



**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 βιοπληροφορικής και αναλυτικής δεδομένων από διαφορετικές τεχνολογικές πλατφόρμες**

**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): ΠΑΝΑΓΙΩΤΗΣ ΒΛΑΜΟΣ (ΙΠ)**



## Ανάπτυξη τεχνολογιών βιοπληροφορικής και αναλυτικής δεδομένων από διαφορετικές τεχνολογικές πλατφόρμες

### 1. Εισαγωγή στις τεχνολογίες βιοπληροφορικής και αναλυτικής δεδομένων στη νόσο Πάρκινσον

Η νόσος του Πάρκινσον είναι μια πολύπλοκη νευροεκφυλιστική ασθένεια, τα αίτια της οποίας δεν έχουν διασαφηνιστεί πλήρως, με την κληρονομικότητα να διαδραματίζει έναν μικρό ρόλο στην εκδήλωση της νόσου (*Scholz et al. 2012*). Ο ρόλος της βιοπληροφορικής αφορά αυτή τη στιγμή στην ανάδειξη και συλλογή μοριακών αποδείξεων και μαρτύρων, οι οποίοι απαρτίζουν τη βιομοριακή και βιοχημική υπόσταση της νόσου, λαμβάνοντας υπόψη κληρονομικούς και επιγενετικούς παράγοντες με στόχο την καθιέρωση βιοδεικτών και θεραπευτικών στόχων. Εργαλεία της βιοπληροφορικής χρησιμοποιούν διάφορα είδη πληροφορίας. Πληροφορίες όπως δεδομένα μεταγραφώματος το οποίο εξήχθη από σωματικούς ιστούς (*Planken et al. 2017*) ή προϊόντα αίματος (*X. Wang et al. 2022*) και απεικονιστικούς ελέγχους (*Salvatore et al. 2014*) είναι μερικά παραδείγματα πηγών, από τις οποίες ερευνητές προσπαθούν να αποκομίσουν χρήσιμες πληροφορίες. Εργαλεία με κομβικό ρόλο είναι προγράμματα και βιβλιοθήκες στατιστικής ανάλυσης και αλγόριθμοι μηχανικής μάθησης αλλά ακόμη και μοντέλα βαθιάς μάθησης (*W. Wang et al. 2020*).

### 2. Βιοπληροφορική μελέτη και ανάλυση δεδομένων σε δείγματα αλληλούχισης μεταγραφώματος ολικού αίματος ατόμων με νόσο Πάρκινσον

Χρησιμοποιήθηκαν δεδομένα RNA από το Project 133 IR3 με τελευταία έκδοση την 4η Φεβρουαρίου 2021, σύμφωνα με την πλατφόρμα του PPMI. Η αλληλούχιση του μεταγραφώματος έγινε με τη χρήση δειγμάτων ολικού αίματος, τα οποία περισυλλέχτηκαν από νοσούντες και υγιείς, σχηματίζοντας έτσι τις ομάδες της μελέτης. Το σύνολο των δειγμάτων καθώς και αρχείο CSV το οποίο περιλαμβάνει μέτα-πληροφορίες αναφορικά με τα δείγματα ανακτήθηκαν μέσω Download μετά από εγγραφή στις σελίδες του παρόχου δεδομένων της μελέτης του PPMI, IDA. Το συμπιεσμένο αρχείο περιλαμβάνει μεμονωμένα αρχεία για κάθε δείγμα. Υπάρχουν αρχεία με τιμές έκφρασης με κανονικοποίηση TPM (*Zhao et al. 2021*) και με μετατροπή σε μορφή feature counts (*Liao et al. 2014*).

Στην παρούσα εργασία έγινε χρήση των feature counts μιας και η κανονικοποίηση σε TPM θα περιόριζε την καταλληλότητα των δεδομένων ως προς την ανάλυση, ειδικά σε μεθόδους όπως την ανάλυση διαφορικής έκφρασης. Αυτό οφείλεται στο γεγονός ότι η συγκεκριμένη κανονικοποίηση γενικά λειτουργεί σε επίπεδο δείγματος, καθιστώντας τα δείγματα

ακατάλληλα για μια μεταξύ τους σύγκριση. Ιδανικά, τα πακέτα λογισμικού ανάλυσης γονιδιακής έκφρασης όπως DESeq απαιτούν τα δεδομένα σε μορφή RAW, δηλαδή σε ακατέργαστη μορφή, χωρίς κάποια κανονικοποίηση. Δεδομένου των προϋποθέσεων για μια ανάκτηση των αντίστοιχων πληροφοριών, όπως αναφέρθηκε και παραπάνω, η μορφή feature counts φάνηκε ως η πιο κατάλληλη, καθώς είναι ότι πιο κοντινό σε ακατέργαστα δεδομένα μεταγραφικής πληροφορίας (Liao *et al.* 2014).

Τα δεδομένα της μελέτης του PPMI περιέχουν ξεχωριστά αρχεία ανά δείγμα. Για την ανάλυση είναι χρήσιμη η διάθεση ενός μοναδικού συνόλου με όλα τα δείγματα, τις τιμές έκφρασης ανά δείγμα αλλά και μετα-πληροφορίες όπως σε ποια ομάδα, ελέγχου ή νοσούντων, ανήκει ένα δείγμα, το φύλο, η ηλικία κ.ά. για τη συγκρότηση ενός τέτοιου συνόλου συγγράφηκε το σενάριο στον κώδικα 1 και το σενάριο στον κώδικα 2. Στο πρώτο συλλέγονται τα αρχεία-δείγματα και καταγράφονται σε ένα CSV αρχείο, ενώ στο δεύτερο γίνεται ανάγνωση αυτού του αρχείου, επικύρωση των μετα-δεδομένων ανά δείγμα και εκκαθάριση του συνόλου εφαρμόζοντας φίλτρο σε στήλη η οποία ορίζει αν ένα δείγμα πληροί τα ποιοτικά κριτήρια του πειράματος ή όχι. Σαφώς στο σενάριο γίνεται εκκαθάριση όσων δειγμάτων δεν φέρουν την τιμή "pass" στην αντίστοιχη στήλη.

Για την καλύτερη διαχείριση, σε οποιοδήποτε αναλυτικό σκέλος της εργασίας, θεωρήθηκε σκόπιμη η συγκρότηση των δεδομένων σε αντικείμενο AnnData (Virshup *et al.*, 2024) καθώς και η αποθήκευση σε αρχείο τύπου h5ad. Ακολουθεί μια ιεραρχική δομή όπου παρέχεται εύκολη πρόσβαση σε παρατηρήσεις (*obs*) και σε μεταβλητές (*var*) αλλά και σε άλλες συλλογές πληροφορίας, όπως κανονικοποιημένες τιμές έκφρασης και πίνακες αντιστοίχισης με σύμβολα γονιδίων. Επιπλέον, οι μεμονωμένες ηλικίες ανά άτομο συγκροτήθηκαν στις ηλικιακές ομάδες 30-50, 50-70, 70-80 και >80. Το σύνολο των δεδομένων που επιλέχθηκε περιέχει μερικές αξιοσημείωτες προκλήσεις για την ανάλυση. Αφενός, αυτές προκύπτουν από τη φύση των δειγμάτων αίματος, όπου η ποικιλία μεταγραφικού προϊόντος είναι πιθανόν να εμπεριέχει και βιολογικό θόρυβο, ο οποίος δύναται να υπερκαλύπτει τυχόν σχετική με τη νόσο του Πάρκινσον πληροφορία. Αφετέρου, η ηλικία και το φύλο μπορεί να συμβάλουν σημαντικά σε μία διαφοροποίηση των δειγμάτων με χρήση στατιστικών μεθόδων ή αλγορίθμων μηχανικής μάθησης. Επιπλέον, τα άτομα που συμμετείχαν στη μελέτη κλήθηκαν σε επανειλημμένη δειγματοληψία, συνολικά σε πέντε επισκέψεις, σε απόσταση δύο μηνών από την προηγούμενη. Εξ αυτού κρίθηκε αναγκαία η στρωματοποίηση (*Stratification*) των δειγμάτων σύμφωνα με το φύλο, την ηλικία αλλά και τις επισκέψεις.

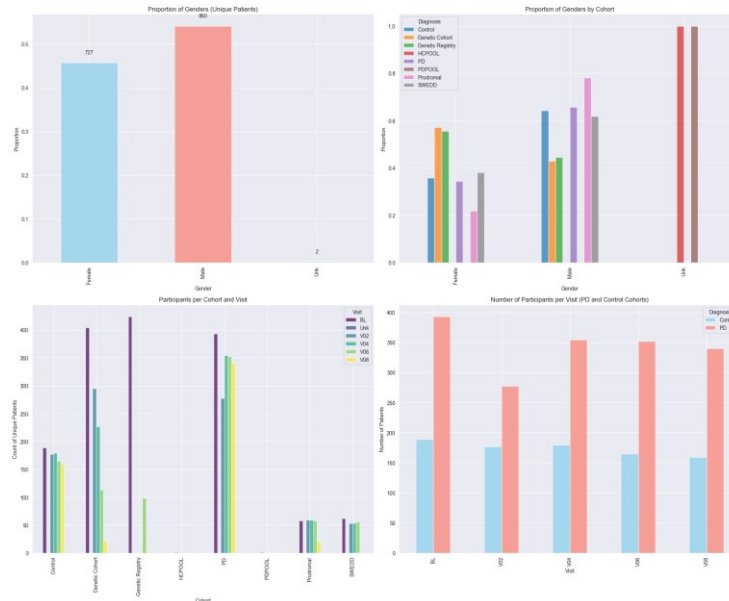
Με στόχο την λεπτομερή ανάλυση της γονιδιακής έκφρασης για την εξερεύνηση πιθανών λειτουργικών δικτύων, συγκροτήθηκε πρωτόκολλο ανάλυσης αποτελούμενο από:

- Την ανάλυση κυρίων συνιστωσών PCA, για την εξερεύνηση των δεδομένων, σύμφωνα με διαφορετικούς παράγοντες που πιθανώς επηρεάζουν τις τιμές έκφρασης.
- Την ανάλυση διαφορικής έκφρασης ανά στρώμα (*stratum*), για την ανάδειξη διαφορικά εκφραζόμενων γονιδίων.
- Την εφαρμογή ευρέως διαδομένων αλγορίθμων μηχανικής μάθησης στους τομείς της βιοπληροφορικής και την εξαγωγή χαρακτηριστικών που κρίθηκαν από τον εκάστοτε αλγόριθμο ως σημαντικά για την κατηγοριοποίηση.
- Την εφαρμογή της ανάλυσης εμπλουτισμού GSEA σε γονίδια που αναγνωρίστηκαν ως διαφορικά εκφραζόμενα διασταύρωση με γονίδια, τα οποία χαρακτηρίστηκαν σημαντικά κατά την κατηγοριοποίηση.
- Την ανάλυση και κατασκευή/οπτικοποίηση πιθανών δικτύων συνέκφρασης σε γονίδια που βρέθηκαν ως διαφορικά εκφραζόμενα καθώς και από την σημαντικότητα χαρακτηριστικών των αποτελεσμάτων των αλγορίθμων μηχανικής μάθησης.
- Την αξιολόγηση των αποτελεσμάτων και τις πιθανές σχέσεις με τον φαινότυπο της νόσου του Πάρκινσον.

### 3. Αποτελέσματα

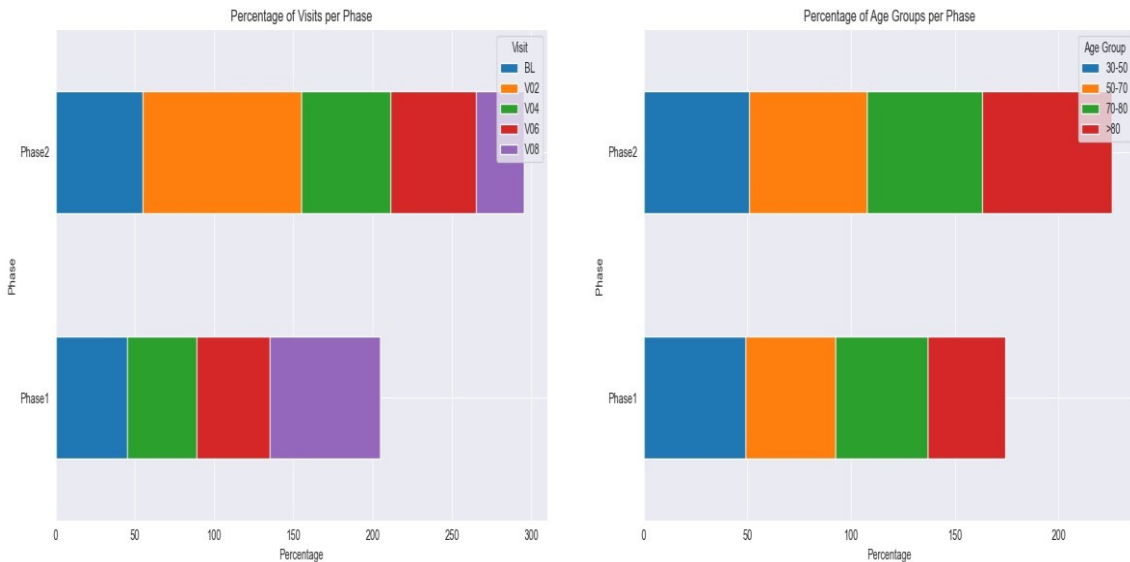
#### 3.1 Χαρακτηρισμός του συνόλου των δεδομένων

Στην εικόνα 3.1 παρουσιάζονται οι αριθμοί συμμετεχόντων από κάθε φύλο, ανεξάρτητα από την ομάδα και τον τελικό αριθμό δειγμάτων. Πλην δύο μόνο συμμετεχόντων αγνώστου φύλου, ο χαρακτηρισμός των υπολοίπων είναι πλήρης για κάθε ομάδα. Το γράφημα στα δεξιά παρουσιάζει την ποσοστιαία κατανομή των συμμετεχόντων ανά φύλο και ομάδα. Στο κάτω μέρος παρουσιάζονται οι αριθμοί συμμετεχόντων της κάθε ομάδας ανά επίσκεψη (Visit). Στην παρούσα εργασία χρησιμοποιήθηκαν μόνο δείγματα των ομάδων PD και Control. Το πλήθος συμμετεχόντων για αυτές τις δύο ομάδες παρουσιάζεται στο γράφημα κάτω δεξιά, όπου απεικονίζονται οι απόλυτοι αριθμοί ατόμων ανά ομάδα και επίσκεψη. Χαρακτηριστικό είναι ότι η ομάδα ελέγχου υστερεί σε πλήθος της ομάδας PD, γεγονός το οποίο πρέπει να λαμβάνεται υπόψη ιδιαίτερα στη μοντελοποίηση κατηγοριοποιητών μηχανικής μάθησης.



**Εικόνα Error! No text of specified style in document..1** Επάνω αριστερά: Συμμετέχοντες ανά φύλο συνολικά, Επάνω δεξιά: Συμμετέχοντες ανά ομάδα.

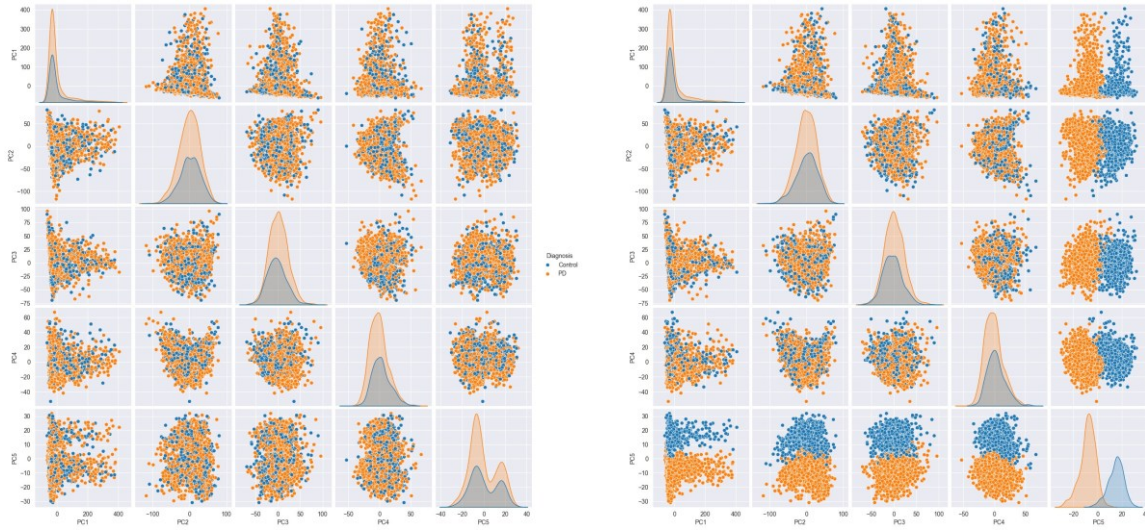
Πρέπει να σημειωθεί ότι η αλληλούχιση διενεργήθηκε σε δύο διαφορετικές φάσεις. Αυτό δεν σημαίνει πως η δειγματοληψία ακολούθησε επίσης αυτόν τον διαχωρισμό, έτσι ώστε να προκύψουν για κάθε συμμετέχοντα δέκα δείγματα, ήτοι πέντε επισκέψεις επί δύο φάσεις, αλλά δείγματα από τις πέντε (ή και λιγότερες) διαφορετικές επισκέψεις μπορεί να υποβλήθηκαν σε αλληλούχιση σε διαφορετικές φάσεις. Σαφώς αυτό καθιστά έναν τεχνικό παράγοντα που μπορεί να οδηγήσει σε ανακριβή αποτελέσματα σε αναλύσεις. Ειδικά η αλληλούχιση των δειγμάτων της δεύτερης επίσκεψης έχει γίνει αποκλειστικά στη δεύτερη φάση, σύμφωνα με την εικόνα 3.2i. Ένα επίσης ερώτημα μπορεί να απαντηθεί ως προς το αν τα δείγματα όλων των ηλικιακών ομάδων εκπροσωπούνται και στις δύο φάσεις. Με το γράφημα στην εικόνα 3.2ii η απάντηση μπορεί να δοθεί καταφατικά, όπως επίσης και ότι δεν υπάρχει αξιοσημείωτη ασυμμετρία στην κατανομή των ηλικιακών ομάδων ανά φάση αλληλούχισης.



**Εικόνα** Error! No text of specified style in document..2 Οπτικοποίηση δειγμάτων ανά φάση αλληλούχισης. (i) Κατανομή δειγμάτων ανά επίσκεψη και φάση αλληλούχισης. (ii) Κατανομή ηλικιακών ομάδων ανά φάση αλληλούχισης.

### 3.2 Ανάλυση κυρίων συνιστωσών

Στην εικόνα 3.3 παρουσιάζεται το αποτέλεσμα της PCA για τις πρώτες πέντε συνιστώσες, όπου ο χρωματισμός πραγματοποιήθηκε σύμφωνα με τη διάγνωση. Δεν διακρίνονται ξεχωριστές συστάδες, γεγονός που συνιστά πως η μεταβλητότητα των εκφράσεων δεν διαχωρίζει σε αυτό το επίπεδο ανάλυσης δείγματα νοσούντων από υγιείς. Ενδιαφέρουσα είναι η απεικόνιση κυρίων συνιστωσών με χρωματισμό με βάση το φύλο, όπου διακρίνονται συστάδες μεταξύ της πέμπτης συνιστώσας και όλων των υπολοίπων. Το συμπέρασμα είναι πως υπάρχει μεταβλητότητα σύμφωνα με γονιδιακές εκφράσεις στα δείγματα ολικού αίματος, με την οποία δύναται να διακριθούν τα δείγματα των δύο φύλων. Πρόκειται για πολύ μικρό μέρος εκ του συνόλου της μεταβλητότητας (Πίνακας 3.1). Η στρωμάτωση των δειγμάτων σε παρακάτω βήματα χρησιμεύει προς αποφυγήν της οποίας πιθανής παρεμβολής στις διάφορες αναλύσεις.

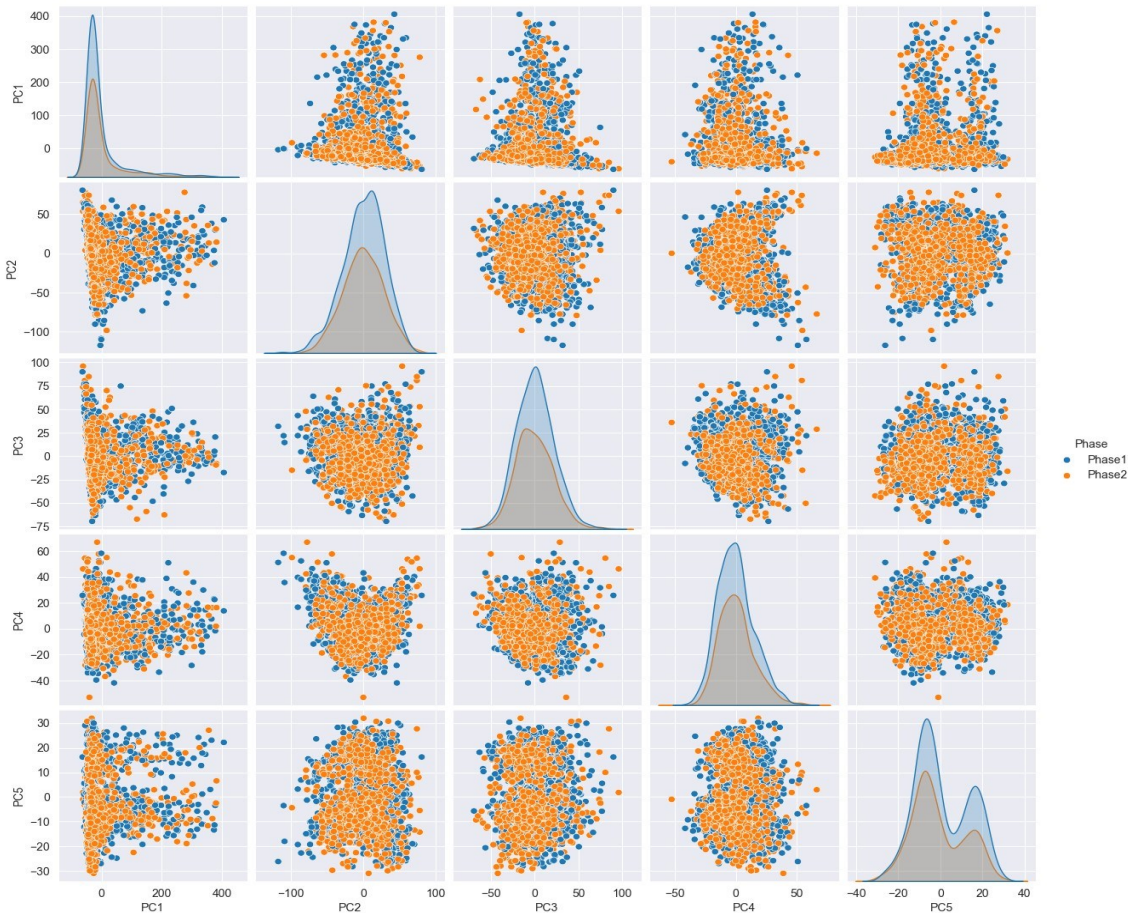


**Εικόνα** Error! No text of specified style in document..3 Αποτελέσματα PCA. (i) PCA σύμφωνα με την διάγνωση, (ii) PCA σύμφωνα με το φύλο.

**Πίνακας** Error! No text of specified style in document..1 Μεταβλητότητα ανά συνιστώσα

PC 1	PC 2	PC 3	PC 4	PC 5
56,88%	8,77%	5,15%	2,44%	1,78%

Παρακάτω απεικονίζεται το αποτέλεσμα της PCA σύμφωνα με τη φάση αλληλούχισης. Εντός των πρώτων πέντε κύριων συνιστωσών δεν παρατηρείται ομαδοποίηση και μαζί με αυτό προκύπτει και η διαβεβαίωση πως δεν εισάγεται τεχνικός θόρυβος από τη φάση αλληλούχισης.

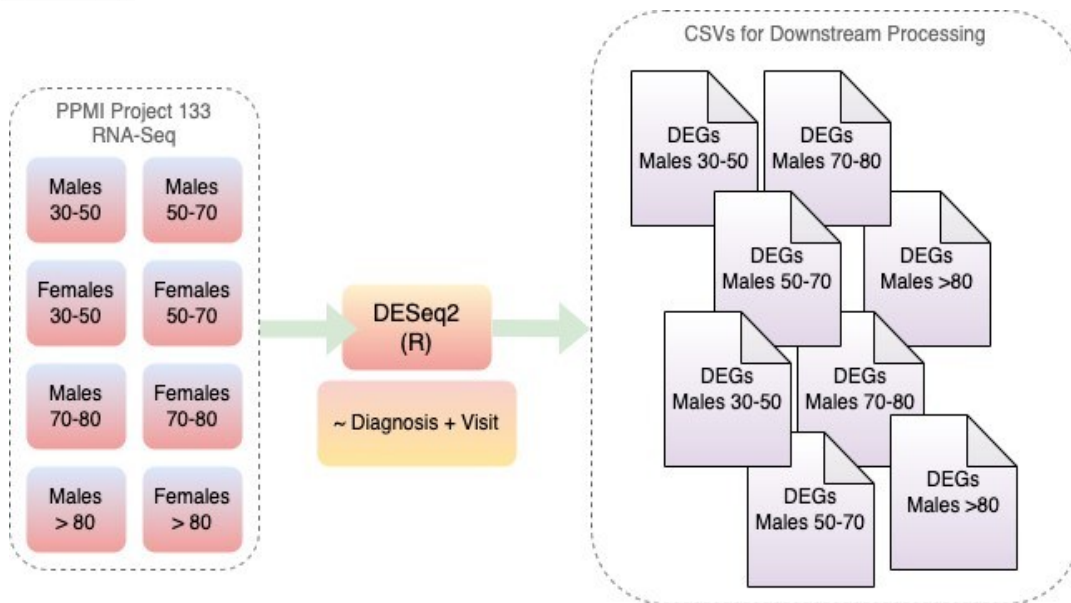


Εικόνα Error! No text of specified style in document..4. PCA σύμφωνα με τη φάση αλληλούχησης

### 3.3 Ανάλυση διαφορικής έκφρασης

Η ανάλυση διαφορικής έκφρασης γονιδίων (*DGEA*) διενεργήθηκε μετά από στρωματοποίηση των δειγμάτων σύμφωνα με το φύλο και την ηλικιακή ομάδα με τη χρήση του πακέτου λογισμικού της R, *DESeq2* (Love *et al.* 2014) και αντίστοιχη οπτικοποίηση με χρήση σεναρίου στην γλώσσα *Rython*. Η επίσκεψη της κάθε δειγματοληψίας συμπεριλήφθηκε ως συμμεταβλητή και θεωρήθηκε ως τεχνικός παράγοντας, χωρίς να ερευνηθεί αν έχει βιολογικό χαρακτήρα ως προς την πρόοδο της νόσου. Ως όριο για την έκφραση  $|\log_2\text{FoldChange}|$  τέθηκε η τιμή  $> 0,5$  ενώ για την τιμή της *padj*  $< 0,05$ .

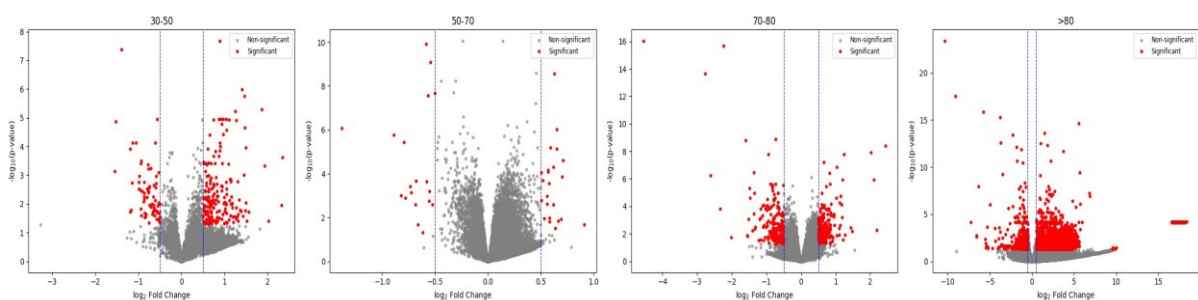
Το αντίστοιχο τμήμα της διαδικασίας ανάλυσης (*Pipeline*) παρουσιάζεται στην εικόνα 3.5. Το κομμάτι αυτό κατακερματίζει το σύνολο δεδομένων σύμφωνα με το φύλο και την ηλικιακή ομάδα των συμμετεχόντων σε υποσύνολα (*stratification*), όπου για κάθε υποσύνολο διεξάγεται η διαφορική ανάλυση μέσω *DESeq2*. Πέραν της διάγνωσης, συμπεριλήφθηκε και η παρατήρηση των επισκέψεων δειγματοληψίας (*Visit*), καθώς καθιστά πηγή μεταβλητότητας η οποία πρέπει να ληφθεί υπόψη ως τεχνικός παράγοντας στην στατιστική ανάλυση (Love *et al.* 2016). Τα αποτελέσματα περιέχουν τα διαφορικά εκφραζόμενα και καταγράφονται σε μορφή CSV σε διαφορετικά αρχεία για περαιτέρω ανάλυση.



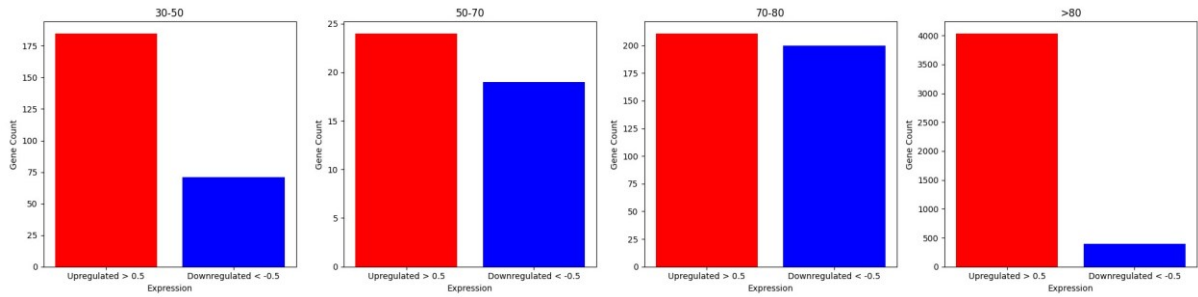
**Εικόνα** Error! No text of specified style in document..5 Pipeline για την ανάλυση διαφορικής έκφρασης με εξαγωγή αποτελεσμάτων σε αρχεία CSV.

### 3.4 Διαφορική έκφραση γονιδίων σε γυναίκες

Το πλήθος των στατιστικά σημαντικών εκφραζόμενων γονιδίων ποικίλει ανά ηλικιακή ομάδα. Στη σειρά διαγραμμάτων της εικόνας 3.6 διακρίνονται περισσότερα γονίδια στατιστικής σημασίας στην ηλικιακή ομάδα 30-50 έναντι της ομάδας 50-70 ετών. Ενδεχομένως, η νόσος να διαφοροποιείται σε γονιδιακό επίπεδο σε περιπτώσεις όπου η νόσος εκδηλώνεται σε μικρότερες ηλικίες. Σε ηλικίες άνω των 70 ετών, το πλήθος των διαφορικά εκφραζόμενων γονιδίων στατιστικής σημασίας αυξάνεται σημαντικά, κάτι που μπορεί να παραπέμπει και στην πρόοδο της νόσου με αυξανόμενη ηλικία. Όσον αφορά το πλήθος των υπέρ-έναντι των υπό-εκφραζόμενων γονιδίων, ισχύει σύμφωνα με την παρακάτω οπτικοποίηση στην εικόνα 3.7, ότι υπάρχουν περισσότερα υπέρ-εκφραζόμενα γονίδια σε όλες τις ηλικιακές ομάδες υπό παρατήρηση.



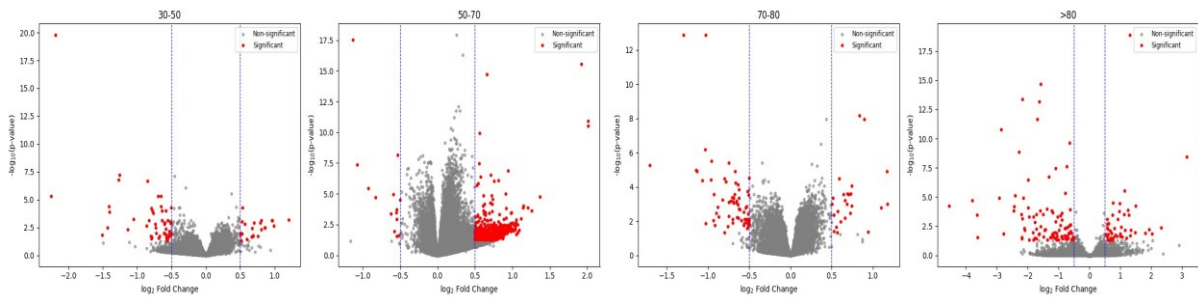
**Εικόνα** Error! No text of specified style in document..6 Διαγράμματα ηφαιστείου ανά ηλικιακή ομάδα γυναικών



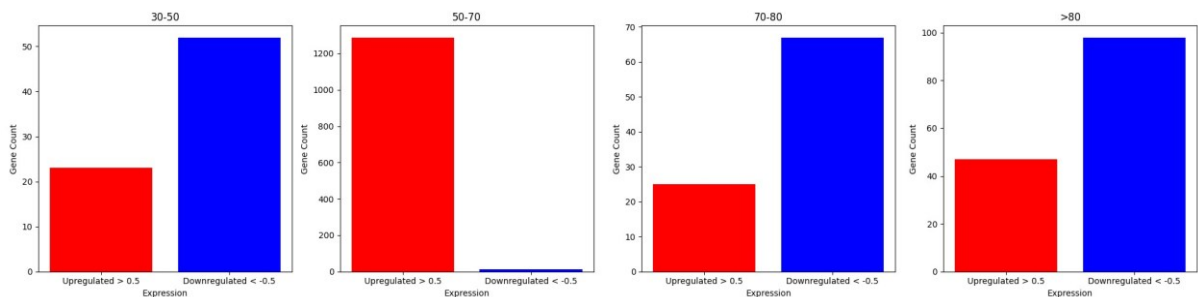
**Εικόνα Error! No text of specified style in document..7** Πλήθος στατιστικά σημαντικών διαφορικά εκφραζόμενων γονιδίων - Γυναίκες

### 3.5 Διαφορική έκφραση γονιδίων σε άνδρες

Συνολικά η διαφορική έκφραση γονιδίων στους άνδρες σε κάθε ηλικιακή ομάδα διαφέρει από τις εκφράσεις των γυναικών. Χαρακτηριστική είναι η παρουσία περισσότερων υπέρ-εκφραζόμενων γονιδίων έναντι ελάχιστων υπό-εκφραζόμενων στην ηλικιακή ομάδα των 50-70, ενώ στις υπόλοιπες περιπτώσεις ο αριθμός των υπό-εκφραζόμενων δείχνει να είναι περίπου διπλάσιος από των υπέρ-εκφραζόμενων.



**Εικόνα Error! No text of specified style in document..8** Διαγράμματα ηφαιστείου ανά ηλικιακή ομάδα ανδρών



**Εικόνα Error! No text of specified style in document..9** Πλήθος στατιστικά σημαντικών διαφορικά εκφραζόμενων γονιδίων – Άνδρες

Στην εικόνα 3.10 παρατίθενται στα εξής διαγράμματα κοινά γονίδια μεταξύ των φύλων ανά ηλικιακή ομάδα, στατιστικά σημαντικά και με τιμές έκφρασης  $|\log_2 \text{FoldChange}| > 0,5$ . Από

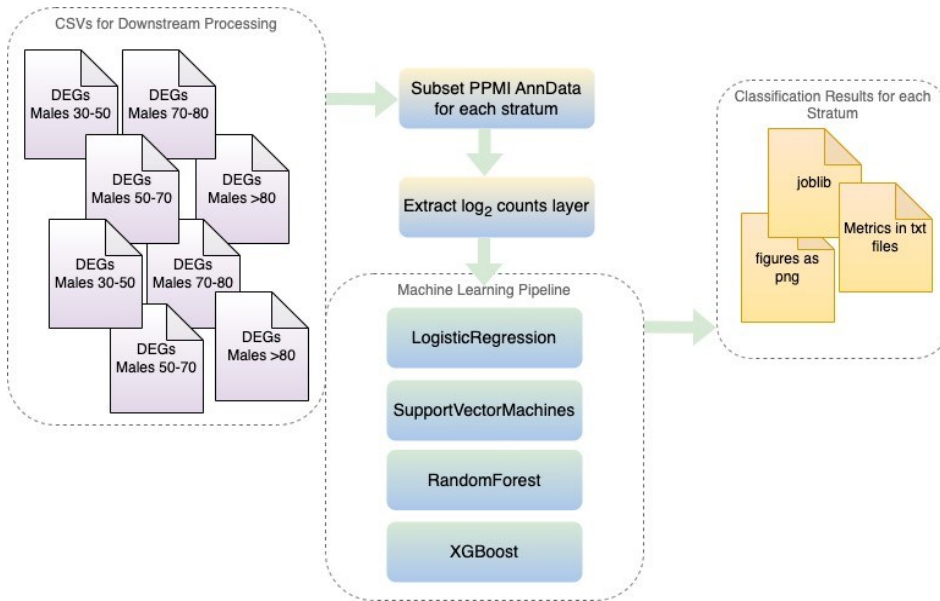


Ηλικιακή ομάδα	Γονίδιο	Εμπλοκή σε βιολογική λειτουργία
30-50	IF27	Ανοσοποιητικό - Ρύθμιση Ουβικουιτίνης ( <i>Xue et al. 2016</i> ), μονοπάτι σηματοδότησης Jak/STAT ( <i>Y. Chen et al. 2017, Lashgari et al. 2021</i> )
	MYOM2	Μυοσκελετικό - «QuickGO::Term GO:0006936», χ.η. Οντολογία σχετίζεται με μυϊκή συστολή.
50-70	-	-
70-80	RPL39	Ριβοσωμική πρωτεΐνη
	NECTIN2	Υποδοχέας ιού του απλού έρπητα ( <i>Martinez και Spear, 2001</i> )
	OLFM1	Νευρικό σύστημα - ( <i>Sultana et al., 2011, Kelly et al., 2019</i> )
	CMBL	Ένζυμο υδρόλυσης
>80	SFRP1	Νευρικό σύστημα - ( <i>Kele et al. 2012</i> )
	TXK	Ανοσοποιητικό - ( <i>Kashiwakura et al. 1999</i> )
	NECTIN2	

### 3.6 Κατηγοριοποίηση δειγμάτων μέσω Μηχανικής Μάθησης

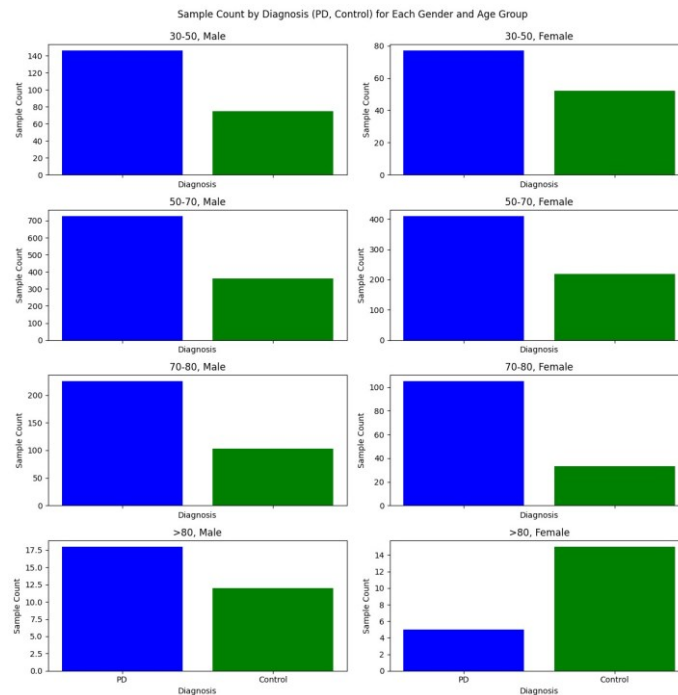
Με την χρήση των δεδομένων της διαφορικής ανάλυσης γονιδιακής έκφρασης εκπαιδεύτηκαν τέσσερις διαφορετικοί κατηγοριοποιητές εποπτευόμενης μάθησης. Χρησιμοποιήθηκαν οι αλγόριθμοι λογιστικής παλινδρόμησης, SVM, Random Forest και XGBoost. Πρόκειται για αλγόριθμους που χρησιμοποιούνται, μεταξύ άλλων, στον τομέα της βιοπληροφορικής. Η εφαρμογή της κατηγοριοποίησης πραγματοποιήθηκε σε όλα τα στρώματα (*strata*) του συνόλου δεδομένων. Ως αποτελέσματα για την αποτίμηση της επίδοσης του κάθε αλγορίθμου ανά στρώμα, χρησιμοποιήθηκαν καθιερωμένες μετρικές όπως ROC-AUC, PR-AUC καθώς και η ευαισθησία και η ειδικότητα της κατηγοριοποίησης κατά την εκτέλεση. Η εκπαίδευση λειτούργησε με τη χρήση ενός εύρους παραμέτρων (*Hyperparameters*) μέσω των οποίων ο κατηγοριοποιητής της κάθε μεθόδου δοκιμάζει και αποφασίζει τον βέλτιστο συνδυασμό παραμέτρων μετά από δεκαπλάσια διασταυρούμενη επικύρωση (*10-fold Cross Validation*). Ο έτσι παραμετροποιημένος βέλτιστος κατηγοριοποιητής εκπαιδεύεται και

μετέπειτα δοκιμάζεται σε κατηγοριοποίηση επί του αντίστοιχου συνόλου. Ανεξάρτητα από την εύρεση του καλύτερου δυνατού κατηγοριοποιητή ανά αλγόριθμο, η δεκαπλάσια διασταυρούμενη επικύρωση λαμβάνει χώρα και ξεχωριστά από την παραπάνω διαδικασία, ως μέτρο παρουσίασης της γενικής επίδοσης του κάθε μοντέλου κατηγοριοποίησης.



**Εικόνα** Error! No text of specified style in document..11 Pipeline για την κατηγοριοποίηση δειγμάτων με χρήση μηχανικής μάθησης

Τα σύνολα εκπαίδευσης και εκτέλεσης είναι διαφορετικά, με το σύνολο εκπαίδευσης να επιλέγεται στο 80% επί του συνόλου δεδομένων ανά στρώμα, όπως και επίσης δείγματα από τα ίδια άτομα δεν βρίσκονται και στα δύο σύνολα, κάτι που θα είχε ως αποτέλεσμα το φαινόμενο του Feature Leakage (Oosterhuis et al. 2024) και θα παραποιούσε σημαντικά τις μετρικές καθώς και την έκβαση της κατηγοριοποίησης. Για την επίτευξη του διαχωρισμού έγινε χρήση της κλάσης GroupShuffleSplit στην Python, από το πακέτο scikit (Buitinck et al. 2013). Παρότι τα δείγματα των νοσούντων είναι σε όλα τα στρώματα περίπου διπλάσια από των υγιών, η ανισοροπία (Class imbalance) δεν είναι αρκετά υψηλή ώστε να δημιουργήσει σημαντικό πρόβλημα, όπως την προνομαϊκή κατηγοριοποίηση με γνώμονα την πολυπληθέστερη ομάδα (εικόνα 3.12). Για το λόγο αυτό και επίσης για την αποφυγή υπέρ-αισιόδοξων αποτελεσμάτων κατηγοριοποίησης, δεν επιδιώχθηκε η χρήση μεθόδων όπως π.χ. SMOTE. Αντί αυτού, τέθηκε παράμετρος `class_weight=balanced` σε όλους τους κατηγοριοποιητές, προς όφελος της ομάδας που περιέχει τα λιγότερα δείγματα, χωρίς τη δημιουργία συνθετικών δειγμάτων.

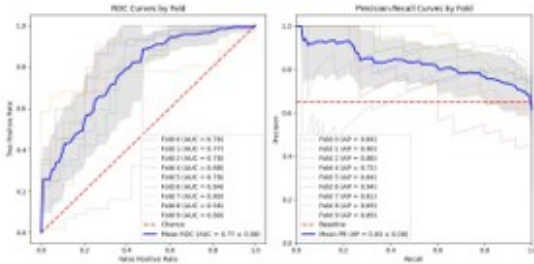


**Εικόνα** Error! No text of specified style in document..12 Στις ομάδες ανδρών 30-50, 50-70 και 70-80 παρατηρείται ανισορροπία κατά περίπου το διπλάσιο της ομάδας ελέγχου. Στις ομάδες γυναικών 50-70 επίσης, ενώ στην ηλικιακή ομάδα 70-80 περίπου το τριπλάσιο. Η ομάδα γυναικών στις ηλικίες άνω των 80 δείχνει σχεδόν τριπλάσια δείγματα ελέγχου αντί νοσούντων.

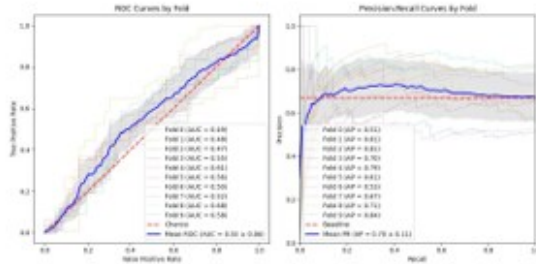
### 3.7 Επιδόσεις κατηγοριοποίησης

Οι επιδόσεις της κατηγοριοποίησης ποικίλουν ανά στρώμα, δηλαδή τον συνδυασμό φύλου και ηλικιακής ομάδας. Η πιο αντιπροσωπευτική ομάδα από άποψη του πλήθους των διαθέσιμων δειγμάτων παρουσιάζεται παρακάτω. Τα αποτελέσματα των λοιπών ομάδων παρατίθενται στο παράρτημα της εργασίας. Αρχικά, παρατηρούνται οι μετρικές της δεκαπλάσιας διασταυρούμενης επικύρωσης των διαφόρων αλγορίθμων κατηγοριοποίησης στην εικόνα 3.13. Απεικονίζονται οι καμπύλες ROC-AUC ως παράσταση των αληθώς θετικών προβλέψεων έναντι των ψευδώς θετικών και η PR-AUC ως παράσταση της ακρίβειας και της ευαισθησίας ή ανάκλησης. Τα αποτελέσματα όλων των κατηγοριοποιητών για το γυναικείο φύλο κυμαίνονται μεταξύ της χαμηλότερης τιμής 0,72 και της υψηλότερης 0,77 για τη μετρική ROC-AUC, όπου η χαμηλότερη μέση τιμή αποδίδεται από τον αλγόριθμο XGBoost και η υψηλότερη από τον αλγόριθμο της λογιστικής παλινδρόμησης. Αντίστοιχα, η μετρική PR-AUC βρίσκεται μεταξύ χαμηλού 0,81 από τους αλγόριθμους Random Forest και XGBoost και 0,83 από την λογιστική παλινδρόμηση και SVM. Στους άνδρες οι υψηλότερες μέσες τιμές για ROC-AUC και PR-AUC αποδίδονται από τον αλγόριθμο XGBoost με 0,73 και 0,84 αντίστοιχα. Η χαμηλότερη επίδοση αποδίδεται από την λογιστική παλινδρόμηση με 0,55 για ROC-AUC και

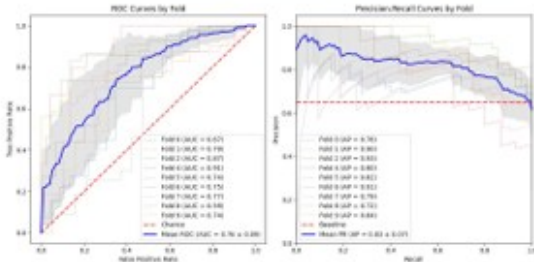
0,70 για PR-AUC. Η χαμηλή ROC-AUC καθιστά την επίδοση της λογιστικής παλινδρόμησης ελαφρώς καλύτερη από ένα μοντέλο τυχαίων προβλέψεων, λόγω της εγγύτητας στο κατώφλι αξιολόγησης των 50%. Η βελτιωμένη επίδοση του XGBoost πιθανώς μπορεί να αποδοθεί στο μεγαλύτερο πλήθος δειγμάτων και με αυτό, μεγαλύτερο πλήθος δεδομένων μάθησης σε σύγκριση με το σύνολο των γυναικών, παρόλο που η ανισορροπία και στις δύο περιπτώσεις κυμαίνεται σε αναλογία 2 προς 1.



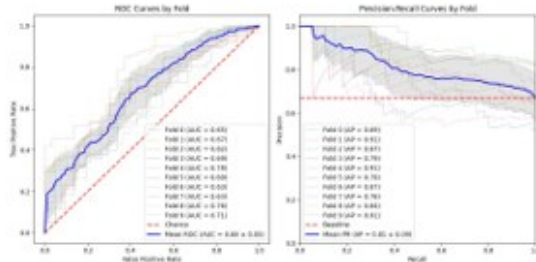
(i) Γυναίκες: Λογιστική παλινδρόμηση



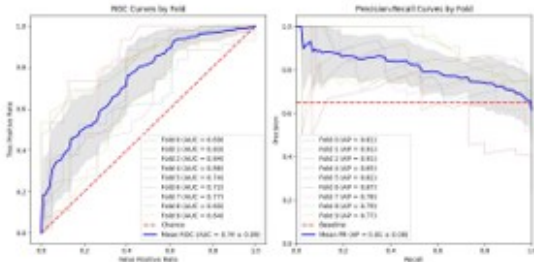
(ii) Άνδρες: Λογιστική παλινδρόμηση



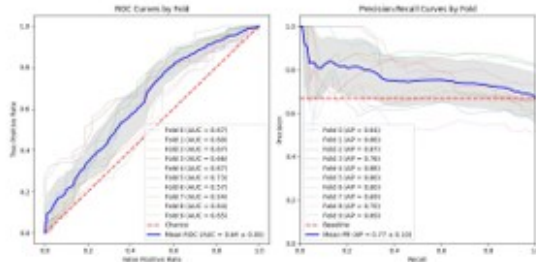
(iii) Γυναίκες: Support Vector Machine



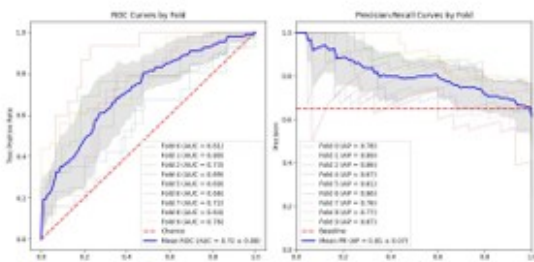
(iv) Άνδρες: Support Vector Machine



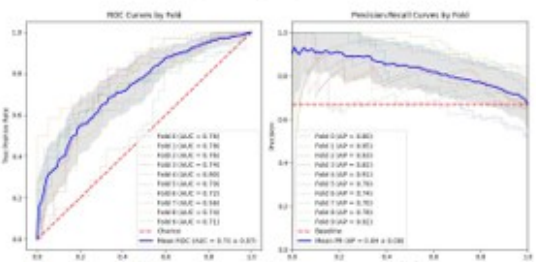
(v) Γυναίκες: Random Forest



(vi) Άνδρες: Random Forest



(vii) Γυναίκες: XGBoost



(viii) Άνδρες: XGBoost

**Εικόνα** Error! No text of specified style in document..13 Αποτελέσματα δεκαπλάσιας δισταυρούμενης επικύρωσης

Ο πίνακας 3.4 περιέχει μετρικές όπως την ακρίβεια (*Precision*), την ανάκληση ή ευαισθησία (*Recall*) για κάθε κλάση ταξινόμησης (*HC=ομάδα υγείων/ελέγχου, PD=ομάδα νοσού Πάρκινσον*) και τις τιμές ROC-AUC και PR-AUC. Στις στήλες βρίσκονται οι αλγόριθμοι κατηγοριοποίησης (*LR=Λογιστική Παλινδρόμηση, SVM=Support Vector Machine, RF=Random Forest, XGB=XGBoost*) κάτω από τους οποίους καταγράφονται τα αποτελέσματα για άνδρες και γυναίκες (*M=Male, F=Female*).

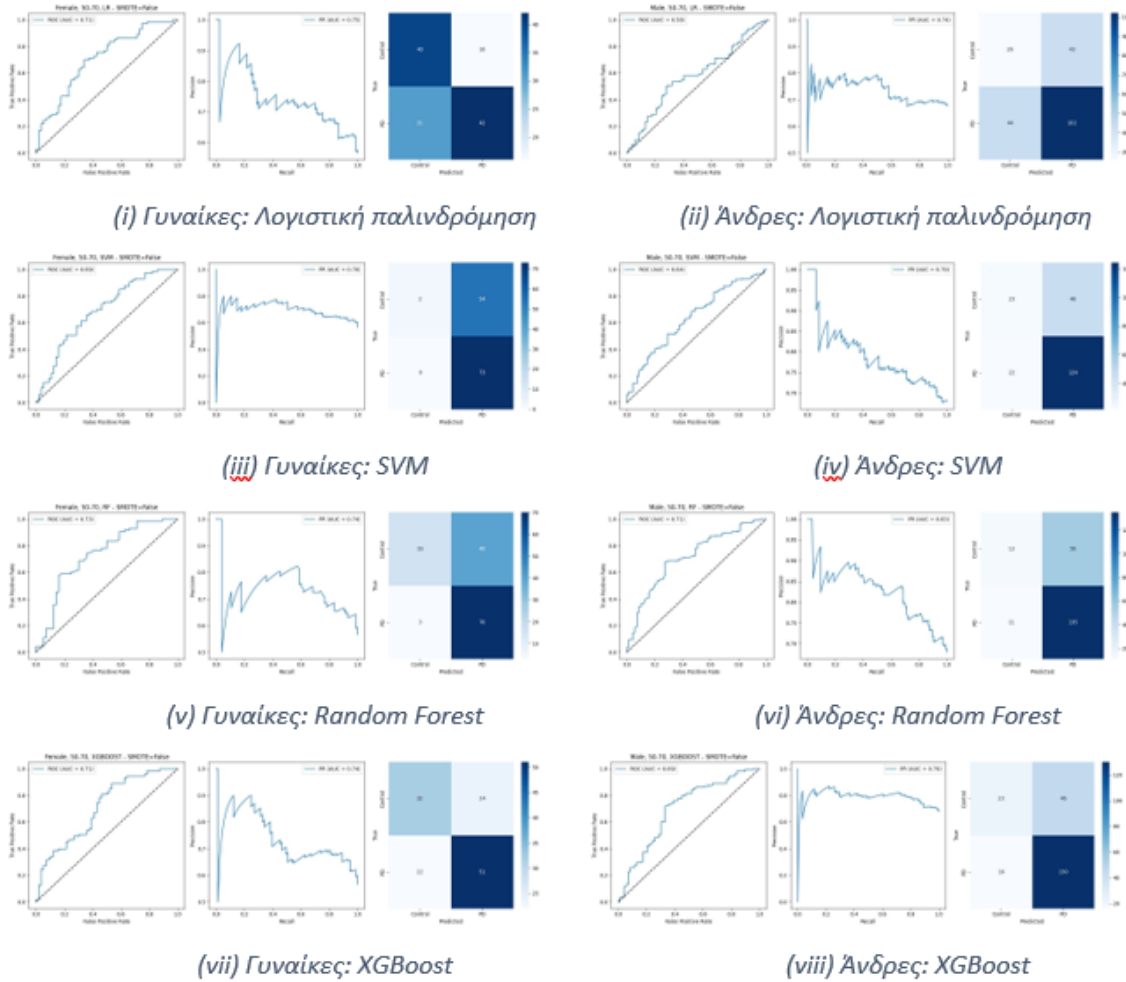
Η τιμή της ακρίβειας ορίζεται από το λόγο των αληθώς θετικών προς το σύνολο των θετικών προβλέψεων και καταγράφει την επίδοση του κατηγοριοποιητή να διακρίνει περιπτώσεις θετικών δειγμάτων ως προς την κλάση αναφοράς, που όπως απεικονίζεται στον πίνακα 3.3 η τιμή υπολογίζεται τόσο για την ομάδα νοσούντων όσο και για την ομάδα ελέγχου. Η τιμή της ευαισθησίας ή ανάκλησης αφορά, όπως και η ακρίβεια, κάθε κλάση στον παρακάτω πίνακα και υπολογίζεται ως λόγος αληθώς θετικών προβλέψεων προς το σύνολο των προβλέψεων.

**Πίνακας** Error! No text of specified style in document..3 Αναλυτικές μετρικές επίδοσης κατηγοριοποίησης

		LR		SVM		RF		XGB	
		M	F	M	F	M	F	M	F
Precision	HC	0,37	0,56	0,52	0,68	0,50	<b>0,80</b>	0,59	0,59
	PD	<b>0,70</b>	<b>0,72</b>	<b>0,75</b>	0,69	<b>0,70</b>	0,65	<b>0,74</b>	0,68
Recall	HC	0,38	<b>0,71</b>	0,45	0,54	0,13	0,36	0,33	0,57
	PD	<b>0,70</b>	0,58	<b>0,80</b>	<b>0,81</b>	<b>0,94</b>	<b>0,93</b>	<b>0,89</b>	<b>0,70</b>
ROC-AUC		0,582	0,707	0,641	0,689	0,714	0,732	0,694	0,707
PR-AUC		0,739	0,747	0,790	0,703	0,828	0,737	0,790	0,743

Η ευαισθησία του μοντέλου Random Forest βρίσκεται σε εξαιρετικά υψηλά επίπεδα και για τα δύο φύλα. Υψηλές τιμές ευαισθησίας παρουσιάζονται επίσης από το μοντέλο SVM. Ο αλγόριθμος XGBoost παρουσιάζει υψηλό σκορ ανάκλησης για άνδρες και μειωμένο για τις γυναίκες. Αντίστοιχα, στη λογιστική παλινδρόμηση για τους άνδρες η ευαισθησία βρίσκεται σε ένα ικανοποιητικό επίπεδο των 70%. Με αυτές τις τιμές ευαισθησίας, τα προαναφερόμενα μοντέλα κατηγοριοποίησης παρουσιάζουν σχετικά ικανοποιητική επίδοση αναφορικά με την ανίχνευση θετικών ως προς την νόσο του Πάρκινσον. Ως προς την ακρίβεια, η λογιστική παλινδρόμηση παρουσιάζει τιμές  $\geq 70\%$  και στα δύο φύλα. Η ακρίβεια στους αλγόριθμους Random Forest και XGBoost βρίσκεται για τους άνδρες στο 70% και 74% αντίστοιχα. Η υψηλότερη τιμή ακρίβειας ως προς την ομάδα ελέγχου αποδίδεται από το μοντέλο Random

Forest με 80% για τις γυναίκες. Οι καμπύλες ROC-AUC, PR-AUC και η αντίστοιχη μήτρα σύγχυσης παρουσιάζονται στην εικόνα 3.14.



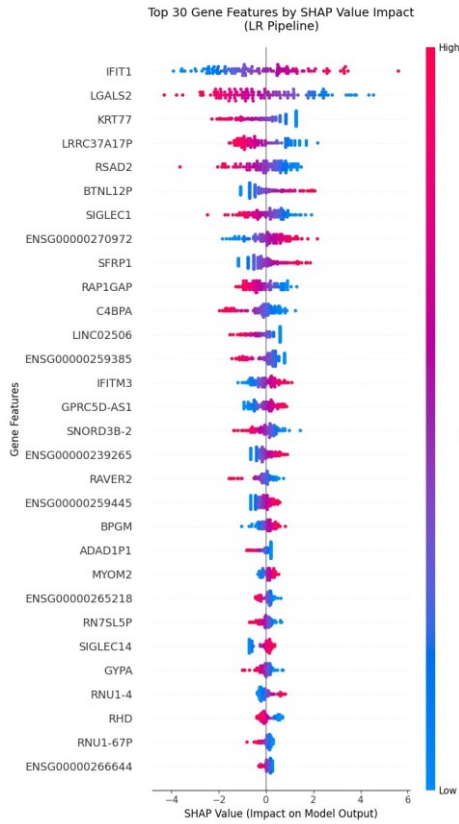
**Εικόνα** Error! No text of specified style in document..14 Αποτελέσματα κατηγοριοποίησης ανά φύλο.

### 3.8 Αποτίμηση των αποτελεσμάτων της κατηγοριοποίησης

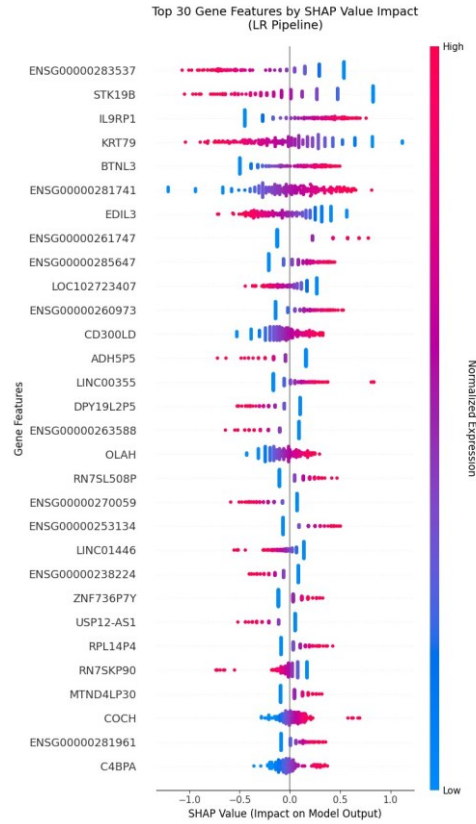
Μετά την εκτέλεση της κατηγοριοποίησης προκύπτει το ερώτημα ποια χαρακτηριστικά θεώρησαν τα μοντέλα ταξινόμησης ως παράγοντες διαφοροποίησης μεταξύ των δύο κλάσεων, των υγιών και των νοσούντων ή αλλιώς, ποια είναι τα χαρακτηριστικά που συνέβαλαν περισσότερο ή και λιγότερο, στη διάκριση των νοσούντων και αντίστοιχα των υγιών. Η ερώτηση μπορεί να απαντηθεί μέσω της εξαγωγής συγκεκριμένων τιμών από τα μοντέλα κατηγοριοποίησης, τα λεγόμενα *feature importances* που αποδίδουν τιμές σημαντικότητας στα χαρακτηριστικά με βάση των οποίων έγινε η ταξινόμηση. Κάθε μοντέλο κατηγοριοποίησης περιλαμβάνει τρόπους εξαγωγής χαρακτηριστικών, στα οποία αποδίδει και αντίστοιχα σκορ, σύμφωνα με το ρόλο που διαδραμάτισαν στην ταξινόμηση. Τα σκορ που αποδίδονται μπορεί να διαφέρουν αρκετά μεταξύ κατηγοριοποιητών και σε περίπτωση όπου στόχος είναι η

σύγκριση των χαρακτηριστικών μεταξύ διάφορων μοντέλων ταξινόμησης, τότε τα σκορ σημαντικότητας πρέπει να κανονικοποιηθούν ώστε να είναι συγκρίσιμα. Επίσης, ο τρόπος με τον οποίο υπολογίζονται τα σημαντικά χαρακτηριστικά της ταξινόμησης μπορεί να διαφέρει από μοντέλο σε μοντέλο και να υπάρχει και μια κλίση σε κατεύθυνση συγκεκριμένων χαρακτηριστικών, η οποία υπόκειται στην αλγοριθμική υλοποίηση, όπως στην περίπτωση του κατηγοριοποιητή Random Forest (Strobl *et al.* 2007).

Η αξιολόγηση της κατηγοριοποίησης μπορεί να επιτευχθεί με ενιαίο και κατανοητό τρόπο με τη χρήση της ανάλυσης SHAP (Lundberg και Lee, 2017). Πρόκειται για μια μέθοδο η οποία λειτουργεί για κάθε μοντέλο ταξινόμησης με την οποία αποδίδονται σκορ για κάθε χαρακτηριστικό, εξηγώντας πώς αυτό συνέβαλε στην κατηγοριοποίηση. Συγκεκριμένα αποδίδουν το σκορ για το πόσο πολύ ή πόσο λίγο ένα χαρακτηριστικό συνέβαλε στο διαχωρισμό των κλάσεων. Η μέθοδος αυτή χρησιμοποιείται ευρέως σε επιστημονικές έρευνες βιοπληροφορικής όπως Meral *et al.* 2025 και Filali Razzouki *et al.* 2025 με στόχο την απάντηση ερωτημάτων σχετικά με τη λήψη αποφάσεων ενός κατηγοριοποιητή και δυνητικά στην ανεύρεση βιοδεικτών, ως πιθανή επέκταση της χρήσης και ερμηνείας αποτελεσμάτων. Στην εικόνα 3.15 παρουσιάζεται σε διάγραμμα σκέδασης, γνωστό με την ονομασία beeswarm plot (διάγραμμα σμήνους), η αξιολόγηση των χαρακτηριστικών κατά τη μέθοδο SHAP για το μοντέλο ταξινόμησης της λογιστικής παλινδρόμησης για τα δύο φύλα. Στον άξονα  $x$  αναγράφεται το αντίκτυπο του κάθε χαρακτηριστικού (άξονας  $y$ ) στην πρόβλεψη. Η τελείες εκπροσωπούν τιμές χαρακτηριστικών, δηλαδή τιμές γονιδιακής έκφρασης. Τελείες σε κοντινές αποστάσεις παραπέμπουν σε παρόμοια μοτίβα έκφρασης και αντίθετα, μεγαλύτερες αποστάσεις σε διαφοροποιημένα μοτίβα έκφρασης. Ο χρωματισμός σε μπλε παραπέμπει σε χαμηλές τιμές έκφρασης και αντίστοιχα, υψηλές τιμές έκφρασης αναπαρίστανται με κόκκινο χρώμα. Η έκταση σε αρνητική κατεύθυνση του άξονα  $x$  σημαίνει πως το αντίστοιχο χαρακτηριστικό συνέβαλε αρνητικά στην ταξινόμηση. Δηλαδή, οι τιμές του χαρακτηριστικού, ανάλογα με τα επίπεδα έκφρασης που παρουσιάζει, συνέβαλαν στην απόφαση να χαρακτηριστεί αντίστοιχο δείγμα ως αρνητικό. Οι εκτάσεις σε θετική κατεύθυνση του άξονα  $x$  εννοούν το ακριβώς αντίθετο, δηλαδή είτε με χαμηλή τιμή (μπλε) ή υψηλή τιμή έκφρασης, το αντίστοιχο χαρακτηριστικό συνέβαλε θετικά ως προς το να ληφθεί θετική απόφαση κατηγοριοποίησης. Από την εικόνα 3.15i μπορεί να ερμηνευθεί ότι χαμηλές τιμές έκφρασης του γονιδίου IFIT1 οδηγούν σε αρνητική ταξινόμηση, ενώ υψηλές σε θετική. Παρακάτω, ψηλές τιμές έκφρασης του γονιδίου LGALS2 οδηγούν σε αρνητικές προβλέψεις και αντίθετα χαμηλές σε θετικές.



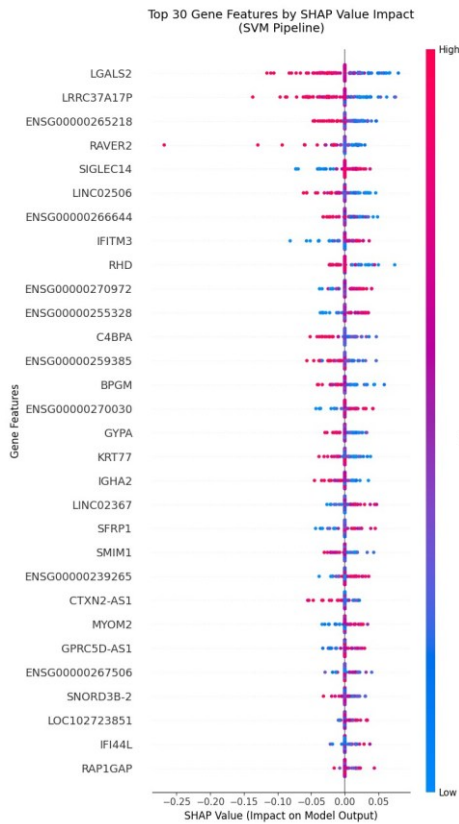
(i) Γυναίκες 50-70 ετών



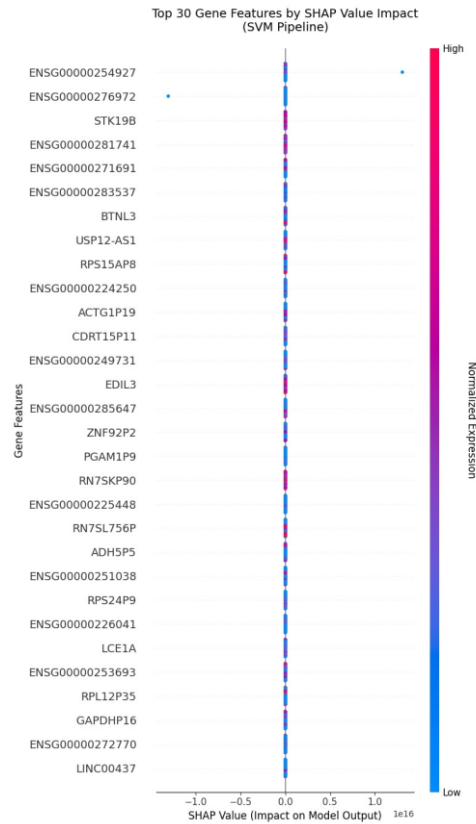
(ii) Άνδρες 50-70 ετών

**Εικόνα** Error! No text of specified style in document..15 Ανάλυση SHAP για μοντέλο Λογιστικής Παλινδρόμησης

Συγκρίνοντας τις γραφικές αναπαραστάσεις των εικόνων 3.15i και 3.15ii, ιδιαίτερα στις πρώτες εγγραφές από την κορυφή, καθώς εκεί βρίσκονται και τα χαρακτηριστικά που είχαν το μεγαλύτερο αντίκτυπο στην ταξινόμηση, διακρίνονται περισσότερα διαφορετικά γονίδια μεταξύ ανδρών και γυναικών παρά κοινά. Αυτό το φαινόμενο επιτρέπει στο συμπέρασμα, πως η νόσος του Πάρκινσον οδηγείται ενδεχομένως από εκφράσεις διαφορετικών γονιδίων μεταξύ ανδρών και γυναικών.



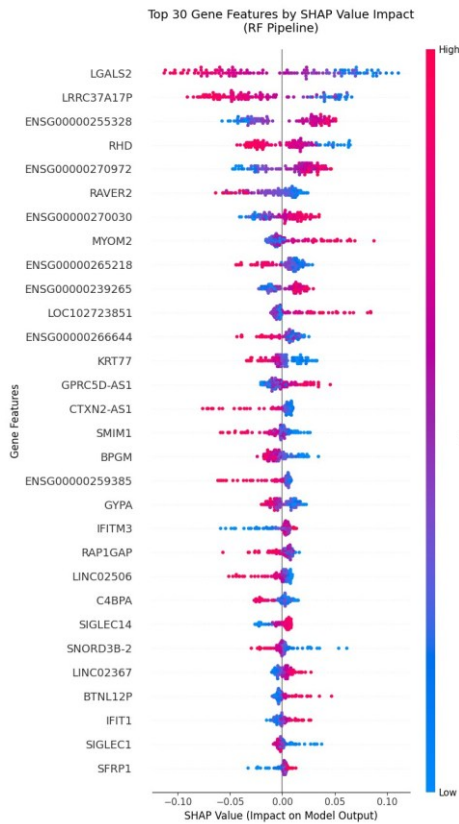
(i) Γυναίκες 50-70 ετών



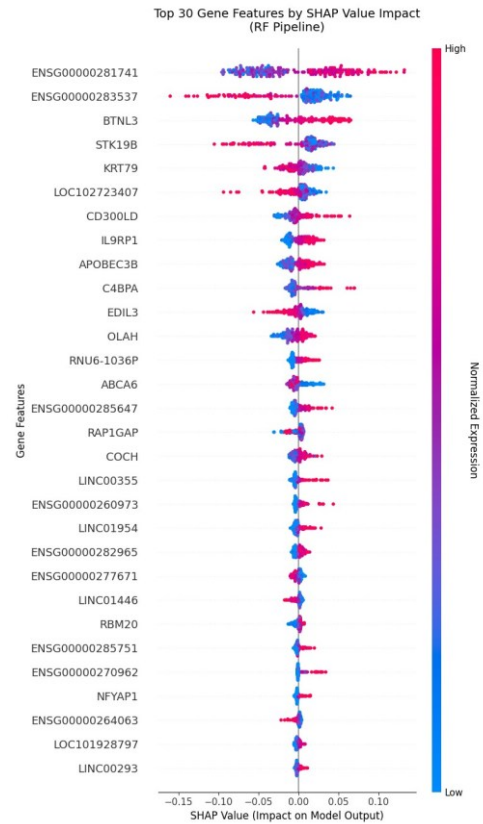
(ii) Άνδρες 50-70 ετών

## Εικόνα Error! No text of specified style in document..16 Ανάλυση SHAP για μοντέλο SVM

Το συμπέρασμα από την εφαρμογή της ανάλυσης SHAP στα αποτελέσματα της ταξινόμησης μέσω της λογιστικής παλινδρόμησης υποστηρίζεται από τα σημαντικά χαρακτηριστικά που εξάγονται από το μοντέλο SVM. Για τις γυναίκες παρουσιάζονται τα γονίδια LGALS2 και LRR37A17P ως τα σημαντικότερα. Επίσης, τα αποτελέσματα δείχνουν και συνοχή ως προς το μοτίβο έκφρασης, δηλαδή και τα δύο αυτά γονίδια βοήθησαν να διακριθούν οι νοσούντες από τους υγιείς μέσω χαμηλών επιπέδων έκφρασης και ταυτόχρονα οι υγιείς από τους νοσούντες μέσω υψηλών επιπέδων έκφρασης (3.16i). Το ίδιο μοτίβο επαναλαμβάνεται και στην περίπτωση του μοντέλου Random Forest, όπου παρουσιάζονται και πάλι τα γονίδια LGALS2 και LRR37A17P στις ανώτερες θέσεις με τα ίδια μοτίβα έκφρασης τα οποία οδήγησαν και στους προηγούμενους κατηγοριοποιητές την ταξινόμηση (3.17i). Στους άνδρες δείχνουν τα γονίδια ENSG00000283537, STK19B, KRT79 σημαντικότητα σε μοτίβο υπο-έκφρασης για την ομάδα νοσούντων. Αντίθετα, τα γονίδια IL9RP1, ENSG00000281741 και BTNL3 παρουσιάζονται ως οδηγοί διάκρισης της ομάδας νοσούντων σε μοτίβο υπέρ-έκφρασης.



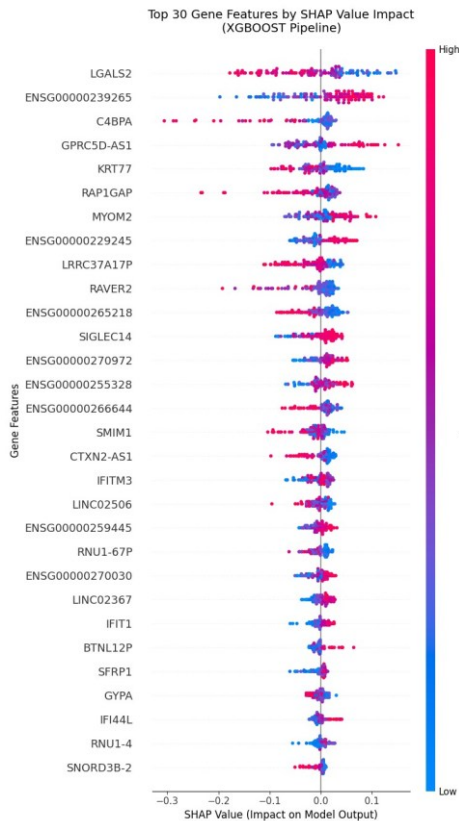
(i) Γυναίκες 50-70 ετών



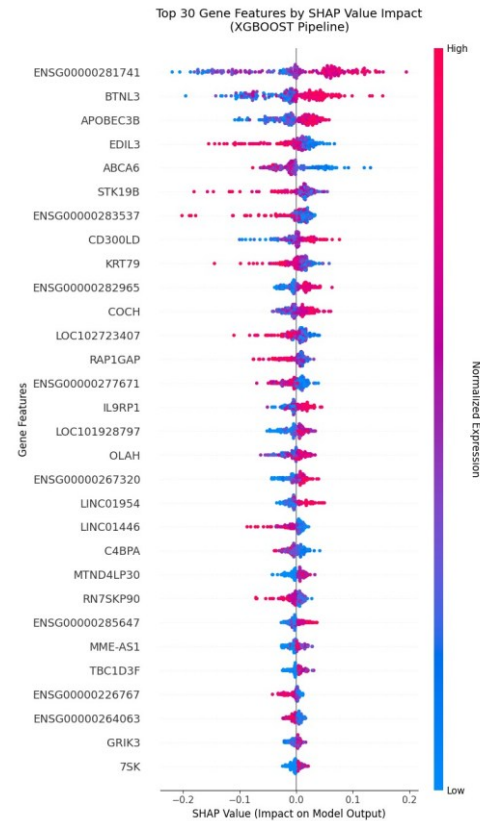
(ii) Άνδρες 50-70 ετών

**Εικόνα** Error! No text of specified style in document..17 Ανάλυση SHAP για μοντέλο Random Forest

Για τον κατηγοριοποιητή XGBoost σημαντικό ρόλο διαδραμάτισε η έκφραση του γονιδίου LGAL2 ενώ το κοινό στους κατηγοριοποιητές της λογιστικής παλινδρόμησης, SVM και Random Forest γονίδιο LRRC37A17P δεν βρίσκεται εντός των πρώτων τριών θέσεων. Αντί αυτού, παρουσιάζονται τα γονίδια ENSG00000239265, C4BPA, GPRC5D-AS1 ως σημαντικότερα, τα οποία στους άλλους κατηγοριοποιητές αποδόθηκαν χαμηλότερη σημασία. Τα σημαντικά γονίδια στην ομάδα των ανδρών παρουσιάζουν ομοιότητα στον κατηγοριοποιητή XGBoost (3.18ii) για την αντίστοιχη ομάδα του κατηγοριοποιητή Random Forest (3.17ii). Συγκεκριμένα, στις ανώτερες θέσεις βρίσκονται και στους δύο κατηγοριοποιητές τα γονίδια ENSG00000281741, BTNL3 και STK19B.



(i) Γυναίκες 50-70 ετών



(ii) Άνδρες 50-70 ετών

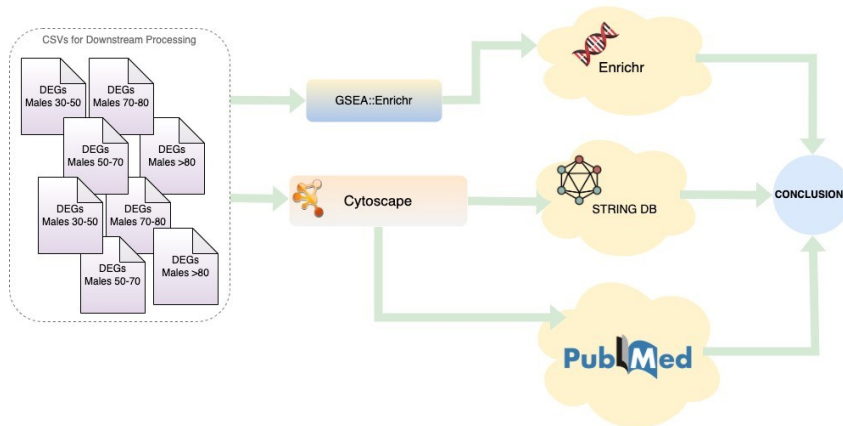
### Εικόνα Error! No text of specified style in document..18 Ανάλυση SHAP για μοντέλο XGBoost

Η εφαρμογή της μεθόδου SHAP στα μοντέλα κατηγοριοποίησης ανά φύλο της ηλικιακής ομάδας 50-70 ετών έδειξε συνοχή ως προς τα γονίδια που συντέλεσαν στην ταξινόμηση και διάκριση μεταξύ δειγμάτων των ομάδων ελέγχου και νοσούντων. Εκτός από την απόδοση παρόμοιων σκορ, διακρίθηκε και η συνοχή στα αντίστοιχα μοτίβα έκφρασης.

### 3.9 Χαρτογράφηση λειτουργικών δικτύων και ανάλυση εμπλουτισμού

Στα πλαίσια της εργασίας διενεργήθηκε η χαρτογράφηση λειτουργικών δικτύων με τη χρήση του λογισμικού Cytoscape (*Shannon κ.ά., 2003*) καθώς και ο εμπλουτισμός των αλληλεπιδράσεων και βιβλιογραφίας μέσω της βάσης STRING (*Szklarczyk et al. 2023*) και η ανάλυση εμπλουτισμού, με τη χρήση του API της βάσης Enrichr (*E. Y. Chen et al. 2013*) μέσω της βιβλιοθήκης Python GSEArq (*Fang et al. 2023*) (κώδικας 10). Επίσης, χρησιμοποιήθηκε η βάση δεδομένων Gene4PD (*Li et al. 2021*) όπου υπάρχουν εμπλουτισμένες καταχωρισμένες συσχετίσεις γονιδίων με τη νόσο του Πάρκινσον από διαδεδομένες βάσεις σχολιασμού. Ο στόχος είναι η αποτίμηση του συνόλου των αποτελεσμάτων, από τα επιμέρους αναλυτικά στάδια που πραγματοποιήθηκαν σε αυτήν

την εργασία. Η χαρτογράφηση γονιδιακών δικτύων και η ανάλυση εμπλουτισμού έγιναν με βάση τα διαφορικά εκφραζόμενα γονίδια που προέκυψαν ανά φύλο και ηλικιακή ομάδα.



**Εικόνα** Error! No text of specified style in document..19 Pipeline για την χαρτογράφηση λειτουργικών δικτύων και ανάλυση εμπλουτισμού

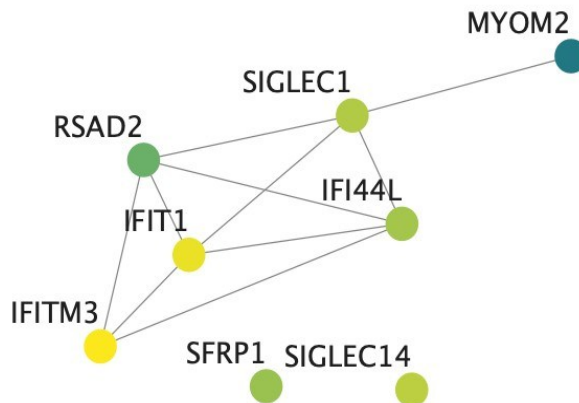
Στον πίνακα 3.4 παρατίθενται τα αποτελέσματα εμπλουτισμού μετά από κατασκευή δικτύου με βάση διαφορικά εκφραζόμενα γονίδια γυναικών, της ηλικιακής ομάδας 50-70 ετών. Έντονη είναι η παρουσία των γονιδίων IFITM3, SIGLEC1, MYOM2 τα οποία απαντώνται και στα αποτελέσματα της εφαρμογής της μεθόδου SHAP. Η εικόνα 3.20 παρουσιάζει το γονιδιακό δίκτυο, όπως αυτό προέκυψε μετά την ανάλυση μέσω Cytoscape, όπου ο σκούρος χρωματισμός των κόμβων αντιπροσωπεύει υψηλές απόλυτες τιμές έκφρασης και ο ανοιχτός αντίστοιχα χαμηλές τιμές. Χαρακτηριστική στο δίκτυο είναι η παρουσία ιντερφερόνων, η έκφραση των οποίων συμβαδίζει με μια απόκριση του ανοσοποιητικού συστήματος σε φλεγμονή ή και σε κυτταρική βλάβη (Kopitar-Jerala, 2017). Στα αποτελέσματα βιβλιογραφικού εμπλουτισμού μέσω STRING διακρίθηκαν αποτελέσματα για το υποδίκτυο των γονιδίων RSAD2, IFIT1, SIGLEC1, IFI44L, σχετικά με την κινάση Janus (Yamaoka et al. 2004), η οποία συμμετέχει στο μονοπάτι Jak/Stat, με το οποίο βρέθηκε συσχετισμός με έκφραση γονιδίου ιντερφερόνης IF27, που αναφέρθηκε στον πίνακα 3.5. Ο συσχετισμός του μονοπατιού Jak/Stat με τη νόσο του Πάρκινσον περιγράφεται από τους Lashgari et al. 2021 ως πιθανός θεραπευτικός στόχος. Ταυτόχρονα το ίδιο υποδίκτυο σχετίζεται με αναφορά σε νευροεκφυλισμό, σύμφωνα με δημοσίευση των Cooray et al. 2023.

**Πίνακας** Error! No text of specified style in document..4 Αποτελέσματα εμπλουτισμού κατά τη χαρτογράφηση γονιδιακών δικτύων γυναικών 50-70 ετών.

Source	Term Name	Intersecting Genes	P-value
WikiPathways	Network map of SARS CoV 2 signaling	IFITM3, SIGLEC1, MYOM2	
Gene Ontology (BP)	biological process involved in interaction with host	IFITM3, 9606, ENSP00000371471, SIGLEC1	1,6 · 10 <sup>-3</sup> 6,6 · 10 <sup>-3</sup> 8,2 · 10 <sup>-3</sup>
Transfac predictions	Factor: IRF-1; motif: INRAAANNGAAASN; match class: 1	SIGLEC1, MYOM2	
mirTarBase	hsa-miR-146a-5p	IFITM3, SIGLEC1, MYOM2	
Gene Ontology (BP)	biological process involved in symbiotic interaction	IFITM3, I9606.ENSP00000371471, SIGLEC1	1,06 · 10 <sup>-2</sup> 2,41 · 10 <sup>-2</sup> 2,41 · 10 <sup>-2</sup>
Gene Ontology (BP)	defense response to virus	IFITM3, SIGLEC1, MYOM2	
Reactome	Interferon alpha/beta signaling	IFITM3, SIGLEC1	
Gene Ontology (BP)	viral life cycle	IFITM3, 9606.ENSP00000371471, SIGLEC1	2,60 · 10 <sup>-2</sup> 2,61 · 10 <sup>-2</sup> 3,91 · 10 <sup>-2</sup>
Transfac predictions	Factor: IRF-4; motif: KRAAANGAAAANYN; match class: 1	SIGLEC1, MYOM2	

**Πίνακας** Error! No text of specified style in document..5 Επιλογή βιβλιογραφικών αναφορών κατά τη χαρτογράφηση γονιδιακών δικτύων γυναικών 50-70 ετών.

Description	FDR	PMID
Janus Kinase Inhibitors in the Treatment of Type I	1,75 ·	
Interferonopathies: A Case Series From a Single Center in China.	10 <sup>-5</sup>	PMID:35418997
JAK inhibitors: a potential treatment for JDM in the context of the role of interferon-driven pathology.	7,98 · 10 <sup>-5</sup>	PMID:34563217
Neuroinflammation, autoinflammation, splenomegaly and anemia caused by bi-allelic mutations in IRAK4.	1,2 · 10 <sup>-4</sup>	PMID:37744344



**Εικόνα** Error! No text of specified style in document..20 Λειτουργικό δίκτυο διαφορεικά εκφραζόμενων γονιδίων σε γυναίκες 50-70 ετών

Τα αποτελέσματα εμπλουτισμού μέσω Enrichr δεν αποκλίνουν από τα παραπάνω, καθώς και σε αυτή την περίπτωση οι σχολιασμοί με υψηλό σκορ στατιστικής σημαντικότητας συγκεντρώνονται γύρω από τη μεταγραφή ιντερφερονών και γενικότερα αφορούν σε λειτουργίες του ανοσοποιητικού. Επιπλέον των βάσεων σχολιασμού οντολογιών και μονοπατιών, συμπεριλήφθηκαν και πηγές δεδομένων από

μεταγραφικούς παράγοντες, όπου και πάλι τα αποτελέσματα σχετίζονται ή παραπέμπουν σε αποκρίσεις φλεγμονής (Πίνακας 3.6). Χαρακτηριστικό είναι το αποτέλεσμα του μεταγραφικού παράγοντα HESX1, για το οποίο βρέθηκαν αντιστοιχίσεις στην βάση Gene4PD με φαινότυπους από την οντολογία HPO (Gargano et al. 2024) που αφορούν σε κινητικά προβλήματα, τρόμους, υποσμία και ανοσμία (Mitchell et al. 2025, Tarakad και Jankonis, 2017), παθολογικά επίπεδα προλακτίνης (που σύμφωνα με Al-kuraishy et al. 2023 δεν υφίσταται ακριβή και εξακριβωμένη συσχέτιση) και κατ' επέκταση σχετικούς με τη νόσο του Πάρκινσον (Πίνακας 3.7).

**Πίνακας Error! No text of specified style in document..6** Ανάλυση εμπλουτισμού μέσω Enrichr - ενδεικτικά αποτελέσματα γυναικών 50-70 ετών

Gene set	Term	Adjusted P-value	Genes
MSigDB_Hallmark_2020	Interferon Gamma Response	$7,0 \cdot 10^{-4}$	IFITM3, RSAD2, BPGM, IFIT1, IFI44L
MSigDB_Hallmark_2020	Interferon Alpha Response	$4,3 \cdot 10^{-3}$	IFITM3, RSAD2, IFI44L
Enrichr_Submissions_TF-Gene_Cooccurrence	<b>HESX1</b>	$5,19 \cdot 10^{-2}$	IFITM3, RSAD2, SIGLEC1, IFIT1, IFI44L

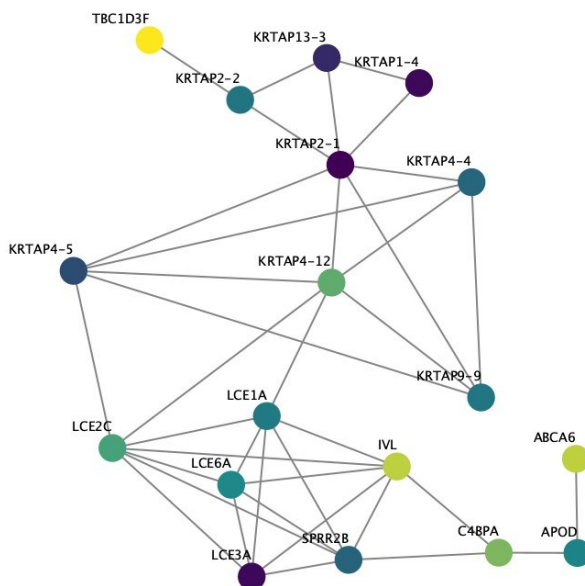
**Πίνακας Error! No text of specified style in document..7** Σχετικοί με τη νόσο του Πάρκινσον Φαινότυποι σε έδαφος του μεταγραφικού παράγοντα HSX1

HP ID	Phenotype
HP:0001337	Tremor
HP:0001288	Gait disturbance

HP:0004409	Hyposmia
HP:0000458	Anosmia
HP:0040086	Abnormal prolactin level
HP:0008202	Reduced circulating prolactin concentration

### 3.10 Αποτελέσματα ανδρών - Ηλικίες 50-70 έτη

Τα αποτελέσματα της ανάλυσης λειτουργικών δικτύων μέσω Cytoscape για τα διαφορικά εκφραζόμενα γονίδια ανδρών της ηλικιακής ομάδας 50-70 ετών χαρακτηρίζονται από την έντονη παρουσία πρωτεϊνών κερατίνης (Εικόνα 3.21). Σύμφωνα με τα ευρήματα της διαφορικής ανάλυσης, το γονίδιο KRT77 βρέθηκε να υπό-εκφράζεται και επίσης συνέβαλε στη διαφοροποίηση μεταξύ παθολογικών και υγιών δειγμάτων στην εφαρμογή ταξινόμησης μηχανικής μάθησης. Παραδόξως, τα παρακάτω απεικονιζόμενα γονίδια παρουσιάζουν υπέρ-έκφραση (*σκούρος χρωματισμός των κόμβων*). Έρευνες που έχουν αποτυπωθεί σε δημοσιεύσεις όπως των Liu *et al.* 2025 και X. Wang *et al.* 2022 αναφέρουν μεταξύ άλλων και το αντίκυτο σε επίπεδα κερατίνης στη νόσο του Πάρκινσον. Χαρακτηριστική είναι η απουσία οποιουδήποτε στοιχείου απόκρισης ανοσοποιητικού συστήματος ή κάποιου σχολιασμού σχετικά με φλεγμονή, σε αντίθεση με τους όρους εμπλουτισμού που σχημάτιζαν την πλειοψηφία του συνόλου του εμπλουτισμού στις γυναίκες. Από τον εμπλουτισμό βιβλιογραφίας μέσω της βάσης STRING δεν προέκυψαν αποτελέσματα, παρά μόνο για τη λειτουργική ανάλυση εμπλουτισμού. Στον πίνακα 3.8 παρατίθενται τα αποτελέσματα της λειτουργικής ανάλυσης εμπλουτισμού μέσω STRING. Όλα τα αποτελέσματα αφορούν βιολογικές διεργασίες σχετικές με την κερατίνη. Στις σχετικές διεργασίες συμβάλλουν κατ'επανάληψη γονίδια όπως LCE1A, LC5AA (*οικογένεια πρωτεϊνών Late Cornified Envelope*) και οι σχετικές με την πρωτεΐνη της κερατίνης KRTAP45, KRTAP21, κοκ.



**Εικόνα** Error! No text of specified style in document..21 Λειτουργικό δίκτυο διαφορικά εκφραζόμενων γονιδίων σε άνδρες 50-70 ετών

**Πίνακας** Error! No text of specified style in document..8 Αποτελέσματα λειτουργικού εμπλουτισμού - Άνδρες 50-70 ετών

Category	Description	FDR value	Genes
Reactome Pathways	Keratinization	$1,36 \cdot 10^{-14}$	LCE1A, LC5AA, KRTAP45, SPRR2B, IVL, LCECC, KRTAP14, KRTAP133, KRTAP21, KRTAP22, LC5BA
Pfam	Keratin, high sulfur B2 protein	$8,27 \cdot 10^{-11}$	KRTAP45, KRTAP14, KRTAP44, KRTAP21, KRTAP99, KRTAP412, KRTAP22
UniProt Keywords	Keratinization	$1,06 \cdot 10^{-9}$	LCE1A, LC5AA, SPRR2B, IVL, LCECC, LC5BA
STRING Clusters	Keratin	$2,32 \cdot 10^{-9}$	KRTAP45, KRTAP14, KRTAP44, KRTAP133, KRTAP21, KRTAP99, KRTAP412

GO Cellular Component	Keratin filament	$7,04 \cdot 10^{-9}$	KRTAP45, KRTAP14, KRTAP44, KRTAP21, KRTAP99, KRTAP412, KRTAP22
GO Cellular Component	Intermediate filament	$7,15 \cdot 10^{-9}$	KRTAP45, KRTAP14, KRTAP44, KRTAP133, KRTAP21, KRTAP99, KRTAP412, KRTAP22
STRING Clusters	Keratinization and Cornified envelope	$1,74 \cdot 10^{-8}$	LCE1A, LC5AA, SPRR2B, IVL, LCECC, LC5BA
STRING Clusters	Keratin, high sulfur B2 protein	$1,95 \cdot 10^{-9}$	KRTAP45, KRTAP44, KRTAP21, KRTAP99, KRTAP412
STRING Clusters	Keratinization	$3,48 \cdot 10^{-8}$	LCE1A, LC5AA, SPRR2B, LCECC, LC5BA
TISSUES	Scalp	$4,42 \cdot 10^{-8}$	KRTAP45, KRTAP44, KRTAP21, KRTAP99, KRTAP412, KRTAP22
STRING Clusters	Late cornified envelope	$7,12 \cdot 10^{-7}$	LCE1A, LC5AA, LCECC, LC5BA
STRING Clusters	Keratin, high sulfur B2 protein	$1,01 \cdot 10^{-8}$	KRTAP45, KRTAP44, KRTAP21, KRTAP99
Reactome Pathways	Formation of the cornified envelope	$1,16 \cdot 10^{-6}$	LCE1A, LC5AA, SPRR2B, IVL, LCECC, LC5BA
GO Biological Process	Keratinization	$1,99 \cdot 10^{-6}$	LCE1A, LC5AA, SPRR2B, IVL, LCECC, LC5BA
STRING Clusters	Late cornified envelope	$9,07 \cdot 10^{-6}$	LCE1A, LCECC, LC5BA
TISSUES	Skin	$1,66 \cdot 10^{-5}$	KRTAP45, APOD, SPRR2B, IVL, KRTAP44, KRTAP21, KRTAP99, KRTAP412, KRTAP22, LC5BA

UniProt Keywords	Keratin	6,13 · 10–5	KRTAP45, KRTAP14, KRTAP133, KRTAP21, KRTAP22
InterPro Domains	Keratin-associated protein	3,80 · 10–4	KRTAP45, KRTAP14, KRTAP21, KRTAP22
Pfam	Late cornified envelope	9,70 · 10–4	LCE1A, LC5AA, LCECC
InterPro Domains	Late cornified envelope protein	3,00 · 10–3	LCE1A, LC5AA, LCECC
STRING Clusters	Keratin, high sulfur B2 protein	5,00 · 10–3	KRTAP45, KRTAP44
STRING Clusters	Mixed, incl. Keratin, high sulfur B2 protein, and Keratin, high-sulphur matrix protein	5,00 · 10–3	KRTAP21, KRTAP99
GO Cellular Component	Cytoskeleton	2,38 · 10–2	KRTAP45, IVL, KRTAP14, KRTAP44, KRTAP133, KRTAP21, KRTAP99, KRTAP412, KRTAP22
STRING Clusters	Keratin-associated protein, PMG type, and Keratin- associated protein, type6/6/16/19/20/21	3,92 · 10–2	KRTAP14, KRTAP133
TISSUES	Scalp	4,42 · 10–8	KRTAP45, KRTAP44, KRTAP21, KRTAP99, KRTAP412, KRTAP22

Τα αποτελέσματα των αναλυτικών σταδίων, όπως αυτά παρουσιάστηκαν στις προηγούμενες ενότητες αυτού του κεφαλαίου, ανέδειξαν σημαντικές διαφορές μεταξύ ανδρών και γυναικών της ίδιας ηλικιακής ομάδας, διαφορεικά εκφραζόμενων γονιδίων

και, κατ' επέκταση, διαφορετικών σημαντικών χαρακτηριστικών στο κομμάτι της μηχανικής μάθησης. Αυτό το εύρημα καλύπτεται και από δημοσίευση των Simunovic κ.ά., 2010, όπου περιγράφονται αποτελέσματα μετά από ανάλυση παθολογικών δειγμάτων νευρώνων της μέλαινας ουσίας. Από τη χαρτογράφηση λειτουργικών δικτύων προέκυψε για τις γυναίκες δίκτυο σχετικό με απόκριση σε φλεγμονή, ενώ στους άνδρες σχετικό με την κερατίνη. Πρόκειται και στις δύο περιπτώσεις για το αντίστοιχα μεγαλύτερο δίκτυο το οποίο κατασκευάστηκε με τη χρήση του συνόλου στατιστικά σημαντικών διαφορικά εκφραζόμενων γονιδίων της εκάστοτε ομάδας. Σύμφωνα με τη διασταύρωση των ευρημάτων με τη βιβλιογραφία, όπως παρατέθηκε ειδικά στην ανάλυση εμπλουτισμού, θα μπορούσε να θεωρηθεί πως το ολικό αίμα μπορεί να περιέχει μάρτυρες μεταγραφωμικής φύσεως οι οποίοι παραπέμπουν στη νόσο του Πάρκινσον, αλλά όχι άμεσα, μιας και δεν υπάρχει ένδειξη η οποία μπορεί να συσχετιστεί αδιάσειστα με τα μέχρι στιγμής γνωστά βιολογικά μονοπάτια, τα οποία διαδραματίζουν κάποιο ρόλο στην αποδόμηση της συνουκλεΐνης-α.

Η έντονη σχέση της κερατίνης στην περίπτωση των ανδρών που παρουσιάζεται από τη χαρτογράφηση λειτουργικών δικτύων, ίσως θα μπορούσε να συσχετιστεί με το γεγονός ότι η αποδόμηση της κερατίνης ρυθμίζεται από το μονοπάτι ουβικουΐτινης-πρωτεασώματος (*Rogel et al. 2010*) το οποίο σχετίζεται και με την παθογένεση της νόσου του Πάρκινσον (*Lim και Tan, 2007*). Από την ανάλυση διαφορικής έκφρασης για την ηλικιακή ομάδα 50-70 ετών των ανδρών προέκυψε πολύ μεγαλύτερος αριθμός διαφορικά υπέρ-εκφραζόμενων γονιδίων σε σύγκριση με τα υπό-εκφραζόμενα της ίδιας ομάδας. Επίσης, η συγκεκριμένη ηλικιακή ομάδα δεν ακολουθεί το μοτίβο έκφρασης των άλλων ομάδων, όπου ο αριθμός υπό-εκφραζόμενων γονιδίων είναι μεγαλύτερος των υπέρ-εκφραζόμενων (εικόνα 3.8). Ο λόγος για τον οποίο προέκυψε αυτή η εικόνα, δηλαδή εάν όντως πρόκειται για λόγο βιολογικής σημασίας ή τεχνητό θόρυβο, αποτελεί ερωτηματικό. Το ίδιο ισχύει και για την ηλικιακή ομάδα γυναικών, άνω των 80 ετών (3.6). Η στρατηγική στρωματοποίησης έθεσε στο επίκεντρο την ηλικιακή ομάδα των 50-70, καθώς σύμφωνα με διάφορες έρευνες καλύπτει το εύρος των ηλικιών, κατά τις οποίες εκδηλώνεται η νόσος (*Bloem et al. 2021, Wirdefeldt et al. 2011*). Η απόφαση ελήφθη μη λαμβάνοντας υπόψη τεχνικούς παράγοντες, όπως αριθμό δειγμάτων ανά ηλικιακή ομάδα, κάτι το οποίο μπορεί και να παραποιεί μερικώς τα αποτελέσματα των αναλύσεων. Μια πιο "δίκαιη" κατανομή των ηλικιακών ομάδων, με επιπλέον γνώμονα τους αριθμούς των δειγμάτων που προκύπτουν, θα μπορούσε ίσως να οδηγήσει σε καλύτερα αποτελέσματα.

Η ανάλυση SHAP για την ομάδα ανδρών, ηλικιών 50-70 ετών για τον κατηγοριοποιητή SVM απεικονίζει για δύο χαρακτηριστικά πολύ υψηλές τιμές σε

σύγκριση με αυτές των υπόλοιπων χαρακτηριστικών, οι οποίες είναι σαφώς πολύ μικρότερες. Η εικόνα παραμένει η ίδια ακόμη και μετά από επανεκτέλεση της κατηγοριοποίησης και της ανάλυσης SHAP. Δεν έχει εξεταστεί εάν η κατηγοριοποίηση λειτουργεί εξίσου ικανοποιητικά και σε διαφορετικά σύνολα δεδομένων αλληλούχησης και σε περίπτωση που οι επιδόσεις αποκλίνουν σημαντικά, να διερευνηθούν πιθανά αίτια, όπως η υπερπροσαρμογή στο σύνολο δεδομένων που χρησιμοποιήθηκε.

### 3.11 Συμπεράσματα

Διενεργήθηκε εκτεταμένη ανάλυση σε δείγματα αλληλούχησης μεταγραφώματος ολικού αίματος ατόμων με τη νόσο του Πάρκινσον, έναντι υγιών. Διενεργήθηκε λεπτομερής στρωματοποίηση ανά φύλο και ηλικιακή ομάδα και η ανάλυση έλαβε χώρα σύμφωνα με τα επί μέρους στρώματα. Η μέθοδος της ανάλυσης διαφορικά εκφραζόμενων γονιδίων χρησιμοποιήθηκε για την επιλογή χαρακτηριστικών της κατηγοριοποίησης μέσω αλγορίθμων μηχανικής μάθησης, αλλά και για τη χαρτογράφηση λειτουργικών δικτύων και την ανάλυση εμπλουτισμού. Επιλέχθηκε η ηλικιακή ομάδα 50-70 ετών και για τα δύο φύλα, καθώς πρόκειται για το πιο κρίσιμο ηλικιακό εύρος για την εκδήλωση της νόσου.

Ένα βασικό συμπέρασμα είναι πως η διαφοροποίηση ασθενών της νόσου Πάρκινσον και υγιών μπορεί να διενεργηθεί με τη χρήση δειγμάτων ολικού αίματος, καθώς κάποια από τα ευρήματα διασταυρώθηκαν επιτυχώς με εγγραφές στη βάση Gene4PD και επίσης έδειξαν να βρίσκονται και σε παραπλήσια βιολογικά μονοπάτια με τα γνωστά της νόσου του Πάρκινσον. Η έντονη διαφορά στα πρότυπα έκφρασης ανά φύλο καθώς και στη διαφορική έκφραση και σε συνέχεια σημαντικών χαρακτηριστικών, τα οποία προέκυψαν από την ανάλυση SHAP, αφήνει να κατανοηθεί πως η νόσος εκφράζεται βιολογικά με διαφορετικούς τρόπους στα δύο φύλα. Επίσης, μπορεί να θεωρηθεί πως υπάρχουν ψήγματα πληροφορίας, τα οποία παραπέμπουν σε κοινά βιολογικά μονοπάτια με αυτά που οδηγούν στην παθογένεση της νόσου. Μια ασφαλής εξαγωγή συμπερασμάτων ως προς την παρουσία βιοδεικτών και ρυθμιστικών σε ολικό αίμα ασθενών της νόσου του Πάρκινσον δεν είναι δυνατή, απλά και μόνο με τα αποτελέσματα που εντοπίστηκαν στα παραπάνω κεφάλαια, παρόλο που υπάρχουν ενδείξεις, σύμφωνα με τις πηγές πληροφοριών με τις οποίες διασταυρώθηκε μέρος των αποτελεσμάτων, και παρατέθηκαν αντίστοιχα εντός των σχετικών με την ανάλυση κεφαλαίων. Η αδυναμία ταυτοποίησης βιοδεικτών οφείλεται κυρίως στο γεγονός ότι πολλά αποτελέσματα, όπως τα σχετικά με τις αποκρίσεις του ανοσοποιητικού σε φλεγμονή, καλύπτουν ένα ευρύ φάσμα πιθανών παθήσεων, καθιστώντας τα σχετικά

γονίδια έτσι ακατάλληλα ως διαγνωστικούς βιοδείκτες. Στον αντίποδα, το εύρημα της κερατίνης ως ρυθμιστικού στόχου του μονοπατιού της ουβικουΐνης είναι ίσως πιο συγκεκριμένο και θα μπορούσε να διερευνηθεί περαιτέρω.

#### 4. Τεχνολογίες βιοπληροφορικής και αναλυτικής δεδομένων στη νόσο Αλτσχάιμερ

##### 4.1 Εισαγωγή

Η νόσος Αλτσχάιμερ είναι μια προοδευτική νευροεκφυλιστική διαταραχή και η πιο συχνή αιτία άνοιας, η οποία χαρακτηρίζεται από απώλεια μνήμης, γνωστική έκπτωση και αλλαγές στη συμπεριφορά. Η πολύπλοκη παθολογία της περιλαμβάνει γενετικούς, μοριακούς και περιβαλλοντικούς παράγοντες, καθιστώντας τις παραδοσιακές πειραματικές προσεγγίσεις ανεπαρκείς από μόνες τους για την πλήρη κατανόηση της νόσου. Η βιοπληροφορική παρέχει ισχυρά υπολογιστικά εργαλεία για την ανάλυση βιολογικών δεδομένων μεγάλης κλίμακας, όπως γονιδιωματικά, μεταγραφωμικά και πρωτεωμικά σύνολα δεδομένων, με στόχο την αποκάλυψη των μοριακών μηχανισμών που διέπουν τη νόσο Αλτσχάιμερ. Μέσω της βιοπληροφορικής ανάλυσης, μπορούν να εντοπιστούν γονίδια που σχετίζονται με τη νόσο, διαταραγμένα βιολογικά μονοπάτια, πιθανοί βιοδείκτες και θεραπευτικοί στόχοι, συμβάλλοντας στην έγκαιρη διάγνωση και στη βελτίωση των θεραπευτικών στρατηγικών.

#### 5. Βιοπληροφορική μελέτη και ανάλυση δεδομένων σε ιστικά δείγματα με νόσο Αλτσχάιμερ

Τα σύνολα δεδομένων που χρησιμοποιήθηκαν ελήφθησαν από τη βάση δεδομένων scREAD, η οποία περιλαμβάνει σύνολα δεδομένων scRNA-seq και snRNA-seq που προέρχονται από μεταθανάτιο ανθρώπινο εγκεφαλικό ιστό που εμφανίζει AD και ζωικά μοντέλα με παθολογία AD (*Jiang et al 2020*). Συμπεριλήφθηκαν επίσης σύνολα δεδομένων ελέγχου που προέρχονται από υγιή δείγματα. Χρησιμοποιώντας προηγμένες υπολογιστικές μεθόδους, οι ερευνητές μπορούν να ενσωματώσουν σύνολα δεδομένων scRNA-seq από διάφορες περιοχές του εγκεφάλου για να σχηματίσουν μια ολιστική άποψη της νόσου. Σε μια πρόσφατη εργασία της ομάδας μας, συγκρίθηκαν δεδομένα scRNA-seq μεταξύ του φλοιού των ποντικών και του υπόκαμπου από υγιή δείγματα και δείγματα AD και παρατηρήθηκαν εκφραζόμενα γονίδια, κυρίως εμπλουτισμένα σε μουσκαρινικούς υποδοχείς ακετυλοχολίνης, υποδοχείς ντοπαμίνης και περιυναπτική εξωκυτταρική μήτρα (*Krokidis et al 2023*). Η παρούσα μελέτη αξιοποιεί υπολογιστικές τεχνικές για την ανάλυση δεδομένων scRNA-seq από πολλαπλές περιοχές του εγκεφάλου

που επηρεάζονται από την AD και ενισχύει τις υπάρχουσες μελέτες ότι η AD εκδηλώνεται διαφορετικά σε διαφορετικές περιοχές του εγκεφάλου και τύπους κυττάρων. Το RNA-seq μπορεί να εξερευνήσει τη γονιδιακή έκφραση σε πολλαπλές περιοχές του εγκεφάλου, παρέχοντας νέες προκλήσεις για τον εντοπισμό βασικών βιολογικών διεργασιών από μοριακή άποψη. Συνθέτοντας δεδομένα στον ενδορινικό φλοιό, τον προμετωπιαίο φλοιό, την άνω μετωπιαία έλικα και τον άνω βρεγματικό λοβό, στοχεύσαμε να δημιουργήσουμε ένα ολοκληρωμένο μοντέλο της επίδρασης της AD στο κυτταρικό και μοριακό τοπίο του εγκεφάλου. Μέσω αυτής της προσέγγισης, τονίζουμε τη δυνατότητα των υπολογιστικών αναλύσεων να εμβαθύνουν την κατανόησή μας για νευροεκφυλιστικές ασθένειες όπως η AD. Αυτό επιτρέπει τη σύγκριση κυτταρικών και μοριακών προφίλ σε διαφορετικές περιοχές, εντοπίζοντας κοινά και διακριτά στοιχεία της παθολογίας.

### 5.1. Συλλογή δεδομένων scRNA-seq

Η παρούσα μελέτη χρησιμοποίησε σύνολα δεδομένων scRNA-seq από τη βάση δεδομένων scREAD, που περιλαμβάνουν κύτταρα από τέσσερις διαφορετικές περιοχές του ανθρώπινου εγκεφάλου για να εξερευνήσει το κυτταρικό και μοριακό τοπίο της AD (Jiang et al 2020). Πιο συγκεκριμένα, η ανάλυση περιελάμβανε επιλεγμένα σύνολα δεδομένων από τον ενδορινικό φλοιό, τον προμετωπιαίο φλοιό, την άνω μετωπιαία έλικα και τον άνω βρεγματικό λοβό, με μεμονωμένα σύνολα δεδομένων που αντιπροσωπεύουν τόσο υγιή όσο και άτομα που έχουν προσβληθεί από AD, όπως περιγράφηκε προηγουμένως. Για τον ενδορινικό φλοιό, συμπεριλάβαμε δεδομένα από υγιή άτομα ελέγχου (σύνολα δεδομένων AD00201 και AD00202) και τα συγκρίναμε με δεδομένα από άτομα με AD (σύνολα δεδομένων AD00203, AD00204, AD00205 και AD00206). Στην περίπτωση του προμετωπιαίου φλοιού, η ομάδα ελέγχου περιελάμβανε σύνολα δεδομένων AD01101 και AD01102, ενώ η ομάδα της νόσου περιελάμβανε AD01103 και AD01104. Τα δεδομένα της ομάδας ελέγχου της ανώτερης μετωπιαίας έλικας προέρχονταν από AD00801 του συνόλου δεδομένων και τα δεδομένα της ομάδας της νόσου αντιπροσωπεύονταν από AD00802 και AD00803. Τέλος, χρησιμοποιήθηκαν έξι σύνολα δεδομένων που περιλαμβάνουν κύτταρα από τον άνω βρεγματικό λοβό. Μεταξύ αυτών, δύο σύνολα δεδομένων αντιστοιχούσαν στον έλεγχο (AD01201 και AD01202) ενώ τα υπόλοιπα τέσσερα αντιπροσώπευαν κύτταρα με παθολογία AD (AD01203, AD01204, AD01205 και AD01206). Όσον αφορά τον αριθμό των δειγμάτων για κάθε περιοχή του εγκεφάλου: Δείγματα ελέγχου ενδορινικού φλοιού: 27.892. Δείγματα νόσου του ενδορινικού φλοιού: 42.635. Δείγματα ελέγχου προμετωπιαίου φλοιού: 29.752. Δείγματα νόσου του προμετωπιαίου φλοιού: 18.326. Δείγματα ελέγχου ανώ μετωπιαίας έλικας:

30.608. Δείγματα νόσου της άνω μετωπιαίας έλικας: 50.256. Δείγματα ελέγχου άνω βρεγματικού λοβού: 411. Δείγματα νόσου του άνω βρεγματικού λοβού: 2486.

### 5.3. Προσδιορισμός του βέλτιστου αριθμού κύριων συνιστωσών για την ανάλυση δεδομένων

Η συνάρτηση `determine_optimal_pcs` βοηθά στην εύρεση του καλύτερου αριθμού κύριων συνιστωσών (PC) για δεδομένα αλληλουχίας RNA ενός κυττάρου. Αυτή η συνάρτηση χρησιμοποιεί την διακύμανση από την Ανάλυση Κύριων Συνιστωσών (PCA) για να βρει τον ελάχιστο αριθμό στοιχείων που απαιτούνται για την εξήγηση ενός συγκεκριμένου ποσού συνολικής διακύμανσης. Αυτή η μέθοδος αποφεύγει τη χειροκίνητη και υποκειμενική μέθοδο του αγκώνα, καθιστώντας τη διαδικασία πιο αντικειμενική και αναπαραγωγίμη. Η συνάρτηση εκτελεί PCA στο σύνολο δεδομένων χρησιμοποιώντας τη συνάρτηση `sc.tl.pca` της βιβλιοθήκης `scanpy`. Ο αριθμός των στοιχείων ορίζεται από `n_pcs` και η διακύμανση που εξηγείται από κάθε στοιχείο αποθηκεύεται στο `adata.uns['pca']['variance_ratio']`. Στη συνέχεια, αθροίζει αυτές τις διακυμάνσεις συσσωρευτικά. Χρησιμοποιώντας το `np.searchsorted`, η συνάρτηση βρίσκει τον μικρότερο αριθμό κυρίων συνιστωσών που απαιτούνται για να υπερβεί το `variance_threshold`. Εάν το όριο δεν επιτευχθεί εντός των υπολογισμένων στοιχείων, χρησιμοποιεί όλα τα στοιχεία. Διαφορετικά, επιστρέφει τον βέλτιστο αριθμό κυρίων συνιστωσών που απαιτούνται για να ξεπεραστεί το όριο.

Με την αυτοματοποίηση της επιλογής του αριθμού των κυρίων συνιστωσών, η λειτουργία μειώνει την ανάγκη για χειροκίνητη παρέμβαση και υποκειμενική κρίση. Η χρήση μιας συνεπούς και αντικειμενικής μεθόδου για τον προσδιορισμό του βέλτιστου αριθμού κυρίων συνιστωσών διασφαλίζει ότι τα αποτελέσματα είναι αναπαραγωγίμα σε διαφορετικά σύνολα δεδομένων και αναλύσεις. Η αυτοματοποίηση της διαδικασίας επιτρέπει ταχύτερη ανάλυση, η οποία είναι ιδιαίτερα ωφέλιμη όταν πρόκειται για μεγάλα σύνολα δεδομένων τυπικά στον προσδιορισμό αλληλουχίας RNA ενός κυττάρου. Επιπλέον, οι χρήστες μπορούν εύκολα να προσαρμόσουν τις παραμέτρους `n_pcs` και `variance_threshold` για να ταιριάζουν στις συγκεκριμένες ανάγκες και προτιμήσεις τους

### 5.4. Προσδιορισμός και διόρθωση batch effect σε σύνολα δεδομένων scRNA-seq

Η διαχείριση των επιδράσεων του batch effect έγινε χρησιμοποιώντας αντικείμενα `AnnData` ώστε να ενσωματωθούν ομαλά στην ανάλυση scRNA-seq. Τα βασικά βήματα περιελάμβαναν τον εντοπισμό, την αξιολόγηση και τη διόρθωση του batch effect για τη διασφάλιση αξιόπιστων αναλύσεων. Αρχικά, διενεργήθηκαν έλεγχοι για τον

εντοπισμό πολλαπλών batches, καθορίζοντας εάν χρειαζόταν διόρθωση. Εάν υπήρχε μόνο ένα batch, η διαδικασία σταματούσε για εξοικονόμηση πόρων. Το UMAP χρησιμοποιήθηκε για την αρχική οπτικοποίηση, ακολουθούμενο από το ANOVA για την ανίχνευση του batch effect, με τιμές p προσαρμοσμένες μέσω της μεθόδου Benjamini-Hochberg. Μέθοδοι διόρθωσης, όπως «combat» και «harmony», χρησιμοποιήθηκαν για την διόρθωση. Η οπτικοποίηση UMAP μετά τη διόρθωση επιβεβαίωσε την αποτελεσματικότητα εξαλείφοντας διακριτές συστάδες ομάδων, διασφαλίζοντας την ομογενοποίηση δεδομένων για επόμενες αναλύσεις.

#### 5.5. Επισκόπηση ενοποίησης και προεπεξεργασίας δεδομένων scRNA-seq

Για να διασφαλιστεί η κατάλληλη ενοποίηση και ιχνηλασιμότητα, κάθε σύνολο δεδομένων αντιγράφηκε και επισημάνθηκε με την περιοχή προέλευσής του στον εγκέφαλο. Αυτά τα σύνολα δεδομένων στη συνέχεια συγχωνεύτηκαν σε ένα ενοποιημένο αντικείμενο AnnData χρησιμοποιώντας τη συνάρτηση `sc.concat`, διατηρώντας τη ετερογένεια και την προέλευση κάθε σημείου δεδομένων. Πραγματοποιήθηκε φιλτράρισμα των γονιδίων για την εναρμόνιση του συνόλου γονιδίων σε όλα τα σύνολα δεδομένων, επιτρέποντας αξιόπιστες συγκριτικές αναλύσεις. Εφαρμόστηκαν μετρήσεις αυστηρού ποιοτικού ελέγχου (QC) για την αφαίρεση ακραίων τιμών, στοχεύοντας κύτταρα με μη φυσιολογικό αριθμό γονιδίων ή υψηλή έκφραση μιτοχονδριακών γονιδίων. Συγκεκριμένα, τα κύτταρα φιλτραρίστηκαν με βάση τον αριθμό γονιδίων (300 έως 2500) και την έκφραση μιτοχονδριακών γονιδίων (<10%) για να αποκλειστούν μη βιώσιμα κύτταρα και πιθανά διπλά, διασφαλίζοντας την ποιότητα του συνόλου δεδομένων για ακριβή ανάλυση.

#### 5.6. Ενσωμάτωση συνόλων δεδομένων από διάφορες περιοχές του εγκεφάλου

Η διαδικασία ξεκίνησε με κανονικοποίηση και λογαριθμικό μετασχηματισμό για την τυποποίηση των επιπέδων γονιδιακής έκφρασης στα κύτταρα, διασφαλίζοντας τη συμβατότητα των δεδομένων. Η κανονικοποίηση (με χρήση `sc.pp.normalize_total`) με άθροισμα στόχο  $1 \times 10^4$  και ο λογαριθμικός μετασχηματισμός (μέσω `sc.pp.log1p`) σταθεροποίησαν τη διακύμανση, καθιστώντας τα δεδομένα πιο ομοιογενή και έτοιμα για ανάλυση μετριάζοντας τις τεχνικές ανισότητες. Στη συνέχεια, εντοπίστηκαν και επιλέχθηκαν γονίδια υψηλής διακύμανσης για να επισημάνουν βιολογικά σήματα εν μέσω τεχνικού θορύβου, εστιάζοντας σε γονίδια με σημαντική διακύμανση. Αυτό το βήμα είναι μεγάλης σημασίας για την κατανόηση της υποκείμενης βιολογίας του δείγματος. Πραγματοποιήθηκε ανίχνευση και απομάκρυνση διπλών κυττάρων (doublets), δηλαδή

περιπτώσεων όπου περισσότερα από ένα κύτταρα αλληλουχούνται ως ένα. Η συνάρτηση *remove\_doublets*, αξιοποιώντας το εργαλείο *Scrublet*, προέβλεψε και απέκλεισε τα doublets από το σύνολο δεδομένων. Το *Scrublet* προσομοιώνει τον σχηματισμό doublets και αποδίδει σε κάθε κύτταρο μια αντίστοιχη βαθμολογία (doublet score). Παράμετροι όπως *min\_counts*, *min\_cells*, *min\_gene\_variability\_pctl* και *n\_prin\_comps* προσαρμόστηκαν για να βελτιστοποιήσουν την ακρίβεια της ανίχνευσης. Οι βαθμολογίες doublet και οι σχετικές προβλέψεις καταχωρήθηκαν στο αντικείμενο *AnnData* για περαιτέρω έλεγχο. Στη συνέχεια, τα κύτταρα που χαρακτηρίστηκαν ως doublets απομακρύνθηκαν, αποδίδοντας ένα σύνολο δεδομένων που περιλαμβάνει μόνο τα πραγματικά μεμονωμένα κύτταρα. Το καθαρισμένο αυτό dataset, απαλλαγμένο από διπλές εγγραφές, ήταν πλέον κατάλληλο για ακριβείς και αξιόπιστες αναλύσεις.

### 5.7. Ομαδοποίηση και οπτικοποίηση

Ο αλγόριθμος ομαδοποίησης Leiden (*sc.tl.leiden*) εφαρμόστηκε στο σύνολο δεδομένων ελέγχου scRNA-seq, χρησιμοποιώντας ανάλυση 0,5 και τυχαία κατάσταση 5 για συνεπή ομαδοποίηση. Αυτό οργάνωσε τα κύτταρα σε διακριτές ομάδες με βάση τα προφίλ γονιδιακής έκφρασης, εντοπίζοντας μοναδικούς κυτταρικούς πληθυσμούς. Μετά την ομαδοποίηση, χρησιμοποιήθηκε UMAP για τη μείωση των διαστάσεων, προβάλλοντας δεδομένα γονιδιακής έκφρασης σε δύο διαστάσεις. Τα συμπλέγματα κυττάρων εμφανίστηκαν με διαφορετικά χρώματα, και η οπτικοποίηση βελτιώθηκε με την επισήμανση των σημείων, υπομνήματος και προσαρμοσμένης χρωματικής παλέτας, καθιστώντας τα αποτελέσματα πιο ευανάγνωστα. Για να μελετηθεί η χωρική κατανομή των συστάδων, χρησιμοποιήθηκε ένας θερμοχάρτης για την απεικόνιση της κατανομής τους στις διαφορετικές εγκεφαλικές περιοχές. Ομάδες με κύτταρα από τρεις ή περισσότερες περιοχές του εγκεφάλου έδειξαν ευρύτερη αναπαράσταση και πιθανή βιολογική σημασία. Τα κύτταρα από αυτές τις αντιπροσωπευτικές συστάδες διατηρήθηκαν και ο αλγόριθμος Leiden εφαρμόστηκε ξανά με τις ίδιες παραμέτρους. Η ομαδοποίηση αυτή αντιπροσώπευε με ακρίβεια την υποκείμενη βιολογική ποικιλότητα. Οι συστάδες προβάλλονται εκ νέου στον χώρο UMAP για συγκριτικούς σκοπούς, επιβεβαιώνοντας την εγκυρότητα της ομαδοποίησης και αποσαφηνίζοντας τη χωρική κατανομή των κυτταρικών πληθυσμών.

Ο αλγόριθμος Leiden επιλέχθηκε για τον αποτελεσματικό χειρισμό των δεδομένων scRNA-seq. Έχει τη δυνατότητα να εντοπίζει ξεκάθαρες και βιολογικά ουσιαστικές συστάδες, ακόμη και σε μεγάλα και πολύπλοκα σύνολα δεδομένων. Σε σχέση με άλλες μεθόδους, επιτυγχάνει μεγαλύτερη συνοχή στις ομάδες που σχηματίζει,

διασφαλίζοντας ότι τα προκύπτοντα σύνολα κυττάρων αποτυπώνουν με ακρίβεια την υποκείμενη δομή των δεδομένων. Ο σχεδιασμός του αλγορίθμου Leiden τον βοηθά να βρει μικρότερες ομάδες μέσα στα δεδομένα, δίνοντας μια λεπτομερή εικόνα των κυτταρικών διαφορών. Επιπλέον, παράγει σταθερά και επαναλαμβανόμενα αποτελέσματα, τα οποία είναι σημαντικά για συνεπή αποτελέσματα σε διαφορετικές αναλύσεις, καθιστώντας τα ευρήματά μας πιο αξιόπιστα.

Μια ανάλυση διαφορικής γονιδιακής έκφρασης με τη δοκιμή Wilcoxon (sc.tl.rank\_genes\_groups) εντόπισε βασικά γονίδια-δείκτες για κάθε συστάδα. Η απεικόνιση των δεικτών έγινε με γραφική παράσταση τύπου dot plot, αναδεικνύοντας τα χαρακτηριστικά μοτίβα γονιδιακής έκφρασης και διευκολύνοντας την αναγνώριση του κάθε κυτταρικού τύπου. Οι ομάδες συσχετίστηκαν με συγκεκριμένους κυτταρικούς τύπους, στηριζόμενες στα προφίλ γονιδίων-δεικτών, μεταξύ των οποίων ανασταλτικοί νευρώνες, μικρογλοία, αστροκύτταρα και διεγερτικοί νευρώνες. Οι χαρακτηρισμοί αυτοί οπτικοποιήθηκαν σε ένα διάγραμμα UMAP, χρωματικά κωδικοποιημένο ανά τύπο κυττάρου, απεικονίζοντας την κατανομή τους σε όλο τον χώρο UMAP. Αυτές οι περιεκτικές αναλύσεις και απεικονίσεις παρείχαν μια βαθιά κατανόηση της κυτταρικής ετερογένειας και των μοριακών χαρακτηριστικών στο σύνολο δεδομένων ελέγχου scRNA-seq, προσφέροντας πολύτιμες πληροφορίες για τις κυτταρικές συνθέσεις και τις λειτουργικές οργανώσεις των περιοχών του εγκεφάλου, ενισχύοντας την κατανόησή μας για την υποκείμενη βιολογία. Η μεθοδολογία που ακολουθήθηκε απεικονίζεται στην Εικόνα 5.1.

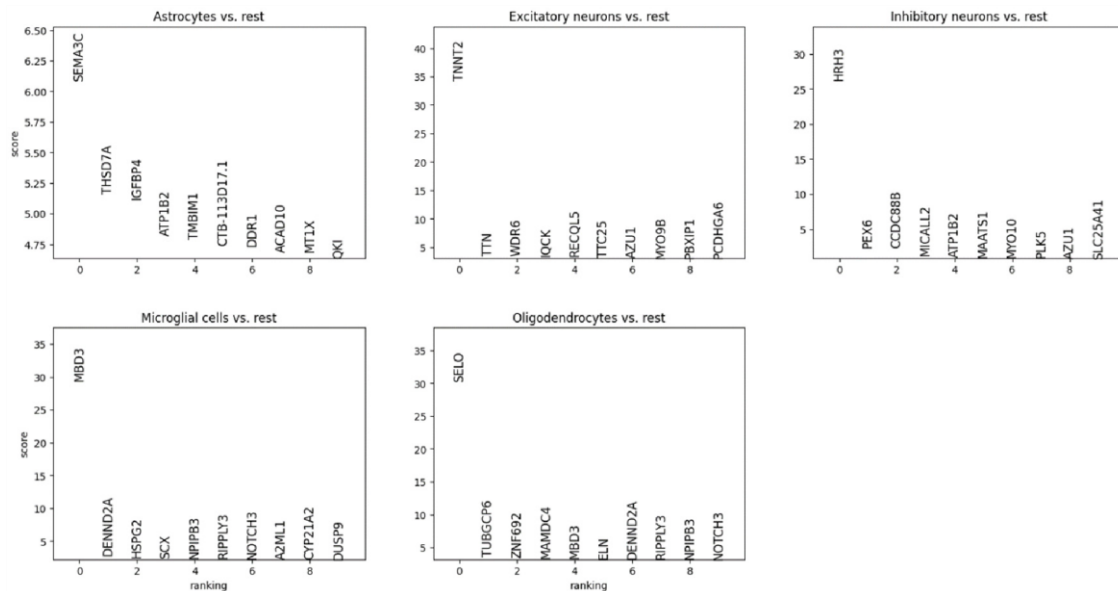


**Εικόνα 5.1** Διάγραμμα ροής ανάλυσης δεδομένων scRNA-Seq που χρησιμοποιήθηκε στην παρούσα ανάλυση.

## 6. Αποτελέσματα

Σύμφωνα με την ανάλυσή, τα ευρήματα οπτικοποιούνται σχολαστικά μέσω μιας σειράς εργαλείων που έχουν σχεδιαστεί για να ενισχύσουν την ερμηνευτική μας ικανότητα. Αυτό περιλαμβάνει την κατάταξη των γονιδίων που σχετίζονται με κάθε συστάδα και τη χρήση ενός θερμικού χάρτη, ο οποίος απεικονίζει τα γονίδια δείκτες και τα μοτίβα έκφρασής τους στις συστάδες. Μέσω αυτών των τεχνικών οπτικοποίησης, Όχι μόνο επισημαίνουμε τις βιολογικές διαφορές μεταξύ διαφορετικών συστάδων, αλλά

υπογραμμίζουμε επίσης την πιθανή ανακάλυψη μοναδικών κυτταρικών ταυτοτήτων ή καταστάσεων. Αυτή η ολοκληρωμένη προσέγγιση εμπλουτίζει σημαντικά την κατανόσή μας για την υποκείμενη βιολογία του συνόλου δεδομένων, ανοίγοντας το δρόμο για περαιτέρω εξερευνησεις στις κυτταρικές περιπλοκές του εγκεφάλου. Η σειρά γραφημάτων αντιπροσωπεύει μια συγκριτική ανάλυση των κατανομών γονιδιακής έκφρασης σε μεμονωμένες ομάδες με φόντο όλα τα άλλα κύτταρα που δεν περιλαμβάνονται σε αυτές τις ομάδες (Εικόνα 6.1). Κάθε διάγραμμα συνοδεύεται από ένα αναγνωριστικό κύτταρο που δείχνει τη συστάδα στην οποία ανήκει, επιτρέποντας τη σύγκριση με τους υπόλοιπους κυτταρικούς πληθυσμούς του συνόλου δεδομένων. Σε κάθε διάγραμμα, ο οριζόντιος άξονας παραθέτει τα γονίδια από αριστερά προς τα δεξιά ανάλογα με τη σχετική τους σημασία για το προφίλ γονιδιακής έκφρασης της συστάδας. Ο κατακόρυφος άξονας αποτυπώνει τον βαθμό διαφορικής έκφρασης, όπου υψηλότερες τιμές υποδηλώνουν ισχυρότερη τάση αυξημένης έκφρασης στη συγκεκριμένη συστάδα σε σχέση με τα υπόλοιπα κύτταρα.



**Εικόνα 6.1** Διαφορική έκφραση των κορυφαίων 10 ταξινομημένων γονιδίων σε κάθε κυτταρική ομάδα σε σύγκριση με τις υπόλοιπες, όπως αναλύθηκε στο σύνολο δεδομένων.

Τα σημεία που εμφανίζονται σε κάθε γράφημα, όπως φαίνεται και στην Εικόνα 6.1 αντιστοιχούν σε μεμονωμένα γονίδια. Η θέση τους αποτυπώνει τόσο τη σχετική συνάφειά τους με τη συγκεκριμένη συστάδα όσο και το επίπεδο έκφρασης που παρουσιάζουν σε αυτή. Για παράδειγμα, ένα γονίδιο που εμφανίζεται προς τα αριστερά με υψηλή βαθμολογία έχει ουσιαστική σημασία εντός αυτού του συμπλέγματος και

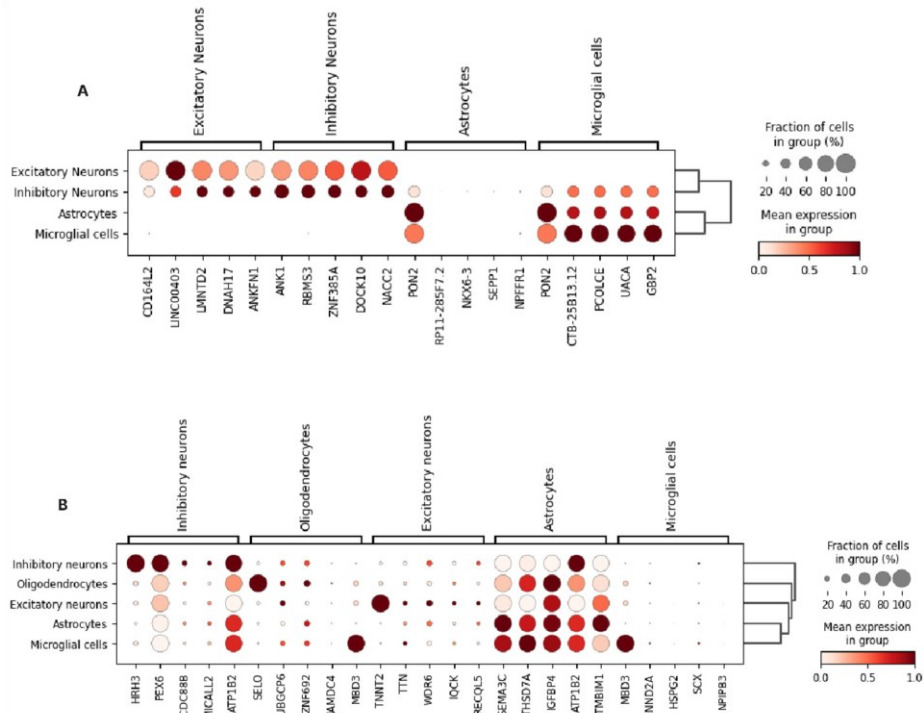
παρουσιάζει ένα σημαντικά διαφορετικό επίπεδο έκφρασης σε σύγκριση με κύτταρα εκτός του συμπλέγματος. Η διακριτή κατανομή των σημείων σε όλο το φάσμα κατάταξης μας επιτρέπει να διακρίνουμε ποια γονίδια είναι πιο χαρακτηριστικά κάθε συστάδας. Αυτή η συγκριτική γραφική προσέγγιση είναι ανεκτίμητη για την ανάδειξη των γονιδίων που διακρίνουν κάθε σύμπλεγμα από τα υπόλοιπα, παρέχοντας έτσι μια λεπτομερή εικόνα της μοριακής ταυτότητας κάθε κυτταρικού πληθυσμού. Παρατηρώντας τα μοτίβα και τις θέσεις αυτών των γονιδίων σε όλη τη σειρά των γραφημάτων, οι ερευνητές μπορούν να εξάγουν συμπεράσματα σχετικά με τις βιολογικές διεργασίες που μπορεί να κυριαρχούν σε κάθε σύμπλεγμα και να εντοπίσουν πιθανούς στόχους για περαιτέρω πειραματική έρευνα.

Επιπλέον, η γονιδιακή έκφραση απεικονίστηκε με χάρτες θερμότητας, προσφέροντας μια εύληπτη γραφική αναπαράσταση των διαφορών μεταξύ των κατηγοριών, όπως φαίνεται στο Σχήμα 3. Αυτές οι τιμές ταξινομούνται συστηματικά και ομαδοποιούνται ανά συγκεκριμένες κατηγορίες, παρέχοντας σαφή διαφοροποίηση μεταξύ διαφορετικών επιπέδων γονιδιακής έκφρασης. Μέσα σε αυτό το διάγραμμα, κάθε στήλη αντιστοιχεί σε μια κατηγορία ή συστάδα, όπως διαφορετικούς τύπους κυττάρων, ιστούς ή πειραματικές συνθήκες. Για κάθε συστάδα στο διάγραμμα, η έκφραση των γονιδίων ποσοτικοποιήθηκε ως λογαριθμική μεταβολή, δείχνοντας πόσο διαφέρει η έκφραση κάθε γονιδίου σε σχέση με την αναφορά. Μια θετική τιμή λογαριθμικής μεταβολής υποδεικνύει ότι το γονίδιο εκφράζεται περισσότερο μέσα σε αυτό το σύμπλεγμα σε σχέση με την αναφορά, δηλαδή είναι πιο ενεργό σε αυτήν την κατάσταση. Αντίθετα, μια αρνητική αλλαγή πτυχής υποδηλώνει μείωση ή μείωση της γονιδιακής έκφρασης, υπονοώντας ότι το γονίδιο είναι λιγότερο ενεργό ή δυνητικά κατασταλμένο στη συγκριτική κατάσταση. Αυτή η ανάλυση διαφορικής έκφρασης, που επισημαίνεται μέσω της Εικόνας 6.2, μας δίνει τη δυνατότητα να εντοπίσουμε γονίδια που παρουσιάζουν σημαντικές αλλαγές στην έκφραση σε διαφορετικές κατηγορίες ή καταστάσεις, διευκολύνοντας τις γνώσεις σχετικά με τις βιολογικές διεργασίες και τους μηχανισμούς ασθενειών.



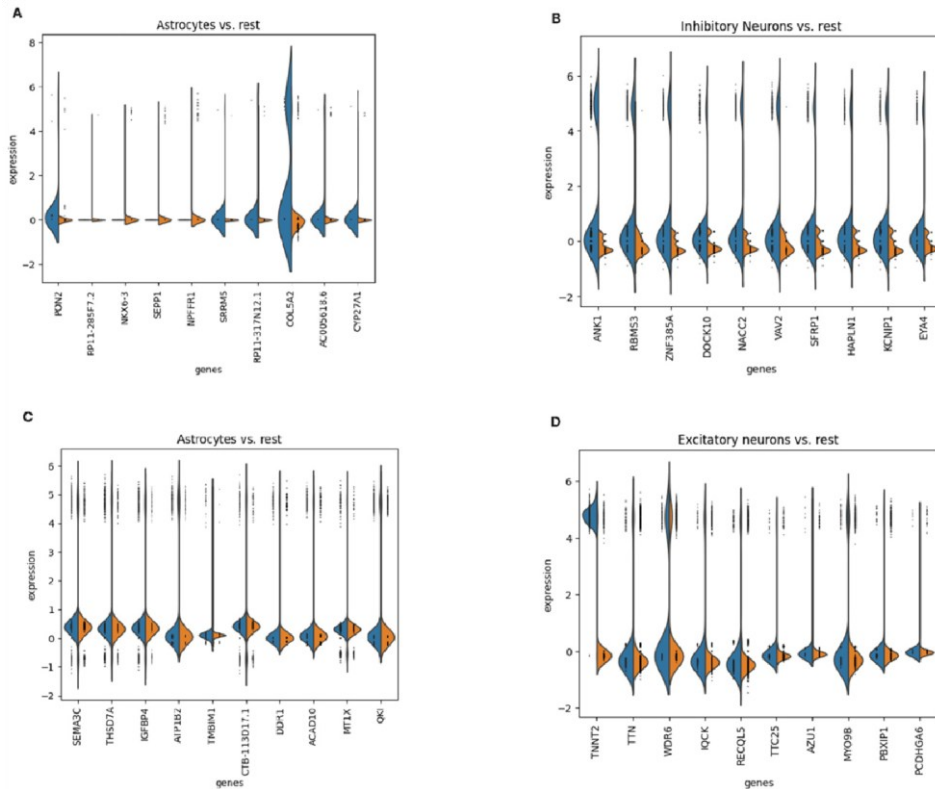
έκφρασης βασικών γονιδίων σε διαφορετικούς τύπους κυττάρων εντός της ομάδας ελέγχου, συμπεριλαμβανομένων των ανασταλτικών νευρώνων, των διεγερτικών νευρώνων, των αστροκυττάρων, των πρόδρομων κυττάρων ολιγοδενδροκυττάρων, των ολιγοδενδροκυττάρων και των μικρογλοιακών κυττάρων. Το Σχήμα 4 παρουσιάζει το dot plot για την ομάδα ασθενειών, επισημαίνοντας αλλαγές στα πρότυπα γονιδιακής έκφρασης λόγω της παθολογικής κατάστασης. Σημαντικές αλλαγές στα επίπεδα έκφρασης και στο ποσοστό κυττάρων που εκφράζουν συγκεκριμένα γονίδια είναι εμφανείς σε όλους τους τύπους κυττάρων σε σύγκριση με την ομάδα ελέγχου. Αυτά τα dot plots είναι ένα ανεκτίμητο εργαλείο για την οπτικοποίηση της πολυπλοκότητας της γονιδιακής έκφρασης σε διαφορετικούς κυτταρικούς πληθυσμούς, προσφέροντας μια λεπτομερή επισκόπηση της κυτταρικής ετερογένειας και των επιπτώσεων της νόσου στη δυναμική της γονιδιακής έκφρασης.

Επιπλέον, τα διαγράμματα βιολιού, όπως φαίνεται στην Εικόνα 6.4, παρουσιάζουν την κατανομή των επιπέδων έκφρασης για τα πέντε πιο εκφραζόμενα γονίδια σε κάθε ομάδα, σε σύγκριση με τις υπόλοιπες ομάδες του συνόλου δεδομένων ελέγχου. Το πλάτος κάθε διαγράμματος απεικονίζει την πυκνότητα των κυττάρων σε διαφορετικά επίπεδα έκφρασης, ενώ η ξεχωριστή προβολή διευκολύνει την οπτική σύγκριση των δύο συνθηκών.



**Εικόνα 6.3** Dot plot scRNA-seq. (A) Το dot plot παρουσιάζει τη διαφορική έκφραση των πέντε κορυφαίων γονιδίων σε καθεμία από τις τέσσερις κυτταρικές ομάδες στο σύνολο δεδομένων ελέγχου. (B) Το dot plot απεικονίζει τη διαφορική έκφραση των πέντε

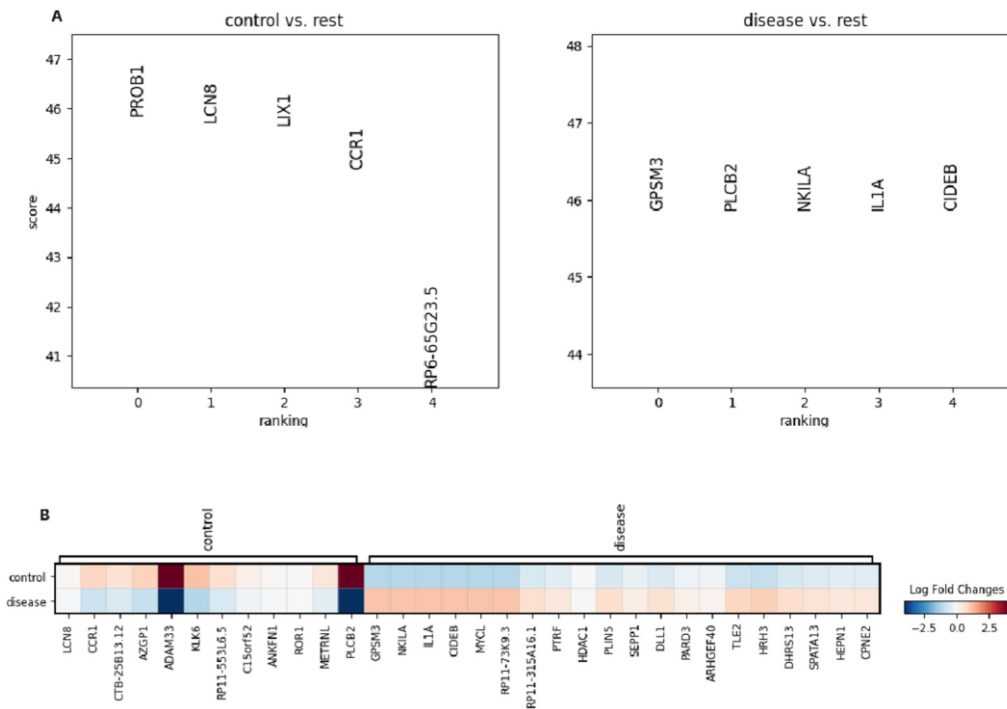
κορυφαίων γονιδίων σε καθεμία από τις κυτταρικές ομάδες στο σύνολο δεδομένων με νόσο Αλτσχάιμερ.



**Εικόνα 6.4** Τα διαγράμματα βιολιού εμφανίζουν την κατανομή των επιπέδων έκφρασης για τα 10 κορυφαία διαφορεικά εκφραζόμενα γονίδια (Α) σε αστροκύτταρα κυτταρικού τύπου σε σύγκριση με τις υπόλοιπες ομάδες στο σύνολο δεδομένων ελέγχου, (Β) σε νευρώνες αναστολέα κυτταρικού τύπου σε σύγκριση με τις υπόλοιπες ομάδες στο σύνολο δεδομένων ελέγχου, (Γ) σε αστροκύτταρα κυτταρικού τύπου σε σύγκριση με τις υπόλοιπες ομάδες στο σύνολο δεδομένων της νόσου και (Δ) σε διεγερτικούς νευρώνες σε σύγκριση με τις υπόλοιπες ομάδες στο σύνολο δεδομένων της νόσου.

Τέλος, συγχωνεύσαμε τα νεοδημιουργηθέντα αντικείμενα AnnData (adata) τόσο για την ομάδα ελέγχου όσο και για την ομάδα ασθένειας σε ένα ενιαίο αντικείμενο adata, ακολουθώντας τις ίδιες αναλυτικές διαδικασίες που είχαν καθοριστεί προηγουμένως (Εικόνα 6.5). Η ενσωμάτωση διευκολύνει μια εστιασμένη σύγκριση των κυττάρων που υπάρχουν τόσο στις συνθήκες ελέγχου όσο και στις συνθήκες Αλτσχάιμερ, επιτρέποντάς μας να εντοπίσουμε γονίδια που παρουσιάζουν σημαντικές διαφορές μεταξύ αυτών των ομάδων. Εστιάζοντας σε αυτά τα συγκεκριμένα κύτταρα, προσπαθήσαμε να αποκαλύψουμε κρίσιμους γενετικούς δείκτες που μπορεί να αποσαφηνίσουν τους υποκείμενους μηχανισμούς εξέλιξης της νόσου ή ανθεκτικότητας. Αυτή η συγκριτική

προσέγγιση είναι καθοριστική για τη διάκριση των γενετικών εκφράσεων που είναι καθοριστικές για την εκδήλωση της νόσου σε σύγκριση με τις φυσιολογικές φυσιολογικές καταστάσεις.



**Εικόνα 6.5** (Α) Διαφορική έκφραση των πέντε κορυφαίων γονιδίων σε κάθε κατάσταση σε σύγκριση με την άλλη. (Β) Θερμικός χάρτης των αλλαγών της λογαριθμικής πτυχής στη γονιδιακή έκφραση μεταξύ ελέγχου και κατάστασης ασθένειας.

Η ανάλυση με διάφορους αλγορίθμους ταξινόμησης, όπως K Neighbors, Extreme Gradient Boosting, Decision Tree και Gradient Boosting, έδειξε σταθερά υψηλή ακρίβεια και υψηλές τιμές AUC. Τα αποτελέσματα υποδεικνύουν ότι τα μοντέλα μπορούν να διακρίνουν αποτελεσματικά τους υγιείς από τα ασθενή δείγματα βάσει των δεδομένων γονιδιακής έκφρασης (Εικόνα 6.6). Επιπλέον, οι αλγόριθμοι παρουσίασαν ισορροπημένη απόδοση ως προς την ευαισθησία, την ακρίβεια και τη βαθμολογία F1, αναδεικνύοντας την ικανότητά τους να εντοπίζουν τα θετικά κρούσματα με περιορισμένα ψευδώς θετικά. Οι δείκτες συμφωνίας Kappa και MCC επιβεβαιώνουν περαιτέρω την αξιοπιστία των μοντέλων, δείχνοντας ότι οι προβλέψεις τους είναι συνεπείς και σταθερές σε διαφορετικά αλγοριθμικά πλαίσια. Αν και παρατηρούνται διακυμάνσεις στους χρόνους εκπαίδευσης, είναι αξιοσημείωτο ότι μοντέλα όπως το Extreme Gradient Boosting συνδυάζουν υψηλή ακρίβεια με αποτελεσματική εκτέλεση, επιτυγχάνοντας εντυπωσιακά αποτελέσματα σε σύντομο χρόνο.

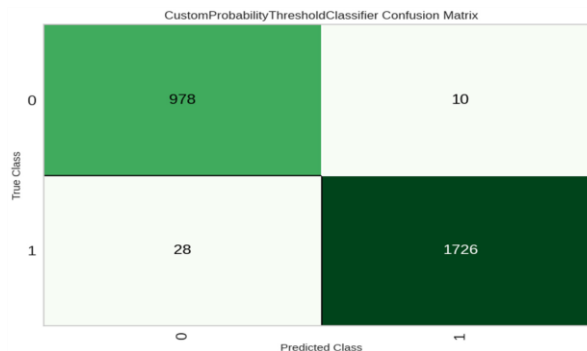
Η ανάλυσή μας υποστηρίχθηκε από τη βιβλιοθήκη PyCaret (έκδοση 3.3.2), ένα ευέλικτο και φιλικό προς τον χρήστη εργαλείο για πειραματισμούς μηχανικής μάθησης. Το PyCaret αυτοματοποιεί πολλές φάσεις της διαδικασίας μηχανικής μάθησης, όπως την προεπεξεργασία των δεδομένων, την επιλογή χαρακτηριστικών, την εκπαίδευση μοντέλων, το συντονισμό υπερπαραμέτρων και την αξιολόγηση των μοντέλων. Η εύχρηστη διεπαφή και οι πολλαπλές δυνατότητες του εργαλείου επιτρέπουν στους ερευνητές να εξετάζουν διαφορετικά μοντέλα, να συγκρίνουν την απόδοσή τους και να αντλούν ουσιαστικά συμπεράσματα από τα δεδομένα. Με τη βοήθεια του PyCaret, πραγματοποιήσαμε μια λεπτομερή αξιολόγηση διαφόρων αλγορίθμων ταξινόμησης στο σύνολο δεδομένων γονιδιακής έκφρασης, διευκολύνοντας τη λήψη τεκμηριωμένων αποφάσεων και επιταχύνοντας την ερευνητική διαδικασία. Η αναλυτική τεκμηρίωση και τα παραδείγματα χρήσης της βιβλιοθήκης την καθιστούν πολύτιμο εργαλείο τόσο για αρχάριους όσο και για έμπειρους ερευνητές. Πέρα από τις ποσοτικές μετρήσεις, η ανάλυσή μας παρέχει βαθύτερη κατανόηση των υποκείμενων βιολογικών μηχανισμών.

Model		Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
<b>lightgbm</b>	Light Gradient Boosting Machine	0.9854	0.9988	0.9854	0.9857	0.9854	0.9683	0.9686
<b>knn</b>	K Neighbors Classifier	0.9825	0.9958	0.9825	0.9832	0.9826	0.9623	0.9629
<b>xgboost</b>	Extreme Gradient Boosting	0.9810	0.9983	0.9810	0.9818	0.9811	0.9592	0.9598
<b>dt</b>	Decision Tree Classifier	0.9796	0.9779	0.9796	0.9796	0.9796	0.9557	0.9557
<b>gbc</b>	Gradient Boosting Classifier	0.9796	0.9981	0.9796	0.9804	0.9797	0.9561	0.9567
<b>nb</b>	Naive Bayes	0.9664	0.9959	0.9664	0.9698	0.9668	0.9290	0.9318
<b>ridge</b>	Ridge Classifier	0.9650	0.9732	0.9650	0.9657	0.9647	0.9230	0.9242
<b>lda</b>	Linear Discriminant Analysis	0.9650	0.9731	0.9650	0.9656	0.9648	0.9234	0.9243
<b>qda</b>	Quadratic Discriminant Analysis	0.9620	0.9935	0.9620	0.9650	0.9624	0.9194	0.9217
<b>svm</b>	SVM - Linear Kernel	0.9562	0.9746	0.9562	0.9574	0.9564	0.9059	0.9068
<b>lr</b>	Logistic Regression	0.9182	0.9783	0.9182	0.9259	0.9193	0.8285	0.8339
<b>rf</b>	Random Forest Classifier	0.9095	0.9997	0.9095	0.9280	0.9111	0.8141	0.8288
<b>et</b>	Extra Trees Classifier	0.8891	0.9999	0.8891	0.9158	0.8912	0.7750	0.7958
<b>ada</b>	Ada Boost Classifier	0.3606	0.9983	0.3606	0.1300	0.1911	0.0000	0.0000
<b>dummy</b>	Dummy Classifier	0.3606	0.5000	0.3606	0.1300	0.1911	0.0000	0.0000

**Εικόνα 6.6.** Πραγματοποιήθηκε μια ολοκληρωμένη ανάλυση της απόδοσης που προέρχονται από διάφορους αλγορίθμους ταξινόμησης, χρησιμοποιώντας δεδομένα γονιδιακής έκφρασης από το σύνολο δεδομένων ασθενειών, με στόχο την κατηγοριοποίηση των ατόμων σε υγιείς ή ασθενείς. Τα διαφορετικά εκπαιδευμένα μοντέλα αξιολογήθηκαν, και οι επισημασμένες μετρήσεις υποδεικνύουν τις υψηλότερες βαθμολογίες. Η ανάλυση πραγματοποιήθηκε με τη βοήθεια της βιβλιοθήκης PyCaret.

Ο πίνακας σύγχυσης (confusion matrix) παρέχει λεπτομερή εικόνα της απόδοσης του αλγορίθμου, εμφανίζοντας τα αληθινά θετικά, τα αληθινά αρνητικά, καθώς και τα ψευδώς θετικά και ψευδώς αρνητικά αποτελέσματα. Στην Εικόνα 6.7 παρουσιάζεται ο πίνακας σύγχυσης για τον Προσαρμοσμένο Αλγόριθμο Κατωφλίου Πιθανοτήτων, ο οποίος αναδεικνύει την ικανότητά του να ταξινομεί σωστά τις περιπτώσεις. Ο αλγόριθμος αυτός βασίζεται στη μετατροπή των προβλεπόμενων πιθανοτήτων σε ετικέτες κλάσης με βάση ένα καθορισμένο όριο. Από προεπιλογή, το όριο ορίζεται στο 0,5, πράγμα που σημαίνει ότι οποιαδήποτε προβλεπόμενη πιθανότητα άνω του 0,5 θεωρείται θετική, ενώ τιμές κάτω του 0,5 θεωρούνται αρνητικές.

Στην ανάλυσή μας, προσαρμόσαμε αυτό το όριο στο 0,7, υποδεικνύοντας ότι μόνο οι προβλέψεις με πιθανότητα πάνω από 0,7 ταξινομούνται ως θετικές, καθιστώντας τα κριτήρια ταξινόμησης πιο αυστηρά. Το μοντέλο εντόπισε 978 αληθινά αρνητικά, υποδεικνύοντας υψηλή ακρίβεια στον εντοπισμό υγιών δειγμάτων. Επιπλέον, τα 1.726 αληθινά θετικά δείχνουν την αξιοπιστία του στην ακριβή ανίχνευση περιπτώσεων ασθένειας. Ο χαμηλός αριθμός ψευδώς αρνητικών (28) και ψευδώς θετικών (10) αναδεικνύει την ακρίβεια του αλγορίθμου, διασφαλίζοντας ότι οι περισσότερες θετικές ταυτοποιήσεις είναι σωστές, ενώ ελάχιστα πραγματικά θετικά κρούσματα παραμένουν αδιάγνωστα. Ο συνδυασμός υψηλής ακρίβειας και ευαισθησίας επιβεβαιώνει επίσης την αυξημένη βαθμολογία F1 που παρατηρείται στις αξιολογήσεις. Ο πίνακας σύγχυσης προσφέρει μια λεπτομερή εικόνα της απόδοσης του αλγορίθμου, αναδεικνύοντας την ικανότητά του να διατηρεί χαμηλό ποσοστό ψευδώς θετικών, κάτι που είναι κρίσιμο για τη μείωση των περιττών εξετάσεων και θεραπειών. Παράλληλα, το χαμηλό ποσοστό ψευδώς αρνητικών διασφαλίζει ότι οι περισσότερες περιπτώσεις ασθένειας εντοπίζονται σωστά, υπογραμμίζοντας την αξιοπιστία και τη σταθερότητα του μοντέλου σε πρακτικές εφαρμογές. Αυτές οι αναλύσεις επιβεβαιώνουν τη χρησιμότητα του αλγορίθμου σε συνθήκες πραγματικού κόσμου, όπου οι ακριβείς και αξιόπιστες προβλέψεις είναι ζωτικής σημασίας.



**Εικόνα 6.7** Πίνακας σύγχυσης για το μοντέλο προσαρμοσμένου ορίου πιθανότητας.

Χρησιμοποιήσαμε επιπλέον διάφορες βιβλιοθήκες για τον εντοπισμό μονοπατιών που σχετίζονται με γονίδια που ρυθμίζονται ανοδικά ή καθοδικά. Η ανάλυση των μονοπατιών αποκάλυψε ότι τα γονίδια με ανοδική ρύθμιση εμπλουτίστηκαν κυρίως σε μονοπάτια όπως η σηματοδότηση Galpha (I), η μεταγραφή νευρωνικών προσδεμάτων, η επεξεργασία ιντερλευκίνης-1, η διάσπαση κυτταροσκελετικών πρωτεϊνών με μεσολάβηση κασπάσης και η σηματοδότηση GPCR. Επιπλέον, εμπλουτισμός σε μονοπάτια νευροφλεγμονής, γλουταμινεργικής σηματοδότησης, ενεργοποίησης NF-κΒ από ιντερλευκίνη-1 και αντιφλεγμονώδους σηματοδότησης IL-10. Η ανάλυση γονιδιακής οντολογίας (GO) διαχώρισε τους όρους σε βιολογικές διεργασίες, αποκαλύπτοντας ότι τα ρυθμισμένα γονίδια εμπλουτίστηκαν κυρίως σε διαδικασίες όπως η απόκριση σε μεταλλικά ιόντα και η θετική ρύθμιση της μεταφοράς πρωτεϊνών. Αντίθετα, τα γονίδια με καθοδική ρύθμιση εμπλουτίστηκαν κυρίως σε μονοπάτια που περιλαμβάνουν τη σηματοδότηση νευροτροφίνης μέσω του υποδοχέα P75 NTR, την παραγωγή IL-4 με μεσολάβηση LTC4-CYSLTR, τη μεταγραφή νευρωνικών προσδεμάτων (Reactome) καθώς και τη σηματοδότηση Wnt, τις οδούς σηματοδότησης χημειοκινών και τον μεταβολισμό λευκοτριενίων. Η ανάλυση των DEG αποκάλυψε σημαντικό εμπλουτισμό σε μονοπάτια όπως η αποπτωτική διαδικασία των λευκοκυττάρων, η καθοδική πορεία της γλουταθειόνης και η θετική ρύθμιση των ERK1 και ERK2. Η μελέτη αυτή παρέχει λεπτομερή εικόνα της νόσου Αλτσχάιμερ, εντοπίζοντας βασικά διαφοροποιημένα γονίδια σε πολλαπλές περιοχές του εγκεφάλου και αποκαλύπτοντας την πολύπλοκη μοριακή δυναμική που σχετίζεται με την παθολογία της νόσου. Ιδιαίτερα, η ανοδική ρύθμιση γονιδίων όπως τα GSN και TTN σε μονοπάτια που σχετίζονται με την αμυλοείδωση, την τεντωτική δραστηριότητα των μυών και τη σύσπαση του καρδιακού μυός υποδεικνύει πιθανούς θεραπευτικούς στόχους.

Επιπλέον, η θετική ρύθμιση γονιδίων όπως τα CCDC88B και IL1A υποδηλώνει νέες δυνατότητες για τη ρύθμιση της γονιδιακής έκφρασης, ειδικά στη διαμόρφωση της ωρίμανσης των T-κυττάρων και της φλεγμονώδους απόκρισης. Ο εντοπισμός μονοπατιών που σχετίζονται με τη νευροφλεγμονή και τη γλουταμινεργική σηματοδότηση, με γονίδια όπως IL1A, GRM4 και SST, τονίζει τον πιθανό ρόλο τους στις συστηματικές παθολογικές διεργασίες της AD. Οι οδοί αυτές εμφανίζουν σταθερές μεταβολές στις μελετημένες περιοχές, υπογραμμίζοντας τη σημασία της στόχευσης αυτών των μοριακών μηχανισμών για τον μετριασμό της εξέλιξης της νόσου. Τέλος, η περιφερειακή έκφραση γονιδίων όπως τα HDAC1 και ARHGEF40, τα οποία συμμετέχουν σε μονοπάτια μείωσης της ρύθμισης όπως η σηματοδότηση μέσω υποδοχέα θανάτου και υποδοχέα p75 NTR, υποδηλώνει

έναν πολύπλοκο ρυθμιστικό μηχανισμό που μπορεί να προσδίδει συγκεκριμένες περιφερειακές ευπάθειες ή ανθεκτικότητα στην παθολογία της νόσου.

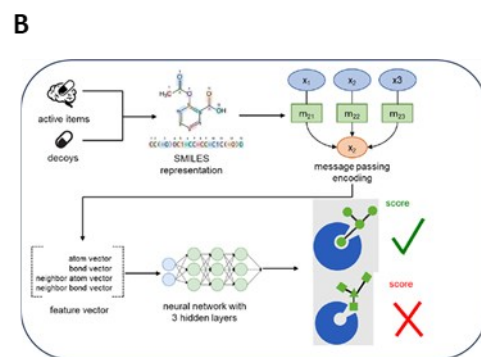
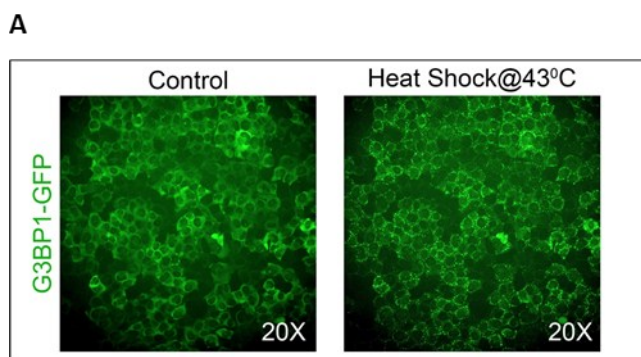
Με την ενσωμάτωση των δεδομένων, η υπολογιστική ανάλυση ανέδειξε νέες μοριακές υπογραφές, επιβεβαιώνοντας ότι τα παρατηρούμενα μοτίβα αντιπροσωπεύουν πραγματικά βιολογικά φαινόμενα και δεν προκύπτουν από τεχνικές παραλλαγές ή σφάλματα στα δεδομένα. Αυτό το βήμα είναι απαραίτητο για να αποφευχθεί οποιαδήποτε επικάλυψη ή σύγχυση κατά τη συγχώνευση, διασφαλίζοντας ότι κάθε κύτταρο, πλέον ενταγμένο σε ένα μεγαλύτερο σύνολο δεδομένων, διατηρεί τη μοναδική του ταυτότητα. Αυτή η σαφής οριοθέτηση είναι θεμελιώδης για μεταγενέστερες αναλύσεις, διασφαλίζοντας ότι τα δεδομένα από διαφορετικά σύνολα δεδομένων μπορούν να συγκριθούν με ακρίβεια. Αυτή η αυστηρή προσέγγιση όχι μόνο ενισχύει την αξιοπιστία των ευρημάτων μας, αλλά δημιουργεί επίσης ένα μεθοδολογικό σχέδιο για μελλοντικές μελέτες με στόχο την αποκωδικοποίηση του πολύπλοκου μοριακού τοπίου της AD. Ενσωματώνοντας γνώσεις από εξωτερικές μελέτες που υποδεικνύουν διακριτά μεταγραφικά δίκτυα στην AD, ιδιαίτερα σε νευρωνικούς και νευρογλοιακούς πληθυσμούς, επιβεβαιώνουμε τα ευρήματά μας σε ένα ευρύτερο επιστημονικό πλαίσιο [Soreq et al 2023]. Επιπλέον, η δυναμική προοπτική της γονιδιακής έκφρασης, που τονίζεται μέσω μελετών ταχύτητας RNA, συμπληρώνει τη στατική μας ανάλυση καταδεικνύοντας τη σημασία της χρονικής δυναμικής στην κατανόηση των κυτταρικών αποκρίσεων στη νόσο Αλτσχάιμερ (Adewale et al 2024).

## 7. Σχεδιασμός και αξιολόγηση εν δυνάμει θεραπευτικών μορίων

Πολλές νευροεκφυλιστικές ασθένειες, συμπεριλαμβανομένου και της νόσου Αλτσχάιμερ χαρακτηρίζονται από τη συσσώρευση πρωτεϊνικών συσσωματωμάτων στο κυτταρόπλασμα των νευρικών κυττάρων των ασθενών, που οδηγούν σε κυτταρικό θάνατο και νευρικό εκφυλισμό. Η πρωτεΐνη G3BP1 (πρωτεΐνη 1 που δεσμεύεται στην πρωτεΐνη ενεργοποίησης της Ras-GTPase) αποτελεί βασικό συστατικό των συσσωματωμάτων στρες (Stress Granules, SGs), κυτταροπλασματικών δομών που σχηματίζονται υπό συνθήκες στρες και απομονώνουν mRNAs και πρωτεΐνες που δεσμεύονται στο RNA (RBPs) (Taylor et al 2016). Τα SGs που περιέχουν G3BP1 εμπλέκονται στην πρόοδο της νόσου Αλτσχάιμερ λόγω της αλληλεπίδρασής τους με mRNA γονιδίων σχετιζόμενων με νόσο Αλτσχάιμερ, συμπεριλαμβανομένων εκείνων που κωδικοποιούν την πρωτεΐνη Tau (MAPT) (Atlas et al 2004). Η G3BP1 δεσμεύεται σε πολυάριθμα mRNAs αλλά και μακρά μη κωδικά RNAs (lncRNAs), συμπεριλαμβανομένων

των mRNAs που σχετίζονται με τη νόσο Αλτσχάιμερ, όπως το MAPT, και τα εγκλωβίζει στα SGs (*Silva et al 2019*). Σε αντίθεση με άλλες πρωτεΐνες SGs όπως η TIA-1, η οποία συσσωματώνεται με συσσωματώματα Ταυ καθώς αυξάνεται η σοβαρότητα της ασθένειας, το G3BP1 σχετίζεται με έναν ξεχωριστό τύπο μοριακής παθολογίας (*Vanderweyde et al 2012*).

Τα SGs θετικά για G3BP1 συχνά δεν περιέχουν κλασικούς δείκτες παθολογίας Ταυ (π.χ., νευροϊνιδιακοί κόμβοι, NFTs), αλλά συσσωρεύονται σε νευρώνες με αυξανόμενη σοβαρότητα της νόσου. Επιπλέον, τα SGs που εμπεριέχουν G3BP1 περιέχουν και mRNAs σχετιζόμενα με τη νόσο Αλτσχάιμερ (π.χ., APP, MAPT), πιθανώς διαταράσσοντας τον μεταβολισμό του RNA (*Sidibé et al 2021*). Αυτή η απομόνωση μπορεί να τροποποιεί την έκφραση πρωτεϊνών, συμβάλλοντας στη νευρωνική δυσλειτουργία. Αναλύσεις δικτύων γονιδίων αποκαλύπτουν σύνδεση μεταξύ της συσώρευσης RNA στους SGs και της διαταραγμένης ομοιόστασης πρωτεϊνών σε εγκεφάλους ασθενών με νόσο Αλτσχάιμερ, τονίζοντας τον ρόλο της G3BP1 στη ρύθμιση της μεταγραφής mRNAs σχετιζόμενων με τη νόσο. Για να εντοπίσουμε καλύτερα στοχευμένες θεραπείες για τη νόσο, χρησιμοποιούμε ένα πειραματικό μοντέλο που σχηματίζει πρωτεϊνικά συσσωματώματα στο κυτταρόπλασμα των κυττάρων, μιμούμενο το σχηματισμό παθολογικών συσσωματωμάτων στην ασθένεια (Εικόνα 7.1). Διάφορα χημικά και περιβαλλοντικά σήματα μπορούν να πυροδοτήσουν το σχηματισμό πρωτεϊνικών συσσωματωμάτων σε αυτά τα κύτταρα (*Yang et al 2020*). Χρησιμοποιώντας αυτά τα κύτταρα που εκφράζουν G3BP1-GFP, στόχο αποτελεί η δημιουργία πειραματικής πλατφόρμας ελέγχου (high throughput screening) για την αξιολόγηση νέων δραστικών μορίων από μεγάλες βιβλιοθήκες ενώσεων παράλληλα με τη χρήση *in silico* τεχνικών για το σχεδιασμό νέων ενώσεων (ενδεικτικά σχεδιάσαμε 75 μικρά εν δυνάμει θεραπευτικά μόρια) και αλγορίθμων μηχανικής μάθησης για την επαναχρησιμοποίηση φαρμάκων μέσω ετερογενών δεδομένων.



**Εικόνα 7.1.** Α) Επαγωγή stress granules μετά από θερμικό στρες σε κύτταρα που εκφράζουν G3BP1 συζευγμένο με GFP. Β) Υπολογιστική μεθοδολογία σχεδιασμού/επανασχεδιασμού/αξιολόγησης εν δυνάμει θεραπευτικών μορίων.

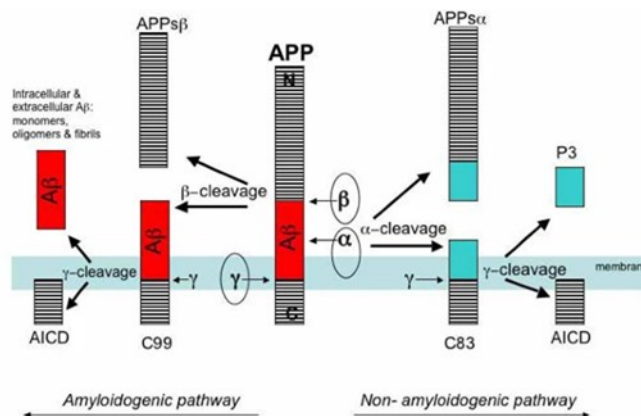
Επίσης, οι εξελίξεις στην υπολογιστική ανάλυση λιπιδικών δεικτών έχουν βελτιώσει σημαντικά την κατανόησή μας σχετικά με τη λιπιδική ρύθμιση στη κυτταρική φυσιολογία και τους μηχανισμούς νόσων. Πραγματοποιήσαμε υπολογιστική ανάλυση λιπιδομικών δεδομένων, με στόχο τη διερεύνηση της λιπιδικής ρύθμισης στην κυτταρική φυσιολογία και στους μηχανισμούς νευροεκφυλιστικών νόσων. Εστίασαμε στην ανάλυση μεταβολών φωσφολιπιδίων σε ανθρώπινα κύτταρα SK-N-SH μετά από έκθεση σε β-αμυλοειδές με στόχο τον εντοπισμό λιπιδικών υπογραφών που θα μπορούσαν να υποδηλώνουν πρώιμες νευροεκφυλιστικές αλλαγές, μέσω της επεξεργασίας και ερμηνείας πολύπλοκων συνόλων δεδομένων. Η ανάλυση των δεδομένων πραγματοποιήθηκε στη γλώσσα προγραμματισμού R, ακολουθώντας ένα αναπαραγώγιμο και κλιμακούμενο υπολογιστικό πλαίσιο εργασίας. Χρησιμοποιήσαμε το πακέτο lipidR για την κανονικοποίηση των δεδομένων, τον σχολιασμό των λιπιδικών ειδών και τη στατιστική αξιολόγηση της διαφορικής αφθονίας, ενώ το ggplot2 αξιοποιήθηκε για την παραγωγή αναλυτικών και υψηλής ποιότητας οπτικοποιήσεων των προτύπων λιπιδικής έκφρασης. Μέσα από την υπολογιστική ανάλυση των δεδομένων, εντοπίσαμε στατιστικά σημαντικές μεταβολές σε συγκεκριμένα φωσφολιπίδια, και συγκεκριμένα στα PE 16:0-18:1, PC 18:2-16:1 και PC 20:2-18:0, υπό συνθήκες έκθεσης στο Αβ πεπτίδιο. Τα αποτελέσματα της ανάλυσης μας υπογραμμίζουν τον καθοριστικό ρόλο της υπολογιστικής επεξεργασίας και της στατιστικής ανάλυσης δεδομένων στη λιπιδωμική έρευνα και αναδεικνύουν την αξία της εφαρμογής τους σε μοντέλα νευρολογικών παθήσεων.

**Πίνακας 7.1** Ανάλυση διαφορικής έκφρασης. Οι τιμές LogFC αποτυπώνουν την κατεύθυνση και το μέγεθος της μεταβολής, με έντονες μειώσεις να παρατηρούνται σε συγκεκριμένες ομάδες PC και PI.

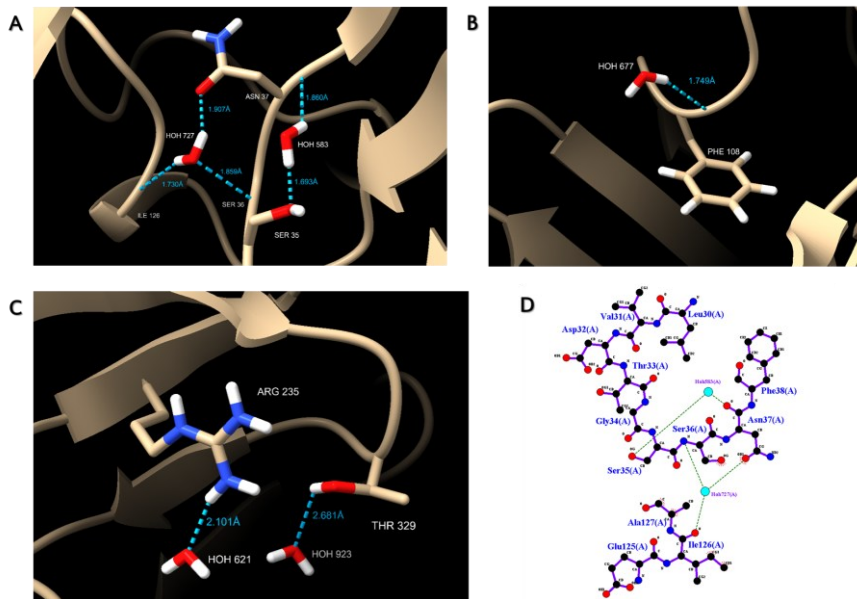
Φωσφολιπίδιο	Ομάδες σύγκρισης	LogFC	p-value
PI 18:0-16:1	Control vs 5nm	0.95	0.019
PI 18:0-18:2	Control vs 5nm	0.89	0.006
PE 16:0-18:1	Control vs 5nm	0.95	0.019
PC 20:2-18:0	Control vs 5nm	0.89	0.006
PI 18:0-18:2	Control vs 50nm	0.92	0.004
PI 18:1-18:2	Control vs 50nm	0.84	0.007
PE 16:0-18:2	Control vs 50nm	0.91	0.002

Τέλος, διερευνήσαμε υπολογιστικά το ρόλο των πεπτιδίων β-αμυλοειδούς, εστιάζοντας στην υπολογιστική αξιολόγηση της αναστολής της β-σεκρετάσης (BACE-1), ενός ενζύμου-

στόχου με ιδιαίτερη σημασία για τη θεραπευτική αντιμετώπιση της AD. Πραγματοποιήσαμε ανάλυση μοριακής πρόσδεσης (molecular docking) σε μια βιβλιοθήκη βάσεων Schiff και ειδικότερα σαλικυλιδενικών ιμινών. Στόχος μας ήταν η εκτίμηση της συγγένειας δέσμευσης τους προς το ένζυμο BACE-1, μέσω υπολογιστικών προσομοιώσεων. Η δομή του BACE-1 που περιείχε συγκρυσταλλωμένο προσδέτη ανακτήθηκε από τη βάση δεδομένων Protein Data Bank (PDB), ενώ οι υπό μελέτη ενώσεις προετοιμάστηκαν με τη χρήση κατάλληλων υπολογιστικών εργαλείων. Διεξήχθησαν δύο ανεξάρτητα υπολογιστικά πειράματα, με και χωρίς τη διατήρηση διατηρημένων μορίων νερού (conserved water molecules, CWM), προκειμένου να αξιολογήσουμε την επίδρασή τους στη διαδικασία της μοριακής πρόσδεσης. Για τον σκοπό αυτό, υπερθέσαμε πολλαπλές δομές και διατηρήσαμε τα μόρια νερού που αλληλεπιδρούν με κρίσιμα κατάλοιπα της θέσης πρόσδεσης. Η επικύρωση του πρωτοκόλλου μοριακής πρόσδεσης πραγματοποιήθηκε μέσω επαναπρόσδεσης (re-docking) του συγκρυσταλλωμένου προσδέτη. Από την ανάλυση των ενεργειών πρόσδεσης και των σταθερών αναστολής προέκυψε ότι ορισμένες ενώσεις από τη βιβλιοθήκη των βάσεων Schiff παρουσιάζουν υπολογιστικά εκτιμώμενη συγγένεια δέσμευσης συγκρίσιμη με εκείνη του συγκρυσταλλωμένου προσδέτη και αλληλεπιδρούν με βασικά κατάλοιπα του ενεργού κέντρου του BACE-1. Τα αποτελέσματα αυτά υποδεικνύουν ότι οι συγκεκριμένες ενώσεις θα μπορούσαν να αποτελέσουν υποψήφιους αναστολείς και να διερευνηθούν περαιτέρω σε πειραματικές in vitro μελέτες.



**Εικόνα 7.2.** Οδοί διάσπασης της APP. Η διάσπαση της APP από την α-σεκρετάση εντός του τμήματος Αβ οδηγεί στη μη αμυλοειδογόνο οδό, ενώ η παθολογική διάσπαση της APP από τη β-σεκρετάση οδηγεί στην αμυλοειδογόνο οδό.



**Εικόνα 7.3.** Τρισδιάστατη (A–C) και δισδιάστατη (D) απεικόνιση των αλληλεπιδράσεων μεταξύ των διατηρημένων μορίων νερού που επιλέχθηκαν για χρήση στα πειράματα μοριακής πρόσδεσης της παρούσας μελέτης και των καταλοίπων της θέσης πρόσδεσης του BACE-1. Οι δεσμοί υδρογόνου απεικονίζονται με διακεκομμένες γραμμές μπλε χρώματος (A–C) ή πράσινου χρώματος (D).

#### 8. Πράκτορες Τεχνητής Νοημοσύνης (AI agents) στη νόσο Αλτσχάιμερ

Οι νευροεκφυλιστικές ασθένειες, όπως η άνοια που σχετίζεται με την νόσο Αλτσχάιμερ και η νόσος του Πάρκινσον, αποτελούν μία από τις μεγαλύτερες προκλήσεις στην παγκόσμια υγειονομική περίθαλψη. Σύμφωνα με την Παγκόσμια Έκθεση για την Νόσο Αλτσχάιμερ, πάνω από 55 εκατομμύρια άνθρωποι ζουν με άνοια, κυρίως λόγω της νόσου Αλτσχάιμερ, ενώ ο αριθμός αυτός προβλέπεται να φτάσει τα 150 εκατομμύρια άτομα παγκοσμίως έως το 2050 (Alzheimer’s Disease International, 2024). Καθώς αυξάνεται η διάρκεια ζωής, όλο και περισσότερα άτομα υποφέρουν από γνωστική έκπτωση ή κινητική δυσλειτουργία, ενώ οι αποτελεσματικές θεραπείες και οι αξιόπιστες μέθοδοι έγκαιρης διάγνωσης παραμένουν περιορισμένες. Αυτό το αυξανόμενο χάσμα μεταξύ του βάρους της νόσου και της θεραπευτικής προόδου υπογραμμίζει την ανάγκη για νέους τρόπους κατανόησης και διαχείρισης αυτών των σύνθετων κλινικών περιστατικών. Η τεχνητή νοημοσύνη (AI) έχει σημειώσει πρόοδο σε αυτόν τον τομέα ως υποστηρικτικό εργαλείο στη διαγνωστική ταξινόμηση, την πρόβλεψη ασθενειών, την ανακάλυψη βιοδεικτών και τη διερεύνηση των μηχανισμών των ασθενειών (van Oostveen και de Lange, 2021). Για παράδειγμα, η ανάλυση απεικονιστικών βιοδεικτών μπορεί να βοηθήσει στην έγκαιρη πρόβλεψη της παθολογίας της νόσου του Αλτσχάιμερ (AD) (Kale et al., 2024).

Ο Πίνακας 8.1 παρουσιάζει μια συνοψη των πρόσφατων εφαρμογών της τεχνητής νοημοσύνης στο πλαίσιο της διαχείρισης της νόσου Αλτσχάιμερ. Αυτό καταδεικνύει τον υποστηρικτικό ρόλο της τεχνητής νοημοσύνης στην έρευνα για τις νευροεκφυλιστικές ασθένειες, από τη διάγνωση των πρώιμων συμπτωμάτων έως την πρόβλεψη των ασθενών με ήπια γνωστική δυσλειτουργία που θα εξελιχθούν σε άνοια (Ai et al., 2025). Ωστόσο, τα περισσότερα από τα τρέχοντα εργαλεία τεχνητής νοημοσύνης στην ιατρική παραμένουν στενά εστιασμένα, με ισχυρή απόδοση σε μεμονωμένες εργασίες, όπως η ταξινόμηση MRI ή η ανάλυση ομιλίας, αλλά με περιορισμένη ικανότητα γενίκευσης χρησιμοποιώντας πολυτροπικά δεδομένα. Πρόσφατα στοιχεία επιβεβαιώνουν ότι τα κλινικά συστήματα τεχνητής νοημοσύνης παραμένουν σε μεγάλο βαθμό μονοτροπικά και στερούνται μηχανισμών για την ενσωμάτωση ετερογενών δεδομένων ή την υποστήριξη διαχρονικής συλλογιστικής (Winchester et al., 2023; Judge et al., 2024). Λειτουργούν ως παθητικοί αλγόριθμοι και όχι ως ενεργοί συνεργάτες και συχνά λειτουργούν ως μαύρα κουτιά με περιορισμένη ερμηνευσιμότητα (Rudin, 2019; Xu και Shuttleworth, 2024). Στο παρόν άρθρο, υποστηρίζουμε ότι το μέλλον της ιατρικής τεχνητής νοημοσύνης είναι να προχωρήσει πέρα από τα στενά, ειδικά για κάθε εργασία μοντέλα προς αυτόνομους παράγοντες που μπορούν να συλλογιστούν σε διάφορες πηγές δεδομένων, να χρησιμοποιήσουν πολλαπλά υπάρχοντα εργαλεία τελευταίας τεχνολογίας για την υποστήριξη της λήψης αποφάσεων και τον υπολογισμό, να προσαρμοστούν μέσω της εμπειρίας και να συνεργαστούν αποτελεσματικά με τους κλινικούς γιατρούς. Εξετάζουμε συγκεκριμένα τον τρέχοντα ρόλο της τεχνητής νοημοσύνης στη φροντίδα της νόσου του Αλτσχάιμερ, περιγράφουμε τα βασικά εμπόδια που σχετίζονται με την ενσωμάτωση, την ερμηνευσιμότητα και τη ρύθμιση, λαμβάνοντας υπόψη τις μελλοντικές κατευθύνσεις για την εισαγωγή αυτών των παραγόντων στην κλινική πρακτική.

Πίνακας 8.1. Πρόσφατες εφαρμογές τεχνητής νοημοσύνης στο πλαίσιο της διαχείρισης της νόσου του Αλτσχάιμερ.

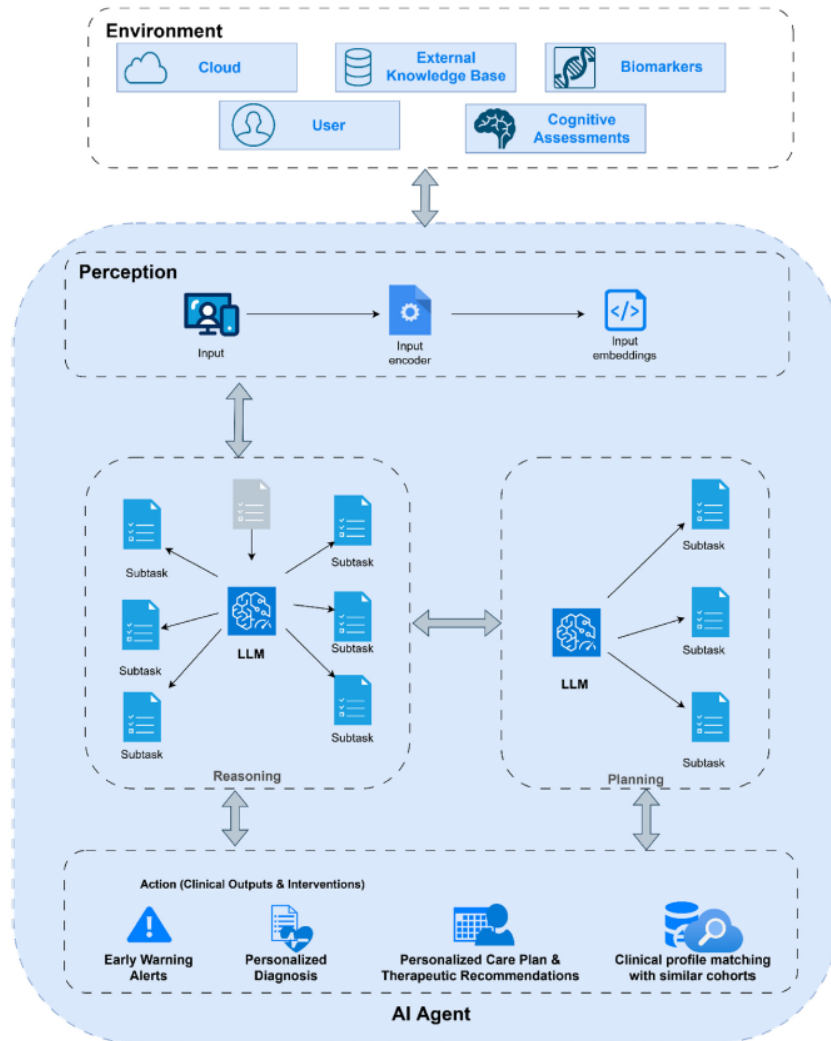
Ερευνητικός τομέας	Μεθοδολογία	Μορφή	Αποτέλεσμα	Αναφορές
Διαγνωστική ή ταξινόμηση	Βαθιά μάθηση με ανάλυση λανθάνοντος χώρου	Δομική μαγνητική τομογραφία	Ταξινομεί τις αλλαγές στο δίκτυο του εγκεφάλου που σχετίζονται με το φύλο με υψηλή ακρίβεια (> 90%).	Wang et al., 2024
	Συγχώνευση πολυτροπικών δεδομένων	MRI + βιοδείκτες πλάσματος	Ανιχνεύει την κατάσταση του αμυλοειδούς με AUC 0,94	Jasodanan et al., 2025

	Μοντέλα EEG/ηλεκτροφυσιολογικής τεχνητής νοημοσύνης	Καταγραφή EEG	Προσδιορισμός των ζωνών συχνότητων EEG που είναι καθοριστικές για τη διάγνωση της νόσου του Αλτσχάιμερ, μετρήσεις ισορροπίας, ανίχνευση της νόσου του Αλτσχάιμερ σε πραγματικό χρόνο	<b>Arikan et al., 2025</b>
Πρόβλεψη ασθένειας	Μάθηση με γνώση-γράφημα	EHR	Προβλέπει την εμφάνιση της νόσου Αλτσχάιμερ 7 χρόνια νωρίτερα (AUC 0,80).	<b>Tang et al., 2024</b>
	Μοντελοποίηση τροχιάς	Διαχρονική μαγνητική τομογραφία + CSF (Aβ42)	Προβλέπει τη μετατροπή της MCI σε AD εντός 3 ετών με ακρίβεια 85%.	<b>Agostinho et al., 2024</b>
Ανακάλυψη βιοδεικτών	Επιλογή χαρακτηριστικών και ταξινομητές μηχανικής μάθησης	Έκφραση γονιδίων στο αίμα από ADNI, ANM1, ANM2	Ταξινόμηση της νόσου Αλτσχάιμερ (AUC: 0,619–0,859)· ταυτοποίηση των μονοπατιών που σχετίζονται με τη νόσο Αλτσχάιμερ (φλεγμονή, μιτοχόνδρια, σηματοδότηση Wnt)	<b>Lee και Lee, 2020</b>
	Κλασική μηχανική μάθηση	Ψηφιακοί βιοδείκτες: κινητική δραστηριότητα, νευρογνωστικές δοκιμασίες, παρακολούθηση των ματιών, ανάλυση ομιλίας	Πρόβλεψη/ταξινόμηση της AD (AUC ~0,887) και της MCI (AUC ~0,821)· ενσωμάτωση στην κλινική πρακτική	<b>Qi et al., 2025</b>
Εξερεύνηση μηχανισμών	Ανάλυση συσχέτισης GS; ενσωμάτωση απεικόνισης-μεταγραφωμικής	fMRI; γονιδιακή έκφραση	Αλλοιωμένη τοπογραφία GS στην AD; συσχέτιση με τις λειτουργικές ιδιότητες του δικτύου και τη γνωστική λειτουργία	<b>Chen et al., 2023</b>
	Χρήση transfer learning με νευρωνικά δίκτυα CNNs	MRI	Ακριβής διάγνωση της νόσου Αλτσχάιμερ (με ακρίβεια έως 96%) με ερμηνεύσιμες οπτικές εξηγήσεις που επισημαίνουν τις βασικές νευρικές περιοχές	<b>Mahmud et al., 2024</b>

Οι πράκτορες τεχνητής νοημοσύνης μπορούν να ταξινομηθούν σε διαφορετικές κατηγορίες με βάση την εσωτερική τους αρχιτεκτονική και τον βαθμό αυτονομίας τους. Οι αντιδραστικοί πράκτορες αποτελούν την απλούστερη μορφή, καθώς ανταποκρίνονται άμεσα στα ερεθίσματα του περιβάλλοντος χωρίς να διατηρούν μνήμη ή να αξιοποιούν προηγούμενες εμπειρίες. Η απουσία μηχανισμών μνήμης τους καθιστά ιδιαίτερα κατάλληλους για εφαρμογές που απαιτούν ταχεία απόκριση ή συνεχή και επαναλαμβανόμενη αξιολόγηση δυναμικά μεταβαλλόμενων συνθηκών (Sarkota et al., 2025). Οι διαλογικοί πράκτορες λειτουργούν σε υψηλότερο επίπεδο πολυπλοκότητας, καθώς διατηρούν εσωτερικές αναπαραστάσεις του περιβάλλοντός τους και είναι σε θέση να σχεδιάζουν μελλοντικές ενέργειες βάσει προβλέψεων για τα πιθανά αποτελέσματα. Πριν επιλέξουν μια συγκεκριμένη πορεία δράσης, εξετάζουν εναλλακτικά σενάρια, γεγονός που τους καθιστά ιδιαίτερα χρήσιμους για την αναπαράσταση και κατανόηση μακροχρόνιων και σύνθετων βιολογικών διεργασιών, όπως η εξέλιξη μιας νόσου. Στο πλαίσιο της έρευνας για τις νευροεκφυλιστικές παθήσεις, τέτοιοι πράκτορες μπορούν να αξιοποιηθούν για την προσομοίωση της αλληλεπίδρασης κυτταρικών, μοριακών και συμπεριφορικών παραγόντων στον χρόνο, υποστηρίζοντας τη διατύπωση και τον έλεγχο υποθέσεων σχετικά με τη γήρανση του εγκεφάλου ή την πορεία της άνοιας. Οι υβριδικοί πράκτορες συνδυάζουν χαρακτηριστικά αντιδραστικών και στοχαστικών προσεγγίσεων, επιτρέποντας αφενός άμεση απόκριση σε μεταβαλλόμενα ερεθίσματα και αφετέρου τη διατήρηση στρατηγικών, μακροπρόθεσμων ικανοτήτων συλλογιστικής και σχεδιασμού. (Lanham, 2025). Η αρχιτεκτονική τους τους επιτρέπει να εξισορροπούν την ερμηνεία δεδομένων σε πραγματικό χρόνο με ευρύτερους αναλυτικούς στόχους. Αυτή η ευελιξία μπορεί να είναι πολύτιμη σε διαχρονικές βιοϊατρικές μελέτες, όπου τα συστήματα πρέπει να προσαρμόζονται σε νέα δεδομένα, διατηρώντας παράλληλα τη συνέχεια με προηγούμενες αναλύσεις. Μπορούν να προσδιοριστούν λειτουργικές διακρίσεις μεταξύ πρακτόρων που βασίζονται σε στόχους και πρακτόρων που βασίζονται στη χρησιμότητα (Matsumoto et al., 2006). Οι πράκτορες που βασίζονται σε στόχους λειτουργούν με βάση προκαθορισμένους στόχους, προσαρμόζοντας συνεχώς τις αποφάσεις τους για την επίτευξη αυτών των στόχων. Είναι ιδιαίτερα χρήσιμοι σε ερευνητικές διαδικασίες που βασίζονται σε δεδομένα και αποσκοπούν στη βελτιστοποίηση της απόδοσης του μοντέλου ή στη βελτίωση της ακρίβειας των προβλέψεων. Οι πράκτορες που βασίζονται στη χρησιμότητα βασίζονται σε αυτήν την αρχή, αξιολογώντας τη σχετική αξία των πιθανών αποτελεσμάτων και επιλέγοντας ενέργειες που μεγιστοποιούν το συνολικό όφελος σύμφωνα με συγκεκριμένα κριτήρια αξιολόγησης. Αυτοί οι πράκτορες είναι ιδιαίτερα πολύτιμοι σε ερευνητικά περιβάλλοντα που απαιτούν συμβιβασμούς μεταξύ ευαισθησίας, ειδικότητας και ερμηνευσιμότητας, για παράδειγμα, διακρίνοντας μεταξύ των πρώιμων και των όψιμων σταδίων της παθολογίας του Αλτσχάιμερ.

Οι μαθησιακοί πράκτορες αποτελούν μια άλλη σημαντική κατηγορία, που χαρακτηρίζεται από την ικανότητά τους να βελτιώνονται μέσω της ανατροφοδότησης (Axelsson et al., 2022). Βελτιώνουν τις διαδικασίες λήψης αποφάσεων καθώς συσσωρεύουν εμπειρία, γεγονός που τους καθιστά ιδιαίτερα σημαντικούς σε περιβάλλοντα όπου το υποκείμενο σύστημα εξελίσσεται ή όπου τα διαθέσιμα δεδομένα αυξάνονται με την πάροδο του χρόνου. Η ενισχυτική μάθηση, μια βασική προσέγγιση αυτής της κατηγορίας, επιτρέπει στους πράκτορες να μαθαίνουν βέλτιστες συμπεριφορές μέσω επαναληπτικής εξερεύνησης και αξιολόγησης (Wang F. et al., 2025). Αν και η πλήρης εφαρμογή της στην έρευνα των νευροεκφυλιστικών ασθενειών βρίσκεται ακόμη σε αρχικό στάδιο, τέτοιες προσαρμοστικές μέθοδοι μάθησης θα μπορούσαν κάποια μέρα να συμβάλουν στη βελτίωση διαγνωστικών εντοπισμό νέων θεραπευτικών στόχων, ενημερώνοντας συνεχώς τη συλλογιστική τους καθώς εμφανίζονται νέα στοιχεία. Τέλος, τα συστήματα πολλαπλών παραγόντων συγκεντρώνουν διάφορες αυτόνομες οντότητες που συνεργάζονται ή συντονίζονται για την επίτευξη κοινών στόχων (Rother et al., 2025). Μπορούν να χωρίσουν σύνθετα αναλυτικά προβλήματα σε μικρότερες, εξειδικευμένες εργασίες και να επικοινωνούν μεταξύ τους για να σχηματίσουν μια συλλογική νοημοσύνη μεγαλύτερη από το άθροισμα των μερών της. Αν και η ενσωμάτωσή τους στη νευροεπιστήμη και τη βιοϊατρική έρευνα παραμένει σε μεγάλο βαθμό θεωρητική, τέτοια συστήματα είναι πολλά υποσχόμενα για τη διαχείριση μεγάλων συνόλων δεδομένων που καλύπτουν μοριακές, απεικονιστικές και κλινικές διαστάσεις, επιτρέποντας ενδεχομένως τη συνεργατική συλλογιστική μεταξύ διασυνδεδεμένων μοντέλων.

Η ανάπτυξη αυτών των διαφόρων μορφών πρακτόρων καταδεικνύει την ευελιξία και το βάθος του παραδείγματος της πρακτικής τεχνητής νοημοσύνης. Ενώ η τρέχουσα χρήση τους στην έρευνα για τις νευροεκφυλιστικές ασθένειες βρίσκεται ακόμη σε διερευνητικό στάδιο, οι θεωρητικές τους δυνατότητες ευθυγραμμίζονται στενά με τις ανάγκες του τομέα (Perry et al., 2024). Προσφέρουν πλαίσια για συνεχή μάθηση, προσαρμοστική συλλογιστική και κατανεμημένη ανάλυση, ιδιότητες που αντικατοπτρίζουν την προοδευτική και πολυπαραγοντική φύση διαταραχών όπως η νόσος του Alzheimer (Εικόνα 8.1). Καθώς αυτές οι τεχνολογίες εξελίσσονται, η κύρια πρόκληση θα είναι η μετατροπή της εννοιολογικής τους υπόσχεσης σε πρακτικά, διαφανή και ερμηνεύσιμα εργαλεία που μπορούν να λειτουργούν υπεύθυνα στο πλαίσιο της επιστημονικής και κλινικής έρευνας.



**Εικονά 8.1.** Αρχιτεκτονικό πλαίσιο πράκτορα τεχνητής νοημοσύνης για τη νόσο του Αλτσχάιμερ. Ο πράκτορας αλληλοεπιδρά με περιβαλλοντικά δεδομένα, συμπεριλαμβανομένων βιοδεικτών και γνωστικών αξιολογήσεων.

Αυτά τα συστήματα επιδεικνύουν περιορισμένη νοημοσύνη με την εφαρμογή επεξεργασίας υψηλού επιπέδου σε μεμονωμένες, περιορισμένες κλινικές εργασίες χωρίς ικανότητα γενίκευσης ή πολυεπίπεδης συλλογιστικής (Adler- Milstein et al., 2022). Τα περισσότερα μοντέλα λειτουργούν με μονοτροπικά δεδομένα, πραγματοποιώντας ανάλυση σε βιοδείκτες νευροαπεικόνισης ή διεξάγοντας κλινικές αξιολογήσεις μεμονωμένα (Shokrour et al., 2025). Ενώ η βαθιά μάθηση επιτυγχάνει ακρίβεια μεγαλύτερη από 80% σε συγκεκριμένες εργασίες ταξινόμησης (Venugoralan et al., 2021), αυτά τα συστήματα παρουσιάζουν υποβάθμιση της απόδοσης όταν συναντούν ελλιπή δεδομένα ή σενάρια βασισμένα στο πλαίσιο που δεν ήταν μέρος του συνόλου εκπαίδευσης. Επιπλέον, κανένα τρέχον σύστημα τεχνητής νοημοσύνης δεν λαμβάνει συνεχώς ανατροφοδότηση σχετικά με τη διαγνωστική απόδοση, περιορίζοντας την ικανότητα κλινικής μάθησης και προσαρμογής (Ahmed et al., 2023).

Οι πράκτορες τεχνητής νοημοσύνης μπορούν να αλλάξουν εντελώς αυτή την κατάσταση. Σε αντίθεση με τις παραδοσιακές μεθοδολογίες μηχανικής μάθησης και βαθιάς μάθησης που αντιστοιχούν τις εισόδους με τις εξόδους, οι πράκτορες τεχνητής νοημοσύνης εκτελούν συμπεριφορές προσανατολισμένες προς τον στόχο, οι οποίες χωρίζουν τους πολύπλοκους στόχους υψηλού επιπέδου σε μικρότερες υπο-εργασίες, διατηρώντας παράλληλα την πιθανοτική λήψη αποφάσεων. Πρόσφατες εργασίες καταδεικνύουν αυτή την αλλαγή. Η Google παρουσίασε το σύστημα ΑΙΜΕ, το οποίο χρησιμοποιεί συλλογιστική με γνώση της κατάστασης μέσω διαγνωστικών φάσεων και υπερτερεί των κλινικών ιατρών σε 29 από τις 32 αξιολογήσεις (Saab et al., 2025). Το MedAgent-Pro εισήγαγε ιεραρχικές ροές εργασίας με σχεδιασμό σε επίπεδο νόσου και συλλογιστική σε επίπεδο ασθενούς (Wang Z. et al., 2025). Για τις νευροεκφυλιστικές νόσους, οι πράκτορες μπορούν να παρέχουν καθοδήγηση που είναι αδύνατο να επιτευχθεί με περιορισμένα εργαλεία. Αντί να παράγει μια απλή ετικέτα ταξινόμησης, ένας πράκτορας μπορεί να ακολουθήσει μια διαδικασία συλλογιστικής πολλαπλών βημάτων. Μπορεί να αναλύσει τα μοτίβα ατροφίας στη μαγνητική τομογραφία, να ερμηνεύσει τις αναλογίες των βιοδεικτών του εγκεφαλονωτιαίου υγρού, να ανακτήσει πληροφορίες από ηλεκτρονικά αρχεία υγείας σχετικά με γενετικούς παράγοντες κινδύνου και να αξιολογήσει την γνωστική πορεία του ασθενούς στο χρόνο. Μπορεί επίσης να εξετάσει τα τρέχοντα διαγνωστικά κριτήρια, να δημιουργήσει μια πιθανοτική διαφορική διάγνωση και να προτείνει επιβεβαιωτικές αξιολογήσεις όταν η αβεβαιότητα παραμένει υψηλή. Μια μελέτη του Nature Medicine του 2024, στην οποία συμμετείχαν περισσότεροι από 51.000 συμμετέχοντες, έδειξε ότι ένα πολυτροπικό μοντέλο έφτασε σε AUROC 0,96 στη διάκριση δέκα τύπων άνοιας, ξεπερνώντας την αξιολόγηση μόνο από νευρολόγους κατά περισσότερο από 26% (Xue et al., 2024). Ωστόσο, ακόμη και αυτό παραμένει ένας στατικός ταξινομητής. Τα συστήματα με πράκτορες βασίζονται σε τέτοια μοντέλα με την εισαγωγή δυναμικής συλλογιστικής που τους επιτρέπει να ζητούν πρόσθετο πλαίσιο για τα παρεχόμενα δεδομένα όταν μειώνεται η εμπιστοσύνη και να παρακολουθούν τις διαχρονικές πορείες της νόσου για να βελτιώσουν την κατανόηση με την πάροδο του χρόνου. Βασικά μοντέλα όπως το ChatGPT, το Claude και ακόμη πιο κλινικά εστιασμένα όπως το Med-PaLM2 παρέχουν πλέον την ευφυΐα που απαιτείται για αυτά τα συστήματα ώστε να επιτύχουν απόδοση επιπέδου εμπειρογνομόνων σε ένα ευρύ φάσμα ιατρικών εργασιών (Singhal et al., 2023, 2025). Αποδεικτικά στοιχεία από μια ανασκόπηση περισσότερων από τετρακόσιων μελετών δείχνουν ότι οι πολυτροπικές προσεγγίσεις υπερτερούν σταθερά των μοντέλων μονής τροπικότητας κατά περίπου έξι ποσοστιαίες μονάδες στο AUC (Schouten et al., 2025). Οι αρχιτεκτονικές με πράκτορες θα μπορούσαν να μετατρέψουν αυτά τα κέρδη απόδοσης σε πραγματική κλινική αξία μέσω της

ενσωμάτωσης της συλλογιστικής, της μνήμης και των προσαρμοστικών διαδικασιών λήψης αποφάσεων που αντικατοπτρίζουν την ανθρώπινη διαγνωστική σκέψη.

## 9. Συμπεράσματα

Πραγματοποιήθηκε μια ολοκληρωμένη προσέγγιση για την ενσωμάτωση συνόλων δεδομένων scRNA-seq από διάφορα δείγματα ελέγχου που αντιστοιχούν σε διαφορετικές περιοχές του ανθρώπινου εγκεφάλου. Η ανάλυση διεξήχθη προκειμένου να εντοπιστούν κοινά σημεία ή αποκλίσεις σε διαφορετικές περιοχές του εγκεφάλου, κάτι που απαιτεί σχολαστική προεπεξεργασία για να διασφαλιστεί η ποιότητα και η συγκρισιμότητα των δεδομένων. Μέσω αυτής της διαδικασίας, αξιοποιήθηκαν οι δυνατότητες του scRNA-seq για να γεφυρωθεί αυτό το κενό, προσφέροντας μια λεπτομερή ανάλυση σε πολλές περιοχές του εγκεφάλου που επηρεάζονται περισσότερο από την AD, όπως ο ενδορινικός φλοιός, ο προμετωπιαίος φλοιός, η άνω μετωπιαία έλικα και ο άνω βρεγματικός λοβός. Χρησιμοποιώντας εξελιγμένες μεθόδους μηχανικής μάθησης, πραγματοποιώντας ενδελεχή αξιολόγηση διαφόρων αλγορίθμων στο σύνολο δεδομένων γονιδιακής έκφρασης και εφαρμόζοντας διεξοδικές μεθόδους αξιολόγησης, η ανάλυσή μας παρέχει λεπτομερή εικόνα των προτύπων γονιδιακής έκφρασης σε διαφορετικές καταστάσεις. Η παρούσα υπολογιστική ανάλυση, που ενισχύεται από τη βιβλιοθήκη PyCaret, προσφέρει πολύτιμες γνώσεις για τις γενετικές υπογραφές που σχετίζονται με διαφορετικές καταστάσεις και μας δίνει τη δυνατότητα να αξιολογήσουμε τις κυτταρικές και μοριακές αλλοιώσεις μεταξύ ατόμων με AD και υγιών ατόμων ελέγχου, επισημαίνοντας τις μοναδικές και αλληλεπικαλυπτόμενες οδούς εκφυλισμού σε διαφορετικές περιοχές του εγκεφάλου. Με τη συγχώνευση και ανάλυση δεδομένων από διαφορετικές περιοχές, η παρούσα μελέτη αξιοποιεί υπάρχοντα σύνολα δεδομένων scRNA-seq, αναδεικνύοντας τις δυνατότητες της υπολογιστικής ανάλυσης για την παροχή βαθιάς κατανόησης της πολύπλοκης βιολογίας της AD σε επίπεδο πολλών περιοχών του εγκεφάλου. Επιπλέον, πραγματοποιήθηκε ενοποίηση συνόλων δεδομένων ελέγχου και ασθενειών όλων των περιοχών και διεξήχθησαν διαφορές γονιδιακής έκφρασης μεταξύ υγιών κυττάρων και εκείνων νοσούντων που AD. Αυτά τα ευρήματα ανοίγουν το δρόμο για περαιτέρω έρευνα στον εντοπισμό και την επικύρωση βασικών γενετικών δεικτών για ασθένειες, προωθώντας έτσι την κατανόησή μας για την παθολογία της νόσου και υποστηρίζοντας την ανάπτυξη μελλοντικών διαγνωστικών και θεραπευτικών προσεγγίσεων.

Παράλληλα, η τεχνητή νοημοσύνη με πράκτορες αντιπροσωπεύει μια νέα φάση των ευφύων συστημάτων, που χαρακτηρίζεται από συλλογιστική, προσαρμοστικότητα

και σκόπιμη δράση. Αν και η εφαρμογή της στη νευροεκφυλιστική έρευνα βρίσκεται ακόμη σε αρχικό στάδιο, οι θεωρητικές της δυνατότητες συνάδουν στενά με τις ανάγκες της έρευνας για τη νόσο του Αλτσχάιμερ. Επιτρέποντας τη δυναμική, ευαίσθητη στο περιβάλλον ανάλυση δεδομένων απεικόνισης, γονιδιωματικής, γνωστικής και συμπεριφοράς, η τεχνητή νοημοσύνη με πράκτορες προσφέρει πρωτοφανείς δυνατότητες για την παρακολούθηση της εξέλιξης της νόσου και τον προσδιορισμό θεραπευτικών στόχων. Ως μελλοντική προοπτική, αυτοί οι παράγοντες ενδέχεται να προχωρήσουν σταδιακά πέρα από την παθητική ανάλυση προς μοντέλα που μπορούν να ανιχνεύουν πότε οι δικές τους πληροφορίες είναι ελλιπείς και να ξεκινήσουν στοχευμένες αιτήσεις δεδομένων για τη μείωση της αβεβαιότητας. Αυτή η αλλαγή θα φέρει τα συστήματα τεχνητής νοημοσύνης πιο κοντά στην πραγματική κλινική συλλογιστική, διατηρώντας παράλληλα τον έλεγχο στους κλινικούς γιατρούς. Η διασφάλιση της ηθικής και της επικεντρωμένης στον ασθενή ανάπτυξης μέσω της στενής συνεργασίας μεταξύ τεχνικών και κλινικών ομάδων παραμένει ουσιαστική, μετατοπίζοντας την εστίαση από το αν η τεχνητή νοημοσύνη θα αναδιαμορφώσει την περίθαλψη στο πώς μπορεί να υποστηρίξει ουσιαστικά την κλινική εμπειρογνωμοσύνη και τα αποτελέσματα των ασθενών.

#### Βιβλιογραφία

1. Starr, P. A., Vitek, J. L., & Bakay, R. A. (1998). Ablative surgery and deep brain stimulation for Parkinson's disease. *Neurosurgery*, 43(5), 989–1013.
2. Kashiwakura, J. I., Suzuki, N., Nagafuchi, H., Takeno, M., Takeba, Y., Shimoyama, Y., & Sakane, T. (1999). Txk, a nonreceptor tyrosine kinase of the Tec family, is expressed in T helper type 1 cells and regulates interferon  $\gamma$  production in human T lymphocytes. *Journal of Experimental Medicine*, 190(8), 1147–1154.
3. Le Couteur, D. G., McLean, A. J., Taylor, M. C., Woodham, B. L., & Board, P. G. (1999). Pesticides and Parkinson's disease. *Biomedicine & Pharmacotherapy*, 53(3), 122–130.
4. Martinez, W. M., & Spear, P. G. (2001). Structural Features of Nectin-2 (HveB) Required for Herpes Simplex Virus Entry. *Journal of Virology*, 75(22), 11185
5. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 13(11), 2498.
6. Van Den Eeden, S. K., Tanner, C. M., Bernstein, A. L., Fross, R. D., Leimpeter, A., Bloch, D. A., & Nelson, L. M. (2003). Incidence of Parkinson's Disease: Variation by Age, Gender, and Race/Ethnicity. *American Journal of Epidemiology*, 157(11), 1015–1022.
7. Yamaoka, K., Saharinen, P., Pesu, M., Holt, V. E., Silvennoinen, O., & O'Shea, J. J. (2004). The Janus kinases (Jaks). *Genome Biology*, 5(12), 1–6.

8. Tanner, C. M., & Goldman, S. M. (2005). EPIDEMIOLOGY OF PARKINSON'S DISEASE. *Neurologic Clinics*, 14(2), 317.
9. Lim, K. L., & Tan, J. M. (2007). Role of the ubiquitin proteasome system in Parkinson's disease.
10. *BMC Biochemistry*, 8(Suppl 1), S13.
11. Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(1), 1–21.
12. Simunovic, F., Yi, M., Wang, Y., Stephens, R., & Sonntag, K. C. (2010). Evidence for Gender Specific Transcriptional Profiles of Nigral Dopamine Neurons in Parkinson Disease. *PLOS ONE*, 5(1), e8856.
13. Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., Coffey, C., Kieburtz, K., Flagg, E., Chowdhury, S., Poewe, W., Mollenhauer, B., Sherer, T., Frasier, M., Meunier, C., Rudolph, A., Casaceli, C., Seibyl, J., Mendick, S., ... Taylor, P. (2011). The Parkinson Progression Marker Initiative (PPMI). *Progress in Neurobiology*, 95(4), 629–635.
14. Wirdefeldt, K., Adami, H.-O., Cole, P., Trichopoulos, D., Mandel, J., Wirdefeldt, K., Adami, H.-O., Trichopoulos, Á. D., Cole, P., & Mandel, J. (2011). Epidemiology and etiology of Parkinson's disease: a review of the evidence. *European Journal of Epidemiology* 2011 26:1, 26(1), 1–58.
15. Croom, E. (2012). Metabolism of Xenobiotics of Human Environments. *Progress in Molecular Biology and Translational Science*, 112, 31–88.
16. Kele, J., Andersson, E. R., Villaescusa, J. C., Cajanek, L., Parish, C. L., Bonilla, S., Toledo, E. M., Bryja, V., Rubin, J. S., Shimono, A., & Arenas, E. (2012). SFRP1 and SFRP2 Dose-Dependently Regulate Midbrain Dopamine Neuron Development In Vivo and in Embryonic Stem Cells. *Stem cells (Dayton, Ohio)*, 30(5), 865.
17. Scholz, S. W., Mhyre, T., Resson, H., Shah, S., & Federoff, H. J. (2012). Genomics and Bioinformatics of Parkinson's Disease. *Cold Spring Harbor Perspectives in Medicine*, 2(7), a009449.
18. Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R., & Ma'ayan, A. (2013). Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, 14.
19. Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general-purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7), 923–930.
20. Salvatore, C., Cerasa, A., Castiglioni, I., Gallivanone, F., Augimeri, A., Lopez, M., Arabia, G., Morelli, M., Gilardi, M. C., & Quattrone, A. (2014). Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and Progressive Supranuclear Palsy. *Journal of Neuroscience Methods*, 222, 230–237.
21. Chen, Y., Jiao, B., Yao, M., Shi, X., Zheng, Z., Li, S., & Chen, L. (2017). ISG12a inhibits HCV replication and potentiates the anti-HCV activity of IFN- $\alpha$  through activation of the Jak/STAT signaling pathway independent of autophagy and apoptosis. *Virus Research*, 227, 231–239.
22. Tarakad, A., & Jankovic, J. (2017). Anosmia and Ageusia in Parkinson's Disease. *International Review of Neurobiology*, 133, 541–556.
23. Tysnes, O. B., & Storstein, A. (2017). Epidemiology of Parkinson's disease. *Journal of Neural Transmission*, 124(8), 901–905.
24. Chiabrando, D., Fiorito, V., Petrillo, S., & Tolosano, E. (2018). Unraveling the role of heme in neurodegeneration. *Frontiers in Neuroscience*, 12(OCT), 414859.

25. Koller, W. C. (2018). Table 1, The clinical diagnostic criteria for Parkinson's disease, based on the Movement Disorder Society guidelines. *Neurology*, 43(2), 1–11.
26. Balestrino, R., & Schapira, A. H. (2020). Parkinson disease. *European Journal of Neurology*, 27(1), 27–42.
27. Jankovic, J., & Tan, E. K. (2020). Parkinson's disease: etiopathogenesis and treatment. *Journal of Neurology, Neurosurgery & Psychiatry*, 91(8), 795–808.
28. Wang, W., Lee, J., Harrou, F., & Sun, Y. (2020). Early Detection of Parkinson's Disease Using Deep Learning and Machine Learning. *IEEE Access*, 8, 147635–147646.
29. Bloem, B. R., Okun, M. S., & Klein, C. (2021). Parkinson's disease. *The Lancet*, 397(10291), 2284–2303.
30. Lashgari, N. A., Roudsari, N. M., Momtaz, S., Sathyapalan, T., Abdolghaffari, A. H., & Sahebkar, A. (2021). The involvement of the JAK/STAT signaling pathway in the treatment of Parkinson's disease. *Journal of Neuroimmunology*, 361.
31. Lee, T. K., & Yankee, E. L. (2021). A review on Parkinson's disease treatment. *Neurosciences* 2021;8:222., 8(0), N/A–N/A.
32. Li, B., Zhao, G., Zhou, Q., Xie, Y., Wang, Z., Fang, Z., Lu, B., Qin, L., Zhao, Y., Zhang, R., Jiang, L., Pan, H., He, Y., Wang, X., Luo, T., Zhang, Y., Wang, Y., Chen, Q., Liu, Z., ... Li, J. (2021). Gene4PD: A Comprehensive Genetic Database of Parkinson's Disease. *Frontiers in Neuroscience*, 15, 679568.
33. Sassone, J., Reale, C., Dati, G., Regoni, M., Pellecchia, M. T., & Garavaglia, B. (2021). The Role of VPS35 in the Pathobiology of Parkinson's Disease. *Cellular and Molecular Neurobiology*, 41(2), 199–227.
34. Yang, C., & Wang, X. (2021). Lysosome biogenesis: Regulation and functions. *Journal of Cell Biology*, 220(6).
35. Zhao, Y., Li, M. C., Konaté, M. M., Chen, L., Das, B., Karlovich, C., Williams, P. M., Evrard, Y. A., Doroshov, J. H., & McShane, L. M. (2021). TPM, FPKM, or Normalized Counts? A Comparative Study of Quantification Measures for the Analysis of RNA-seq Data from the NCI Patient-Derived Models Repository. *Journal of Translational Medicine*, 19(1), 1–15.
36. Maitin, A. M., Romero Muñoz, J. P., & García-Tejedor, Á. J. (2022). Survey of Machine Learning Techniques in the Analysis of EEG Signals for Parkinson's Disease: A Systematic Review *Applied Sciences* 2022, Vol. 12, Page 6967, 12(14), 6967.
37. Smith, L., & Schapira, A. H. (2022). GBA Variants and Parkinson Disease: Mechanisms and Treatments. *Cells*, 11(8).
38. Wang, X., Shi, N., Wu, B., Yuan, L., Chen, J., Ye, C., & Hao, M. (2022). Bioinformatics Analysis of Gene Expression Profiles and Functional Pathways Linking Periodontitis and Parkinson's Disease. *Frontiers in Aging Neuroscience*, 14, 1029637.
39. Wasner, K., Smajic, S., Ghelfi, J., Delcambre, S., Prada-Medina, C. A., Knappe, E., Arena, G., Mulica, P., Agyeah, G., Rakovic, A., Boussaad, I., Badanjak, K., Ohnmacht, J., Gérardy, J. J., Takanashi, M., Trinh, J., Mittelbronn, M., Hattori, N., Klein, C., ... Grünwald, A. (2022). Parkin Deficiency Impairs Mitochondrial DNA Dynamics and Propagates Inflammation. *Movement Disorders*, 37(7), 1405–1415.
40. Yoshino, H., Li, Y., Nishioka, K., Daida, K., Hayashida, A., Ishiguro, Y., Yamada, D., Izawa, N., Nishi, K., Nishikawa, N., Oyama, G., Hatano, T., Nakamura, S., Yoritaka, A., Motoi, Y., Funayama, M., & Hattori, N. (2022). Genotype-phenotype correlation of Parkinson's disease with PRKN variants. *Neurobiology of Aging*, 114, 117–128.

41. Jiang, J.; Wang, C.; Qi, R.; Fu, H.; Ma, Q. scREAD: A single-cell RNA-Seq database for Alzheimer's disease. *iScience* 2020, 23, 101769.
42. Krokidis, M.G.; Vrahatis, A.G.; Lazaros, K.; Skolariki, K.; Exarchos, T.P.; Vlamos, P. Machine Learning Analysis of Alzheimer's Disease Single-Cell RNA-Sequencing Data across Cortex and Hippocampus Regions. *Curr. Issues Mol. Biol.* 2023, 45, 8652–8669.
43. Soreq, L.; Bird, H.; Mohamed, W.; Hardy, J. Single-cell RNA sequencing analysis of human Alzheimer's disease brain samples reveals neuronal and glial specific cells differential expression. *PLoS ONE* 2023, 18, e0277630.
44. Taylor, J.P.; Brown, R.H.; Cleveland, D.W. Decoding ALS: from genes to mechanism. *Nature* 2016, 539, 197-206.
45. Atlas, R.; Behar, L.; Elliott, E.; Ginzburg, I. The insulin-like growth factor mRNA binding-protein IMP-1 and the Ras-regulatory protein G3BP associate with tau mRNA and HuD protein in differentiated P19 neuronal cells. *Journal of neurochemistry* 2004, 89, 613-626.
46. Silva, J.M.; Rodrigues, S.; Sampaio-Marques, B.; Gomes, P.; Neves-Carvalho, A.; Dioli, C.; Soares-Cunha, C.; Mazuik, B.F.; Takashima, A.; Ludovico, P.; et al. Dysregulation of autophagy and stress granule-related proteins in stress-driven Tau pathology. *Cell death and differentiation* 2019, 26, 1411-1427.
47. Sato, K.; Takayama, K.I.; Inoue, S. Stress granules sequester Alzheimer's disease-associated gene transcripts and regulate disease-related neuronal proteostasis. *Aging* 2023, 15, 3984-4011.
48. Vanderweyde, T.; Yu, H.; Varnum, M.; Liu-Yesucevitz, L.; Citro, A.; Ikezu, T.; Duff, K.; Wolozin, B. Contrasting pathology of the stress granule proteins TIA-1 and G3BP in tauopathies. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2012, 32, 8270-8283.
49. Sidibé, H.; Dubinski, A.; Vande Velde, C. The multi-functional RNA-binding protein G3BP1 and its potential implication in neurodegenerative disease. *Journal of neurochemistry* 2021, 157.
50. Buchan, J.R.; Kolaitis, R.M.; Taylor, J.P.; Parker, R. Eukaryotic stress granules are cleared by autophagy and Cdc48/VCP function. *Cell* 2013, 153, 1461-1474.
51. Yang, P.; Mathieu, C.; Kolaitis, R.M.; Zhang, P.; Messing, J.; Yurtsever, U.; Yang, Z.; Wu, J.; Li, Y.; Pan, Q.; et al. G3BP1 Is a Tunable Switch that Triggers Phase Separation to Assemble Stress Granules. *Cell* 2020, 181, 325-345.e328.
52. W. M. van Oostveen and E. C. M. de Lange, "Imaging Techniques in Alzheimer's Disease: A Review of Applications in Early Diagnosis and Longitudinal Monitoring," *Int. J. Mol. Sci.*, vol. 22, no. 4, p. 2110, Jan. 2021.
53. Kale et al., "AI-driven innovations in Alzheimer's disease: Integrating early diagnosis, personalized treatment, and prognostic modelling," *Ageing Res. Rev.*, vol. 101, p. 102497, Nov. 2025/
54. M. Ai, Y. Liu, D. Liu, C. Yan, X. Wang, and X. Chen, "Research progress in predicting the conversion from mild cognitive impairment to Alzheimer's disease via multimodal MRI and artificial intelligence," *Front. Neurol.*, vol. 16, Jun. 2025.
55. C. S. Judge et al., "Multimodal Artificial Intelligence in Medicine," *Kidney360*, vol. 5, no. 11, pp. 1771–1779, Aug. 2024.
56. L. M. Winchester et al., "Artificial intelligence for biomarker discovery in Alzheimer's disease and dementia," *Alzheimers Dement. J. Alzheimers Assoc.*, vol. 19, no. 12, pp. 5860–5871, Dec. 2023.

57. H. Xu and K. M. J. Shuttleworth, “Medical artificial intelligence and the black box problem: a view based on the ethical principle of ‘do no harm,’” *Intell. Med.*, vol. 4, no. 1, pp. 52–57, Feb. 2024.
58. C. Rudin, “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead,” *Nat. Mach. Intell.*, vol. 1, no. 5, pp. 206–215, May 2019.
59. Sapkota, R., Roumeliotis, K.I. and Karkee, M., 2025. Ai agents vs. agentic ai: A conceptual taxonomy, applications and challenges. arXiv preprint arXiv:2505.10468. <https://doi.org/10.1016/j.inffus.2025.103599>
60. M. Lanham, *AI agents in action*, [First edition]. Shelter Island, NY: Manning Publications, 2025
61. Matsumoto, K., Matsumoto, M. and Abe, H., 2006. Goal-based action selection and utility-based action bias. *Neural networks*, 19(8), pp.1315-1320. doi: 10.1016/j.neunet.2006.05.036. Epub 2006 Aug 30.
62. Perry, N., Sun, C., Munro, M., Boulton, K.A. and Guastella, A.J., 2024. AI technology to support adaptive functioning in neurodevelopmental conditions in everyday environments: a systematic review. *NPJ Digital Medicine*, 7(1), p.370. doi: 10.1038/s41746-024-01355-7.
63. S. Shokrpour, A. MoghadamFarid, S. Bazzaz Abkenar, M. Haghi Kashani, M. Akbari, and M. Sarvzadeh, “Machine learning for Parkinson’s disease: a comprehensive review of datasets, algorithms, and challenges,” *Npj Park. Dis.*, vol. 11, no. 1, p. 187, Jul. 2025, doi: 10.1038/s41531-025-01025-9
64. J. Venugopalan, L. Tong, H. R. Hassanzadeh, and M. D. Wang, “Multimodal deep learning models for early detection of Alzheimer’s disease stage,” *Sci. Rep.*, vol. 11, no. 1, p. 3254, Feb. 2021, doi: 10.1038/s41598-020-74399-w.
65. M. I. Ahmed, B. Spooner, J. Isherwood, M. Lane, E. Orrock, and A. Dennison, “A Systematic Review of the Barriers to the Implementation of Artificial Intelligence in Healthcare,” *Cureus*, vol. 15, no. 10, p. e46454,
66. Saab, K., Freyberg, J., Park, C., Strother, T., Cheng, Y., Weng, W.H., Barrett, D.G., Stutz, D., Tomasev, N., Palepu, A. and Liévin, V., 2025. Advancing Conversational Diagnostic AI with Multimodal Reasoning. arXiv preprint arXiv:2505.04653.
67. Singhal, K., Azizi, S., Tu, T., Mahdavi, S.S., Wei, J., Chung, H.W., Scales, N., Tanwani, A., Cole-Lewis, H., Pfohl, S. and Payne, P., 2023. Large language models encode clinical knowledge. *Nature*, 620(7972), pp.172-180.
68. Singhal, K., Tu, T., Gottweis, J., Sayres, R., Wulczyn, E., Amin, M., Hou, L., Clark, K., Pfohl, S.R., Cole-Lewis, H. and Neal, D., 2025. Toward expert-level medical question answering with large language models. *Nature Medicine*, 31(3), pp.943-950.
69. Schouten, D., Nicoletti, G., Dille, B., Chia, C., Vendittelli, P., Schuurmans, M., Litjens, G. and Khalili, N., 2025. Navigating the landscape of multimodal AI in medicine: a scoping review on technical challenges and clinical applications. *Medical Image Analysis*, p.103621.

### Επιστημονικές Δημοσιεύσεις

1. Karanikolaos, P., Krokidis, M.G., Exarchos, T.P. and Vlamos, P., 2024. Computational Analysis of Marker Genes in Alzheimer's Disease across Multiple Brain Regions. *Information*, 15(9), p.523.
2. Perperidis, K., Exarchos, T.P., Vrahatis, A.G., Vlamos, P. and Krokidis, M.G., 2025. Computational analysis of transcriptome data and mapping of functional networks in Parkinson's disease. *Frontiers in Bioinformatics*, 5, p.1690229.
3. Grammenos, G., Vrahatis, A.G., Exarchos, T.P., T Vlamos, P. and Krokidis, M.G., 2025. AI agents in Alzheimer's Disease Management: Challenges and Future Directions. *Frontiers in Aging Neuroscience*, 17, p.1735892.

### Δημοσιεύσεις και αναφορές σε πρακτικά συνεδρίων (Conference papers and proceedings)

1. Papatheodorou M-C, Exarchos, T. P., Vlamos, P. Krokidis, M.G. 2025. Effect of Amyloid-beta Exposure on Phospholipid Dynamics: A Computational and Cellular Perspective, 3rd International Conference on Chemo and Bioinformatics, 25-26 September, 2025. Kragujevac, Serbia (pp. 331-334).
2. Perperidis, K., Exarchos, T.P., Vrahatis, A.G., Vlamos, P. and Krokidis, M.G., 2025. Computational analysis of transcriptome data and mapping of functional networks in Parkinson's disease, 75th Conference of the HSBMB, 5-7 December 2025, Athens, Greece, P118 – 276.
3. Giannopoulos, V., Exarchos, T. P., Vlamos, P. Krokidis M.G. 2025. Computational analysis of selected salicylidene imines for their potential as BACE-1 inhibitors, 75<sup>th</sup> Conference of the HSBMB, 5-7 December 2025, Athens, Greece, P105-275.

