



**Δράση «Εμβληματικές δράσεις σε διαθεματικές επιστημονικές περιοχές με ειδικό ενδιαφέρον για την σύνδεση με τον παραγωγικό ιστό» ID 16618**

Εθνικό δίκτυο έρευνας για την ανάδειξη της γενετικής βάσης των νευροεκφυλιστικών νόσων Alzheimer και Parkinson, την ανίχνευση αξιόπιστων βιοδεικτών και την ανάπτυξη καινοτόμων υπολογιστικών τεχνολογιών και θεραπευτικών στρατηγικών στη βάση της ιατρικής ακριβείας (BRAIN PRECISION, TAEDR-0535850)

**ΤΙΤΛΟΣ ΠΑΡΑΔΟΤΕΟΥ: Ανάπτυξη τεχνολογιών Μηχανικής Μάθησης για εύρεση βιοδεικτών και κατασκευή μοντέλων**

**ΕΝΟΤΗΤΑ ΕΡΓΑΣΙΑΣ 5: Ανάπτυξη νέων υπολογιστικών μοντέλων και τεχνολογιών για την έγκαιρη διάγνωση των νευροεκφυλιστικών νόσων Alzheimer και Parkinson και των πρόδρομων μορφών τους.**

**ΥΠΕΥΘΥΝΗ ΕΡΕΥΝΗΤΙΚΗ ΟΜΑΔΑ (ΦΟΡΕΑΣ): ΙΩΑΝΝΗΣ ΤΣΑΜΑΡΔΙΝΟΣ (ΙΤΕ)**

## Ανάπτυξη τεχνολογιών Μηχανικής Μάθησης για εύρεση βιοδεικτών και κατασκευή μοντέλων

### ΠΕΡΙΓΡΑΦΗ

Η ανάλυση μοριακών και γενετικών δεδομένων σχετικά με νευροεκφυλιστικές ασθένειες, όπως Parkinson και Alzheimer αναδύει ένα πλήθος από προκλήσεις. Τα δεδομένα συνήθως έχουν πολύ μικρό δείγμα, γεγονός που δημιουργεί δυσκολίες από στατιστικής απόψεως. Επίσης, το πλήθος των μετρούμενων ποσοτήτων μπορεί να φτάσει και τα εκατομμύρια, όπως στην περίπτωση μετρήσεων πολυμορφισμών ενός νουκλεοτιδίου (Single Nucleotide Polymorphisms) και GWAS μελετών. Η εύρεση του ελάχιστου πλήθους από βιοδείκτες που είναι ικανά να διαγνώσουν, προγνώσουν ή προβλέψουν ένα αποτέλεσμα ενδιαφέροντος (outcome) (πρόβλημα επιλογής μεταβλητών στην επιστήμη υπολογιστών), όπως την εμφάνιση της νόσου, είναι ένα συνδυαστικό πρόβλημα υψηλής χρονικής πολυπλοκότητας που απαιτεί βασική έρευνα σε υπολογιστικές μεθοδολογίες και αντίστοιχους αλγορίθμους. Η εύρεση αξιόπιστων, και όσο το δυνατόν λιγότερων, απαιτούμενων βιοδεικτών με υπολογιστικές μεθόδους από δεδομένα είναι απαραίτητη στην Ιατρική Ακριβείας και εξατομικευμένη ιατρική. Εστιάζει περαιτέρω βιολογικά πειράματα μόνο στους απαραίτητους βιοδείκτες, αφαιρώντας όχι μόνο τους άσχετους αλλά και τους περιττούς βιοδείκτες για την πρόβλεψη. Το εργαστήριό μας έχει 20ετή εμπειρία σε τέτοιους αλγορίθμους που θα αναπτύξει περαιτέρω ώστε να μπορούν να εφαρμοστούν στο είδος και μέγεθος των δεδομένων που θα παραχθούν στο έργο.

Η ομάδα μας εργάστηκε στην ανάπτυξη και εφαρμογή νέων υπολογιστικών μοντέλων που προκύπτουν από την ανάλυση μοριακών προφίλ με χρήσεις τεχνολογιών Automated Machine Learning (AutoML). Συγκεκριμένα χρησιμοποιήσαμε το εργαλείο JADBio (Just Add Data Bio, jadbio.com) για την κατασκευή βέλτιστων προβλεπτικών μοντέλων και την εύρεση ενός συνόλου από στατιστικά ισοδύναμα σύνολα βιοδεικτών, το κάθε ένα από τα οποία οδηγεί σε ένα βέλτιστο προβλεπτικό ή διαγνωστικό μοντέλο. Δοκιμάσαμε εκατοντάδες αλυδίδες αλγορίθμων μηχανικής μάθησης, που περιλαμβάνουν διαφορετικούς αλγορίθμους για την προεπεξεργασία των δεδομένων, την επιλογή βιοδεικτών (feature selection) και την μοντελοποίηση (classification), επιλέγοντας το καλύτερο με βάση την προβλεπτική τους ικανότητα. Το JADBio αυτόματα εκτελεί πρωτόκολλα εκτίμησης της απόδοσης του καλύτερου μοντέλου με το cross-validation όλων των αλυσίδων μάθησης και την αφαίρεση του στατιστικού bias με τη χρήση του αλγορίθμου Bootstrap Bias Correction CV. Τα αποτελέσματα δοκιμάστηκαν και η χρήση του πρωτοκόλλου ανάλυσης κρίθηκε ικανοποιητική για περαιτέρω εφαρμογή σε δεδομένα σχιζοφρένειας. Η εργασία αυτή δημοσιεύτηκε στο περιοδικό Clinical Epigenetics με αναφορά:

Panagoroulou M, Panou T, Gkountakos A, Tarapatzi G, Karaglanis M, Tsamardinos I, Chatzaki E. BRCA1 & BRCA2 methylation as a prognostic and predictive biomarker in cancer: Implementation in liquid biopsy in the era of precision medicine. Clin Epigenetics. 2024 Dec 6;16(1):178. doi: 10.1186/s13148-024-01787-8.

### ΑΝΑΛΥΣΗ ΤΗΣ ΠΟΡΕΙΑΣ ΥΛΟΠΟΙΗΣΗΣ ΤΗΣ ΕΕ – ΑΝΑΜΕΝΟΜΕΝΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Εφαρμογή τεχνολογίας αιχμής Μηχανικής Μάθησης, τεχνολογία των καλύτερων πρακτικών (best practices) καθώς και τους νέους αλγορίθμους που θα αναπτύξουμε στην ανάλυση δεδομένων Parkinson και Alzheimer. Φυσικά, η πρώτη προτεραιότητα θα είναι η ανάλυση των



δεδομένων που θα παραχθούν στο έργο. Ταυτόχρονα, αν αυτό κρίνεται βοηθητικό από τους ειδικούς στον τομέα, θα αναλύσουμε δημόσια δεδομένα αυτών των ασθενειών ώστε να αντιπαραβάλλουμε τα αποτελέσματα με αυτά από τα νέα δεδομένα. Τα προβλεπτικά και διαγνωστικά μοντέλα καθώς και οι βιοδείκτες που θα εντοπιστούν θα ερμηνευθούν ως προς την βιολογική τους σημασία σε συνεργασία με τους ειδικούς ιατρούς που συμμετέχουν στο έργο.

Εφαρμόσαμε το παραπάνω πρωτόκολλο ανάλυσης δεδομένων σε επιγενετικά προφίλ αίματος ασθενών σχιζοφρένειας με εξαιρετικά αποτελέσματα. Τα αποτελέσματα έχουν δημοσιευτεί στο περιοδικό Translational Psychiatry του Nature:

Karaglani, M., Agorastos, A., Panagopoulou, M. et al. A novel blood-based epigenetic biosignature in first-episode schizophrenia patients through automated machine learning. Transl Psychiatry 14, 257 (2024). <https://doi.org/10.1038/s41398-024-02946-4>.

