
Φυσιολογική **TotalProtein**

Χαμηλά **Λευκά**

Πιθανότητα άνοιας 75%

σε εύρος πενταετίας

Αποτελέσματα



- Το υπέρβαρο σχετίζεται με μειωμένο κίνδυνο άνοιας, ενώ το χαμηλό σωματικό βάρος ενδέχεται να αυξάνει τον κίνδυνο

Υψηλό vs. **Χαμηλό** σωματικό βάρος

- Χαμηλός ΔΜΣ, σωματικό βάρος, ALT, χοληστερόλη και αριθμός ερυθρών αιμοσφαιρίων, σε συνδυασμό με υψηλότερα επίπεδα BUN και κρεατινίνης, σχετίζονται με αυξημένο κίνδυνο εμφάνισης άνοιας

Μεταβολικοί και **οργανοσυστημικοί** βιοδείκτες

- **Αγγειακοί** παράγοντες όπως η υπέρταση και παράγοντες **ψυχικής υγείας** όπως διαταραχές διάθεσης και συναισθήματος, σχετίζονται με αυξημένο κίνδυνο εμφάνισης άνοιας

Κλινικοί βιοδείκτες **συννοσηρότητας**

JOHNS HOPKINS MEDICINE SCHOOL OF MEDICINE

research-day ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

HELLENIC REPUBLIC IONIAN UNIVERSITY 1984

BiHELab

Five-year dementia prediction and decision support system based on real-world data

Κανόνες Απόφασης *Μηχανικής Μάθησης με εφαρμογές στην* ***Κλινική Υποστήριξη***

Κλινική Αξία	• Αποτελεσματική λύση που συνδυάζει σαφήνεια αποτελέσματος, ταχύτητα και ακρίβεια στην υποστήριξη κλινικής απόφασης
Αντίκτυπος στη φροντίδα ασθενών	• Υποστηρίζει τον έγκαιρο εντοπισμό ασθενών υψηλού κινδύνου, αυξάνοντας την αποτελεσματικότητα των παρεμβάσεων και της αποδοτικής διαχείρισης πόρων
Ενσωμάτωση	• Τα **μονοπάτια λήψης αποφάσεων** λειτουργούν ως πλαίσιο αναφοράς για τη αξιοποίηση των δεδομένων των ιατρικών φακέλων σε έγκαιρες κλινικές παρεμβάσεις, γεφυρώνοντας το κενό ανάμεσα στην **προγνωστική ανάλυση** και την **κλινική πράξη**
Ευρύτερη Αξία	• Αναδεικνύει ότι οι κανόνες απόφασης που εξάγονται από **αλγόριθμους μηχανικής μάθησης** μπορούν να υποστηρίξουν τη διαχείριση πολύπλοκων και πολυπαραγοντικών νοσημάτων, όπως η άνοια, χωρίς όμως να υποκαθιστούν τον ρόλο της κλινικής κρίσης