

**Γεφυρώνοντας μεγάλα ομικά, γενετικά και ιατρικά
δεδομένα για την ευρεία εφαρμογή της Ιατρικής
Ακριβείας στην Ελλάδα**

ΠΑΡΑΔΟΤΕΟ Π8.2

«Τεχνική αναφορά με τις μεθοδολογίες βαθιάς μάθησης»

Φορέας	Πανεπιστήμιο Θεσσαλίας
Τύπος Παραδοτέου	Έκθεση
Ημερομηνία Υποβολής Παραδοτέου	31 Δεκεμβρίου 2025
Ενότητα Εργασίας	Ενότητα Εργασίας 8 «Βάση γενωμικών και βιοϊατρικών δεδομένων»

1	Εισαγωγή	4
2	Ορισμός Προβλήματος και Χαρακτηριστικά Ομικών Δεδομένων	5
3	Γενική Περιγραφή της Προτεινόμενης Μεθοδολογίας Βαθιάς Μάθησης	6
4	Μείωση Διασπασιμότητας για Ομικά Δεδομένα.....	7
	4.1 Τυχαίες Προβολές (Random Projections)	7
	4.2 Ανάλυση Κύριων Συνιστωσών ως Μηχανισμός Ευθυγράμμισης.....	8
	4.3 Συνδυαστική Προσέγγιση RP-PCA	9
5	Στρατηγική Εμπλουτισμού Δεδομένων μέσω Προβολών (Projection-Based Data Augmentation).....	10
	5.1 Εμπλουτισμός μέσω Πολλαπλών Τυχαίων Προβολών	10
	5.2 Ρόλος της PCA στη Δημιουργία Συνεπών Εμπλουτισμένων Αναπαραστάσεων	11
	5.3 Εκπαίδευση Νευρωνικού Δικτύου σε Εμπλουτισμένο Χώρο	11
	5.4 Συνοπτική Περιγραφή της Διαδικασίας Εμπλουτισμού.....	12
6	Αρχιτεκτονική Νευρωνικού Δικτύου και Διαδικασία Εκπαίδευσης.....	12
	6.1 Επιλογή Απλής και Ελεγχόμενης Αρχιτεκτονικής	13
	6.2 Διαδικασία Εκπαίδευσης σε Εμπλουτισμένο Σύνολο Δεδομένων	13
	6.3 Σταθερότητα και Γενικευσιμότητα.....	14
7	Εξαγωγή Προβλέψεων και Δημιουργία Γενετικού Προφίλ	14
	7.1 Πολλαπλές Προβλέψεις μέσω Εμπλουτισμένων Αναπαραστάσεων	14
	7.2 Μηχανισμός Πλειοψηφικής Ψήφου και Σταθεροποίηση Απόφασης.....	15
	7.3 Γενετικό Προφίλ και Πολυδιάστατη Ερμηνεία	15
8	Πειραματική Επικύρωση και Αξιολόγηση της Μεθοδολογίας.....	16
	8.1 Πειραματικό Πρωτόκολλο	16
	8.2 Συγκριτική Αξιολόγηση	17
	8.3 Σταθερότητα Προβλέψεων	18
	8.4 Δυνατότητα Γενίκευσης	18
	8.5 Σύνοψη Πειραματικών Ευρημάτων	19
9	Συζήτηση και Προοπτικές Αξιοποίησης	19
	9.1 Ερμηνεία των Αποτελεσμάτων	19
	9.2 Δημιουργία Γενετικών Προφίλ και Πολυδιάστατη Πληροφόρηση	20
	9.3 Περιορισμοί της Προσέγγισης.....	20
	9.4 Προοπτικές Μελλοντικής Αξιοποίησης	21
10	Συμπεράσματα.....	22
	Βιβλιογραφία	23

Πίνακας Εικόνων

Εικόνα 1: Οι προτεινόμενες τεχνικές προεπεξεργασίας και επαύξησης (augmentation) εφαρμόζονται στο σύνολο δεδομένων Campbell, με ρυθμό επαύξησης 5, αποδίδοντας σε κάθε δημιουργούμενο υποχώρο ένα διακριτό χρώμα. Παρουσιάζουμε 10 τυχαία επιλεγμένα σημεία δεδομένων από το επαυξημένο σύνολο δεδομένων Campbell (οι διαφορετικές χρωματικές αποχρώσεις αναπαριστούν διαφορετικούς υποχώρους).

1 Εισαγωγή

Η ραγδαία εξέλιξη των τεχνολογιών μοριακής βιολογίας και η ευρεία υιοθέτηση τεχνικών υψηλής απόδοσης, όπως η αλληλούχιση RNA σε επίπεδο μεμονωμένου κυττάρου (single-cell RNA sequencing – scRNA-seq), έχουν οδηγήσει στη δημιουργία συνόλων δεδομένων εξαιρετικά υψηλής διαστασιμότητας. Τα δεδομένα αυτά παρέχουν πλούσια πληροφορία για τη γονιδιακή έκφραση και τη λειτουργική κατάσταση των κυττάρων, δημιουργώντας νέες προοπτικές για την κατανόηση της παθογένειας και την ανάπτυξη εξατομικευμένων θεραπευτικών στρατηγικών στο πλαίσιο της ιατρικής ακριβείας.

Κεντρικός στόχος του παρόντος παραδοτέου (Π8.2) είναι η συστηματική και τεχνικά τεκμηριωμένη παρουσίαση των μεθοδολογιών βαθιάς μάθησης που αναπτύχθηκαν στο πλαίσιο του έργου, με σκοπό την αξιοποίηση δεδομένων γονιδιακής έκφρασης για την πρόβλεψη φαινοτύπων και τη δημιουργία σταθερών και αξιόπιστων γενετικών προφίλ. Τα προφίλ αυτά δεν περιορίζονται στην υποστήριξη μιας μεμονωμένης διάγνωσης, αλλά μπορούν να προσφέρουν πρόσθετη πληροφόρηση σχετικά με την πιθανή εμφάνιση άλλων ασθενειών ή φαινοτυπικών καταστάσεων.

Η εφαρμογή μοντέλων βαθιάς μάθησης σε ομικά δεδομένα συνοδεύεται από σημαντικές προκλήσεις, κυρίως λόγω του φαινομένου «λίγα δείγματα – πολλές μεταβλητές» (small n, large p). Παρότι τα σύγχρονα νευρωνικά δίκτυα έχουν επιδείξει εντυπωσιακή απόδοση σε άλλους τομείς, η άμεση εφαρμογή τους σε δεδομένα γονιδιακής έκφρασης συχνά οδηγεί σε υπερπροσαρμογή και ασταθείς προβλέψεις.

Το παρόν παραδοτέο παρουσιάζει μια ενοποιημένη μεθοδολογική προσέγγιση που αντιμετωπίζει τα παραπάνω ζητήματα μέσω συνδυασμού μείωσης διαστασιμότητας, ελεγχόμενου εμπλουτισμού δεδομένων και μηχανισμών συνάθροισης προβλέψεων. Η προσέγγιση αυτή έχει ήδη επικυρωθεί επιστημονικά μέσω δημοσιευμένης εργασίας [1], ωστόσο εδώ η έμφαση δίνεται στη σαφή τεχνική περιγραφή και στην ανάδειξη της συνεισφοράς της στο πλαίσιο του έργου.

2 Ορισμός Προβλήματος και Χαρακτηριστικά Ομικών Δεδομένων

Τα δεδομένα γονιδιακής έκφρασης, και ιδιαίτερα τα δεδομένα scRNA-seq, χαρακτηρίζονται από εξαιρετικά υψηλή διαστασιμότητα, καθώς κάθε δείγμα περιγράφεται από την έκφραση δεκάδων χιλιάδων γονιδίων. Ταυτόχρονα, ο αριθμός των διαθέσιμων δειγμάτων είναι συχνά περιορισμένος, είτε λόγω κόστους είτε λόγω πειραματικών και βιολογικών περιορισμών. Η αναντιστοιχία αυτή μεταξύ αριθμού μεταβλητών και αριθμού δειγμάτων αποτελεί βασική πρόκληση για την εφαρμογή μεθόδων μηχανικής μάθησης.

Ένα από τα κυριότερα προβλήματα που προκύπτουν είναι η αραιότητα των δεδομένων, καθώς μεγάλος αριθμός γονιδίων εμφανίζει μηδενική ή πολύ χαμηλή έκφραση στα περισσότερα κύτταρα. Η αραιότητα αυτή, σε συνδυασμό με τον πειραματικό θόρυβο, δυσχεραίνει τον εντοπισμό σταθερών και γενικεύσιμων προτύπων. Επιπλέον, η υψηλή διαστασιμότητα οδηγεί στο φαινόμενο της «κατάρρας της διαστασιμότητας», μειώνοντας τη διακριτική ικανότητα των αποστάσεων μεταξύ δειγμάτων.

Στο πλαίσιο της πρόβλεψης φαινοτύπων από δεδομένα γονιδιακής έκφρασης, τα παραπάνω προβλήματα μεταφράζονται σε ασταθείς προβλέψεις και περιορισμένη δυνατότητα γενίκευσης. Το γεγονός αυτό καθιστά αναγκαία την ανάπτυξη νέων μεθοδολογικών προσεγγίσεων, οι οποίες θα μειώνουν αποτελεσματικά τη διαστασιμότητα, θα περιορίζουν τον θόρυβο και θα ενισχύουν τη σταθερότητα της διαδικασίας μάθησης.

3 Γενική Περιγραφή της Προτεινόμενης Μεθοδολογίας Βαθιάς Μάθησης

Η μεθοδολογία που αναπτύχθηκε στο πλαίσιο του έργου βασίζεται στην παραδοχή ότι η αξιόπιστη αξιοποίηση δεδομένων γονιδιακής έκφρασης από μοντέλα βαθιάς μάθησης απαιτεί συνδυασμό μείωσης διαστασιμότητας και εμπλουτισμού των διαθέσιμων δεδομένων. Αντί της άμεσης εκπαίδευσης ενός μοντέλου στο αρχικό, υψηλής διάστασης σύνολο δεδομένων, προτείνεται μια πολυφασική διαδικασία προεπεξεργασίας, εκπαίδευσης και εξαγωγής προβλέψεων. Σε πρώτο στάδιο, τα δεδομένα μετασχηματίζονται μέσω πολλαπλών τυχαίων προβολών σε χώρους χαμηλότερης διάστασης, διατηρώντας κατά προσέγγιση τις μεταξύ τους γεωμετρικές σχέσεις. Ακολούθως, εφαρμόζεται ανάλυση κύριων συνιστωσών (PCA) σε κάθε προβολή, με στόχο την ευθυγράμμιση των παραγόμενων αναπαραστάσεων σε ένα κοινό σύστημα συντεταγμένων. Η διαδικασία αυτή οδηγεί στη δημιουργία πολλαπλών συνεπών αναπαραστάσεων για κάθε δείγμα, οι οποίες λειτουργούν ως ελεγχόμενες παραλλαγές του αρχικού γενετικού προφίλ. Οι αναπαραστάσεις αυτές χρησιμοποιούνται για την εκπαίδευση ενός νευρωνικού δικτύου, το οποίο μαθαίνει να συσχετίζει σταθερά μοτίβα γονιδιακής έκφρασης με τον αντίστοιχο φαινότυπο.

Κατά το στάδιο της πρόβλεψης, κάθε νέο δείγμα οδηγεί σε πολλαπλές εκτιμήσεις, οι οποίες συνδυάζονται μέσω μηχανισμού συνάθροισης για την παραγωγή μιας τελικής απόφασης αυξημένης αξιοπιστίας. Με τον τρόπο αυτό, η μεθοδολογία υποστηρίζει τη δημιουργία γενετικών προφίλ που δεν βασίζονται σε μία μοναδική πρόβλεψη, αλλά σε συλλογική αξιολόγηση πολλαπλών αναπαραστάσεων.

Στα επόμενα κεφάλαια παρουσιάζονται αναλυτικά τα επιμέρους στάδια της μεθοδολογίας, οι σχεδιαστικές επιλογές που υιοθετήθηκαν και τα πειραματικά αποτελέσματα που τεκμηριώνουν την αποτελεσματικότητά της.

4 Μείωση Διαστασιμότητας για Ομικά Δεδομένα

Η μείωση της διαστασιμότητας αποτελεί κρίσιμο στάδιο στην ανάλυση ομικών δεδομένων και ειδικότερα δεδομένων γονιδιακής έκφρασης, όπου ο αριθμός των μεταβλητών (γονίδια) είναι κατά τάξεις μεγέθους μεγαλύτερος από τον αριθμό των διαθέσιμων δειγμάτων. Στο πλαίσιο του παρόντος έργου, η μείωση διαστασιμότητας δεν αντιμετωπίζεται ως απλή τεχνική συμπίεσης, αλλά ως βασικό συστατικό μιας συνολικής στρατηγικής που στοχεύει στη σταθερή και αξιόπιστη εκπαίδευση μοντέλων βαθιάς μάθησης.

Οι κλασικές προσεγγίσεις μείωσης διαστασιμότητας, όταν εφαρμόζονται απευθείας σε δεδομένα γονιδιακής έκφρασης, συχνά αποτυγχάνουν να προσφέρουν ικανοποιητικά αποτελέσματα. Αυτό οφείλεται τόσο στη μεγάλη κλίμακα των δεδομένων όσο και στη δομική τους πολυπλοκότητα. Για τον λόγο αυτό, υιοθετήθηκε μια συνδυαστική προσέγγιση δύο σταδίων, η οποία αξιοποιεί τις τυχαίες προβολές (Random Projections) και την ανάλυση κύριων συνιστωσών (Principal Component Analysis – PCA).

4.1 Τυχαίες Προβολές (Random Projections)

Οι τυχαίες προβολές αποτελούν μια υπολογιστικά αποδοτική μέθοδο μείωσης διαστασιμότητας, βασισμένη στο θεώρημα Johnson–Lindenstrauss. Η βασική ιδέα είναι ότι ένα σύνολο σημείων σε έναν χώρο πολύ υψηλής διάστασης μπορεί να προβληθεί σε έναν χαμηλότερης διάστασης χώρο, έτσι ώστε οι μεταξύ τους αποστάσεις να διατηρούνται κατά προσέγγιση με υψηλή πιθανότητα.

Στην περίπτωση των ομικών δεδομένων, οι τυχαίες προβολές παρουσιάζουν σημαντικά πλεονεκτήματα. Πρώτον, δεν απαιτούν εκπαίδευση ή βελτιστοποίηση παραμέτρων, γεγονός που τις καθιστά ιδιαίτερα κατάλληλες για μεγάλης κλίμακας δεδομένα. Δεύτερον, επιτρέπουν την παραγωγή πολλαπλών εναλλακτικών

αναπαραστάσεων των ίδιων δεδομένων, μέσω της χρήσης διαφορετικών τυχαίων πινάκων προβολής.

Ωστόσο, ένα βασικό μειονέκτημα των τυχαίων προβολών είναι η αστάθεια που μπορεί να εισάγουν, καθώς κάθε προβολή αντιστοιχεί σε έναν διαφορετικό γραμμικό μετασχηματισμό του αρχικού χώρου. Οι αναπαραστάσεις που προκύπτουν από διαφορετικές προβολές δεν είναι άμεσα συγκρίσιμες μεταξύ τους, γεγονός που δυσχεραίνει τη χρήση τους σε ενιαία μοντέλα μάθησης. Το πρόβλημα αυτό είναι ιδιαίτερα έντονο όταν ο στόχος είναι η εκπαίδευση ενός νευρωνικού δικτύου, το οποίο προϋποθέτει ένα συνεπές και σταθερό σύνολο χαρακτηριστικών.

4.2 Ανάλυση Κύριων Συνιστωσών ως Μηχανισμός Ευθυγράμμισης

Για την αντιμετώπιση της αστάθειας που εισάγουν οι τυχαίες προβολές, στο δεύτερο στάδιο της προτεινόμενης μεθοδολογίας εφαρμόζεται ανάλυση κύριων συνιστωσών (PCA) σε κάθε προβολή ξεχωριστά. Η PCA χρησιμοποιείται όχι μόνο για περαιτέρω μείωση της διαστασιμότητας, αλλά κυρίως ως μηχανισμός ευθυγράμμισης των παραγόμενων αναπαραστάσεων. Συγκεκριμένα, μετά την εφαρμογή μιας τυχαίας προβολής, τα δεδομένα βρίσκονται σε έναν χαμηλότερης διάστασης χώρο, ο οποίος όμως είναι αυθαίρετα περιστραμμένος ως προς τον αρχικό χώρο. Η PCA επιβάλλει μια κανονικοποιημένη αναπαράσταση, όπου οι άξονες αντιστοιχούν στις κατευθύνσεις μέγιστης διακύμανσης των δεδομένων. Με τον τρόπο αυτό, διαφορετικές τυχαίες προβολές οδηγούνται σε συγκρίσιμα συστήματα συντεταγμένων. Η επιλογή της PCA έναντι μη γραμμικών τεχνικών μείωσης διαστασιμότητας, όπως οι t-SNE ή UMAP, είναι συνειδητή. Παρότι οι μη γραμμικές μέθοδοι είναι ιδιαίτερα δημοφιλείς για σκοπούς οπτικοποίησης, συχνά αλλοιώνουν τη συνολική γεωμετρία των δεδομένων και δεν προσφέρουν σταθερές αναπαραστάσεις κατάλληλες για εκπαίδευση μοντέλων πρόβλεψης. Αντίθετα, η PCA διατηρεί τη γραμμική δομή και επιτρέπει την ερμηνευσιμότητα των παραγόμενων

συνιστωσών, στοιχείο ιδιαίτερα σημαντικό στο πλαίσιο της βιοϊατρικής πληροφορικής.

4.3 Συνδυαστική Προσέγγιση RP–PCA

Ο συνδυασμός τυχαίων προβολών και PCA αποτελεί βασικό πυλώνα της προτεινόμενης μεθοδολογίας. Οι τυχαίες προβολές εξασφαλίζουν την υπολογιστική αποδοτικότητα και τη δυνατότητα παραγωγής πολλαπλών εναλλακτικών αναπαραστάσεων, ενώ η PCA λειτουργεί ως μηχανισμός σταθεροποίησης και ενοποίησης αυτών των αναπαραστάσεων.

Το αποτέλεσμα είναι ένα σύνολο χαμηλής διάστασης χαρακτηριστικών που διατηρεί τη βασική πληροφορία της γονιδιακής έκφρασης, μειώνει σημαντικά τον θόρυβο και επιτρέπει την αποτελεσματική εκπαίδευση μοντέλων βαθιάς μάθησης. Παράλληλα, η διαδικασία αυτή θέτει τις βάσεις για τον τεχνητό εμπλουτισμό των δεδομένων, καθώς κάθε αρχικό δείγμα αντιστοιχίζεται σε πολλαπλές, συνεπείς αναπαραστάσεις. Η έννοια αυτή αναλύεται διεξοδικά στο επόμενο κεφάλαιο, όπου παρουσιάζεται η στρατηγική εμπλουτισμού δεδομένων που αναπτύχθηκε στο πλαίσιο του έργου.

5 Στρατηγική Εμπλουτισμού Δεδομένων μέσω Προβολών (Projection-Based Data Augmentation)

Η στρατηγική εμπλουτισμού δεδομένων που αναπτύχθηκε στο πλαίσιο του παρόντος έργου αποτελεί το βασικό καινοτόμο στοιχείο της προτεινόμενης μεθοδολογίας. Σε αντίθεση με κλασικές τεχνικές data augmentation που εφαρμόζονται κυρίως σε δεδομένα εικόνας ή σήματος, η παρούσα προσέγγιση είναι ειδικά σχεδιασμένη για δεδομένα υψηλής διαστασιμότητας και έντονης αραιότητας, όπως τα δεδομένα γονιδιακής έκφρασης.

Ο στόχος του εμπλουτισμού είναι διττός: (α) η αύξηση του αποτελεσματικού αριθμού δειγμάτων για την εκπαίδευση μοντέλων βαθιάς μάθησης και (β) η ενίσχυση της σταθερότητας των μαθησιακών αναπαραστάσεων. Ο εμπλουτισμός επιτυγχάνεται χωρίς αυθαίρετη εισαγωγή θορύβου ή δημιουργία συνθετικών δειγμάτων, αλλά μέσω ελεγχόμενων γραμμικών μετασχηματισμών που διατηρούν τη δομή των δεδομένων.

5.1 Εμπλουτισμός μέσω Πολλαπλών Τυχαίων Προβολών

Η βασική ιδέα έγκειται στη χρήση πολλαπλών τυχαίων προβολών του ίδιου συνόλου δεδομένων. Κάθε προβολή παράγει μια εναλλακτική χαμηλής διάστασης αναπαράσταση των ίδιων βιολογικών δειγμάτων, διατηρώντας κατά προσέγγιση τις μεταξύ τους αποστάσεις. Με την επανάληψη της διαδικασίας για διαφορετικούς τυχαίους πίνακες προβολής, προκύπτει ένα σύνολο αναπαραστάσεων που μπορούν να θεωρηθούν ως διαφορετικές «όψεις» του ίδιου γενετικού προφίλ.

Σε αντίθεση με τεχνικές επαναδειγματοληψίας (π.χ. bootstrap ή bagging), εδώ η ποικιλία προκύπτει από τον μετασχηματισμό του χώρου χαρακτηριστικών. Κάθε αρχικό δείγμα αντιστοιχίζεται σε πολλαπλά διανύσματα χαρακτηριστικών με κοινή

ετικέτα φαινοτύπου, εμπλουτίζοντας το σύνολο εκπαίδευσης χωρίς αλλοίωση της σημασιολογίας των κλάσεων.

Η χρήση πολλαπλών προβολών λειτουργεί ως μορφή κανονικοποίησης, εκθέτοντας το νευρωνικό δίκτυο σε ελαφρώς διαφορετικές εκδοχές των ίδιων δεδομένων και ωθώντας το στη μάθηση πιο γενικών και ανθεκτικών προτύπων.

5.2 Ρόλος της PCA στη Δημιουργία Συνεπών Εμπλουτισμένων Αναπαραστάσεων

Για την ενοποίηση των πολλαπλών προβολών εφαρμόζεται PCA σε κάθε προβολή. Η PCA λειτουργεί ως μηχανισμός ευθυγράμμισης, διασφαλίζοντας ότι οι εμπλουτισμένες αναπαραστάσεις βρίσκονται σε συγκρίσιμα συστήματα συντεταγμένων. Έτσι, τα πολλαπλά διανύσματα που αντιστοιχούν στο ίδιο αρχικό δείγμα μπορούν να θεωρηθούν ως θορυβώδεις αλλά συνεπείς παραλλαγές ενός κοινού υποκείμενου γενετικού προφίλ.

Από βιολογική σκοπιά, η διαδικασία αυτή ερμηνεύεται ως έλεγχος της σταθερότητας του φαινοτύπου υπό μικρές διακυμάνσεις της γονιδιακής έκφρασης, ενισχύοντας τη δυνατότητα γενίκευσης του μοντέλου.

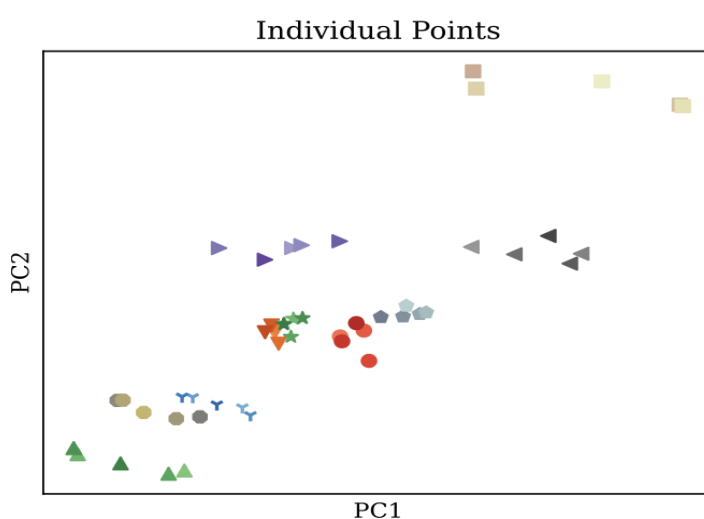
5.3 Εκπαίδευση Νευρωνικού Δικτύου σε Εμπλουτισμένο Χώρο

Μετά τον εμπλουτισμό, το αρχικό σύνολο δεδομένων μετασχηματίζεται σε διευρυμένο σύνολο εκπαίδευσης, όπου κάθε δείγμα εμφανίζεται πολλαπλές φορές με συνεπείς αναπαραστάσεις. Το σύνολο αυτό χρησιμοποιείται για την εκπαίδευση ενός νευρωνικού δικτύου πλήρους σύνδεσης.

Η εκπαίδευση σε εμπλουτισμένο χώρο μειώνει τον κίνδυνο υπερπροσαρμογής και οδηγεί στη μάθηση αναπαραστάσεων σταθερών ως προς μικρές μεταβολές των χαρακτηριστικών—ιδιότητα ιδιαίτερα κρίσιμη για δεδομένα γονιδιακής έκφρασης.

5.4 Συνοπτική Περιγραφή της Διαδικασίας Εμπλουτισμού

Συνοψίζοντας, η στρατηγική εμπλουτισμού δεδομένων περιλαμβάνει: (1) πολλαπλές τυχαίες προβολές, (2) ευθυγράμμιση μέσω PCA, (3) αντιστοίχιση πολλαπλών αναπαραστάσεων ανά δείγμα και (4) εκπαίδευση σε εμπλουτισμένο σύνολο. Η διαδικασία αυτή θέτει τις βάσεις για αξιόπιστη πρόβλεψη φαινοτύπων και δημιουργία γενετικών προφίλ, όπως περιγράφεται στα επόμενα κεφάλαια. Στην Εικόνα 1 εμφανίζεται μία οπτική αναπαράσταση της στρατηγικής εμπλουτισμού.



Εικόνα 1: Οι προτεινόμενες τεχνικές προεπεξεργασίας και επαύξησης (augmentation) εφαρμόζονται στο σύνολο δεδομένων Campbell, με ρυθμό επαύξησης 5, αποδίδοντας σε κάθε δημιουργούμενο υποχώρο ένα διακριτό χρώμα. Παρουσιάζουμε 10 τυχαία επιλεγμένα σημεία δεδομένων από το επαυξημένο σύνολο δεδομένων Campbell (οι διαφορετικές χρωματικές αποχρώσεις αναπαριστούν διαφορετικούς υποχώρους).

6 Αρχιτεκτονική Νευρωνικού Δικτύου και Διαδικασία Εκπαίδευσης

Η επιλογή της αρχιτεκτονικής του μοντέλου βαθιάς μάθησης έγινε με γνώμονα τη σταθερότητα, τη γενικευσιμότητα και τη συμβατότητα με τα ιδιαίτερα

χαρακτηριστικά των ομικών δεδομένων. Βασική επιδίωξη δεν ήταν η μεγιστοποίηση της πολυπλοκότητας του μοντέλου, αλλά η αξιόπιστη αξιοποίηση των εμπλουτισμένων αναπαραστάσεων που προκύπτουν από τα προηγούμενα στάδια.

Σε αντίθεση με εφαρμογές βαθιάς μάθησης σε εικόνες ή χρονικές ακολουθίες, τα δεδομένα γονιδιακής έκφρασης δεν διαθέτουν εγγενή χωρική ή χρονική δομή. Για τον λόγο αυτό, επιλέχθηκε ένα πλήρως συνδεδεμένο (fully connected) νευρωνικό δίκτυο, το οποίο λειτουργεί ως γενικός προσεγγιστής συναρτήσεων σε χαμηλής διάστασης χώρους χαρακτηριστικών.

6.1 Επιλογή Απλής και Ελεγχόμενης Αρχιτεκτονικής

Το προτεινόμενο νευρωνικό δίκτυο αποτελείται από επίπεδο εισόδου, δύο κρυφά επίπεδα και επίπεδο εξόδου. Το περιορισμένο βάθος της αρχιτεκτονικής είναι σκόπιμο και αποσκοπεί στον έλεγχο της υπερπροσαρμογής, η οποία αποτελεί συχνό φαινόμενο σε προβλήματα με περιορισμένο αριθμό δειγμάτων. Η χρήση μη γραμμικών συναρτήσεων ενεργοποίησης επιτρέπει τη μοντελοποίηση σύνθετων σχέσεων μεταξύ των χαρακτηριστικών, ενώ μηχανισμοί κανονικοποίησης, όπως η τυχαία απενεργοποίηση νευρώνων (dropout), ενισχύουν περαιτέρω τη γενικευσιμότητα του μοντέλου.

Η σκόπιμη απλότητα της αρχιτεκτονικής επιτρέπει την καθαρή ανάδειξη της συνεισφοράς της στρατηγικής εμπλουτισμού δεδομένων, απομονώνοντας την επίδραση της αναπαραστάσης από την πολυπλοκότητα του ίδιου του μοντέλου.

6.2 Διαδικασία Εκπαίδευσης σε Εμπλουτισμένο Σύνολο Δεδομένων

Η εκπαίδευση του νευρωνικού δικτύου πραγματοποιείται στο εμπλουτισμένο σύνολο δεδομένων, στο οποίο κάθε αρχικό δείγμα συνεισφέρει πολλαπλές συνεπείς αναπαραστάσεις με κοινή ετικέτα φαινοτύπου. Η διαδικασία αυτή αυξάνει το

αποτελεσματικό πλήθος των δειγμάτων και βελτιώνει τη στατιστική ισχύ της μάθησης.

Για τη βελτιστοποίηση του μοντέλου χρησιμοποιούνται καθιερωμένες συναρτήσεις κόστους για πολυκατηγορική πρόβλεψη, σε συνδυασμό με αλγορίθμους στοχαστικής καθόδου κλίσης. Το μοντέλο εκπαιδεύεται έτσι ώστε να μαθαίνει σχέσεις ανθεκτικές σε μικρές μεταβολές της γονιδιακής έκφρασης, περιορίζοντας την εξάρτηση από ιδιομορφίες των δεδομένων εκπαίδευσης.

6.3 Σταθερότητα και Γενικευσιμότητα

Η εκπαίδευση σε εμπλουτισμένο και ευθυγραμμισμένο χώρο οδηγεί σε προβλέψεις χαμηλής διακύμανσης και αυξημένης αξιοπιστίας. Η ιδιότητα αυτή είναι ιδιαίτερα κρίσιμη για βιοϊατρικές εφαρμογές, όπου η επαναληψιμότητα και η συνέπεια των αποτελεσμάτων αποτελούν βασικές προϋποθέσεις για περαιτέρω αξιοποίηση. Στο επόμενο κεφάλαιο περιγράφεται ο τρόπος με τον οποίο οι επιμέρους προβλέψεις συνδυάζονται για την εξαγωγή τελικών αποφάσεων αυξημένης αξιοπιστίας.

7 Εξαγωγή Προβλέψεων και Δημιουργία Γενετικού Προφίλ

Το στάδιο εξαγωγής προβλέψεων αποτελεί το τελικό και επιχειρησιακά κρίσιμο βήμα της προτεινόμενης μεθοδολογίας. Στο στάδιο αυτό, η πληροφορία που έχει ενσωματωθεί στο εκπαιδευμένο μοντέλο αξιοποιείται με τρόπο ώστε να παραχθούν σταθερές και ερμηνεύσιμες προβλέψεις, οι οποίες συνθέτουν ένα ολοκληρωμένο γενετικό προφίλ για κάθε δείγμα.\

7.1 Πολλαπλές Προβλέψεις μέσω Εμπλουτισμένων Αναπαραστάσεων

Κάθε νέο δείγμα γονιδιακής έκφρασης υποβάλλεται στην ίδια διαδικασία πολλαπλών προβολών και ευθυγράμμισης που εφαρμόστηκε στο σύνολο εκπαίδευσης. Με τον τρόπο αυτό προκύπτει ένα σύνολο εμπλουτισμένων και συνεπών αναπαραστάσεων του ίδιου βιολογικού δείγματος.

Το νευρωνικό δίκτυο παράγει προβλέψεις υπό τη μορφή πιθανοτήτων για κάθε μία από τις αναπαραστάσεις αυτές, παρέχοντας τη δυνατότητα αποτίμησης της συμφωνίας ή της απόκλισης μεταξύ των επιμέρους εκτιμήσεων.

7.2 Μηχανισμός Πλειοψηφικής Ψήφου και Σταθεροποίηση Απόφασης

Οι επιμέρους προβλέψεις συνδυάζονται μέσω μηχανισμού πλειοψηφικής ψήφου, ο οποίος λειτουργεί ως μηχανισμός συνάθροισης πληροφορίας. Η διαδικασία αυτή μειώνει την επίδραση μεμονωμένων ασταθών προβολών και οδηγεί σε τελικές αποφάσεις αυξημένης σταθερότητας και αξιοπιστίας.

7.3 Γενετικό Προφίλ και Πολυδιάστατη Ερμηνεία

Πέρα από την τελική κατηγορία φαινοτύπου, το γενετικό προφίλ περιλαμβάνει την κατανομή πιθανοτήτων μεταξύ των διαθέσιμων κατηγοριών. Η πολυδιάστατη αυτή πληροφορία επιτρέπει την εκτίμηση της αβεβαιότητας της πρόβλεψης και υποστηρίζει την αναγνώριση εναλλακτικών ή συνυπαρχόντων φαινοτυπικών σεναρίων, όπως η πιθανή εμφάνιση άλλων ασθενειών.

Το στάδιο αυτό ολοκληρώνει τη μεθοδολογική αλυσίδα του Π8.2 και αποτελεί τη βάση για την πειραματική επικύρωση και αξιολόγηση που παρουσιάζεται στο επόμενο κεφάλαιο.

8 Πειραματική Επικύρωση και Αξιολόγηση της Μεθοδολογίας

Η πειραματική επικύρωση της προτεινόμενης μεθοδολογίας αποτελεί κρίσιμο στάδιο για την τεκμηρίωση της αποτελεσματικότητάς της και την αξιολόγηση της πρακτικής της αξίας στο πλαίσιο της ανάλυσης ομικών δεδομένων. Στο παρόν κεφάλαιο παρουσιάζεται συνοπτικά η διαδικασία αξιολόγησης, με έμφαση στη συγκριτική απόδοση, τη σταθερότητα των προβλέψεων και τη δυνατότητα γενίκευσης της μεθοδολογίας σε διαφορετικά σύνολα δεδομένων.

Η πειραματική ανάλυση βασίστηκε σε δεδομένα γονιδιακής έκφρασης προερχόμενα από πολλαπλά σύνολα scRNA-seq, τα οποία παρουσιάζουν μεγάλη ποικιλομορφία ως προς τον αριθμό δειγμάτων, τη διαστασιμότητα και τον αριθμό των φαινοτυπικών κατηγοριών. Η επιλογή ετερογενών συνόλων δεδομένων διασφαλίζει ότι η αξιολόγηση δεν περιορίζεται σε ειδικές περιπτώσεις, αλλά αντανακλά ρεαλιστικά σενάρια εφαρμογής.

8.1 Πειραματικό Πρωτόκολλο

Για κάθε σύνολο δεδομένων εφαρμόστηκε διαχωρισμός σε σύνολο εκπαίδευσης και σύνολο ελέγχου, με το μεγαλύτερο μέρος των δεδομένων να χρησιμοποιείται για την εκπαίδευση του μοντέλου. Η διαδικασία εμπλουτισμού δεδομένων εφαρμόστηκε αποκλειστικά στο σύνολο εκπαίδευσης, ενώ το σύνολο ελέγχου χρησιμοποιήθηκε για την αξιολόγηση της γενικευσιμότητας.

Η αξιολόγηση επαναλήφθηκε πολλαπλές φορές με διαφορετικούς τυχαίους αρχικοποιήσεις, ώστε να εκτιμηθεί η σταθερότητα της μεθοδολογίας. Η απόδοση μετρήθηκε με καθιερωμένες μετρικές ταξινόμησης, όπως η ακρίβεια (accuracy) και ο δείκτης F1, οι οποίες είναι κατάλληλες για πολυκατηγορικά προβλήματα και ανισορροπημένες κλάσεις.

8.2 Συγκριτική Αξιολόγηση

Η προτεινόμενη μεθοδολογία συγκρίθηκε με μια σειρά από καθιερωμένες προσεγγίσεις μηχανικής μάθησης και εξειδικευμένες μεθόδους για δεδομένα scRNA-seq. Στις συγκρίσεις περιλαμβάνονται τόσο γενικά μοντέλα ταξινόμησης που χρησιμοποιούνται ευρέως σε δεδομένα πινάκων, όσο και μέθοδοι που ενσωματώνουν βιολογική γνώση ή ιεραρχικές δομές κυτταρικών τύπων.

Πίνακας 1: Σύνοψη των συγκρίσεων μεταξύ των μελετώμενων μεθόδων. Η καταχώριση $c_{i,j}$ υποδεικνύει τη συχνότητα με την οποία η μέθοδος της γραμμής i υπερτερεί της μεθόδου της στήλης j ως προς τη βαθμολογία F1 (F1-score). Ο αριθμός στην παρένθεση υποδηλώνει τον αριθμό των στατιστικά σημαντικών διαφορών.

	PANN-EC	RF-Majority	DNN	XGBoost	CHETAH	SingleR	scClassify
PANN-EC	-	10 (7)	12 (9)	9 (7)	15 (14)	13 (6)	10 (8)
RF-Majority	7 (5)	-	11 (7)	11 (4)	15 (15)	13 (7)	14 (7)
DNN	5 (4)	6 (1)	-	7 (2)	14 (14)	10 (6)	10 (5)
XGBoost	8 (6)	6 (3)	10 (8)	-	15 (15)	11 (9)	12 (7)
CHETAH	2 (1)	2 (2)	3 (3)	2 (0)	-	2 (2)	3 (3)
SingleR	3 (2)	4 (1)	7 (6)	6 (3)	15 (15)	—	8 (5)

Τα πειραματικά αποτελέσματα που παρουσιάζονται στον πίνακα 1 δείχνουν ότι η προτεινόμενη μεθοδολογία επιτυγχάνει συστηματικά υψηλότερη ή συγκρίσιμη απόδοση σε σχέση με τις εναλλακτικές προσεγγίσεις, ιδιαίτερα σε σύνολα

δεδομένων υψηλής διαστασιμότητας και περιορισμένου αριθμού δειγμάτων. Η βελτίωση αυτή αποδίδεται κυρίως στη συνδυαστική επίδραση της μείωσης διαστασιμότητας και του εμπλουτισμού δεδομένων, η οποία επιτρέπει στο νευρωνικό δίκτυο να μάθει πιο σταθερές και γενικεύσιμες αναπαραστάσεις.

8.3 Σταθερότητα Προβλέψεων

Ένα από τα σημαντικότερα ευρήματα της πειραματικής ανάλυσης αφορά τη σταθερότητα των προβλέψεων. Η χρήση πολλαπλών προβολών και ο μηχανισμός πλειοψηφικής ψήφου οδηγούν σε σημαντική μείωση της διακύμανσης της απόδοσης μεταξύ διαφορετικών εκτελέσεων.

Η παρατηρούμενη σταθερότητα είναι ιδιαίτερα κρίσιμη για εφαρμογές βιοϊατρικής πληροφορικής, όπου μικρές αλλαγές στα δεδομένα εισόδου ή στην αρχικοποίηση του μοντέλου μπορούν να οδηγήσουν σε μεγάλες αποκλίσεις στις προβλέψεις. Η προτεινόμενη μεθοδολογία αποδεικνύεται ανθεκτική σε τέτοιες διακυμάνσεις, ενισχύοντας την αξιοπιστία των παραγόμενων γενετικών προφίλ.

8.4 Δυνατότητα Γενίκευσης

Η αξιολόγηση σε πολλαπλά και ετερογενή σύνολα δεδομένων καταδεικνύει τη δυνατότητα γενίκευσης της προτεινόμενης προσέγγισης. Η μεθοδολογία δεν βασίζεται σε ειδικές ιδιότητες ενός συγκεκριμένου συνόλου δεδομένων, αλλά σε γενικές αρχές αναπαράστασης και μάθησης που μπορούν να εφαρμοστούν σε ευρύ φάσμα ομικών δεδομένων.

Το γεγονός αυτό καθιστά τη μεθοδολογία κατάλληλη για μελλοντική επέκταση σε άλλα είδη ομικών δεδομένων ή σε σενάρια όπου η πρόβλεψη φαινοτύπων απαιτεί τη συνδυαστική ανάλυση πολλαπλών πηγών πληροφορίας.

8.5 Σύνοψη Πειραματικών Ευρημάτων

Συνοψίζοντας, η πειραματική επικύρωση καταδεικνύει ότι η προτεινόμενη μεθοδολογία:

- Βελτιώνει συστηματικά την απόδοση μοντέλων βαθιάς μάθησης σε δεδομένα υψηλής διαστασιμότητας.
- Παρουσιάζει αυξημένη σταθερότητα προβλέψεων σε σύγκριση με κλασικές προσεγγίσεις.
- Γενικεύεται αποτελεσματικά σε διαφορετικά σύνολα ομικών δεδομένων.

Τα παραπάνω αποτελέσματα επιβεβαιώνουν ότι η μεθοδολογία που αναπτύχθηκε στο πλαίσιο του Π8.2 αποτελεί μια αξιόπιστη και πρακτικά εφαρμόσιμη λύση για την πρόβλεψη φαινοτύπων και τη δημιουργία γενετικών προφίλ από δεδομένα γονιδιακής έκφρασης. Στο επόμενο κεφάλαιο συζητούνται οι επιπτώσεις των ευρημάτων αυτών και οι προοπτικές μελλοντικής αξιοποίησης της προσέγγισης.

9 Συζήτηση και Προοπτικές Αξιοποίησης

Στο παρόν κεφάλαιο συνοψίζονται και ερμηνεύονται τα κύρια ευρήματα της προτεινόμενης μεθοδολογίας, ενώ παράλληλα συζητούνται οι δυνατότητες αξιοποίησής της στο πλαίσιο της βιοϊατρικής πληροφορικής και της ιατρικής ακριβείας. Η ανάλυση δεν περιορίζεται στην αποτίμηση της υπολογιστικής απόδοσης, αλλά επεκτείνεται στη λειτουργική σημασία των αποτελεσμάτων και στη συμβολή τους στη δημιουργία αξιόπιστων γενετικών προφίλ.

9.1 Ερμηνεία των Αποτελεσμάτων

Τα αποτελέσματα της πειραματικής αξιολόγησης καταδεικνύουν ότι ο συνδυασμός μείωσης διαστασιμότητας και εμπλουτισμού δεδομένων αποτελεί καθοριστικό παράγοντα για την επιτυχή εφαρμογή μοντέλων βαθιάς μάθησης σε δεδομένα

19

γονιδιακής έκφρασης. Η βελτίωση της απόδοσης δεν προκύπτει από την αύξηση της πολυπλοκότητας του μοντέλου, αλλά από την ποιοτική αναδιάρθρωση του χώρου αναπαράστασης των δεδομένων.

Η αυξημένη σταθερότητα των προβλέψεων, όπως αυτή αποτυπώνεται στη μειωμένη διακύμανση των μετρικών απόδοσης, αποτελεί ιδιαίτερα σημαντικό εύρημα. Στο πλαίσιο βιοϊατρικών εφαρμογών, η σταθερότητα αυτή είναι εξίσου σημαντική με την ακρίβεια, καθώς καθορίζει τον βαθμό εμπιστοσύνης που μπορεί να αποδοθεί στις προβλέψεις ενός υπολογιστικού μοντέλου.

9.2 Δημιουργία Γενετικών Προφίλ και Πολυδιάστατη Πληροφόρηση

Ένα από τα βασικά πλεονεκτήματα της προτεινόμενης προσέγγισης είναι η δυνατότητα μετάβασης από τη μονοδιάστατη πρόβλεψη φαινοτύπου στη δημιουργία πολυδιάστατων γενετικών προφίλ. Η αξιοποίηση της κατανομής πιθανοτήτων που προκύπτει από τον μηχανισμό συνάθροισης προβλέψεων επιτρέπει την αποτύπωση της αβεβαιότητας και της ετερογένειας που χαρακτηρίζει τα βιολογικά συστήματα.

Η πληροφορία αυτή μπορεί να αποτελέσει τη βάση για την εκτίμηση κινδύνου εμφάνισης επιπλέον ασθενειών ή φαινοτυπικών εκδηλώσεων, ιδιαίτερα σε περιπτώσεις όπου το γενετικό προφίλ δεν αντιστοιχεί με απόλυτη βεβαιότητα σε μία μοναδική κατηγορία. Με τον τρόπο αυτό, η μεθοδολογία υποστηρίζει σενάρια πολυνοσηρότητας και σύνθετης κλινικής εικόνας, τα οποία αποτελούν συχνό φαινόμενο στην πράξη.

9.3 Περιορισμοί της Προσέγγισης

Παρά τα θετικά αποτελέσματα, η προτεινόμενη μεθοδολογία παρουσιάζει και ορισμένους περιορισμούς. Ένας από τους βασικότερους αφορά το υπολογιστικό

κόστος που συνεπάγεται η εφαρμογή πολλαπλών τυχαίων προβολών και αναλύσεων κύριων συνιστωσών, ιδιαίτερα σε σύνολα δεδομένων μεγάλης κλίμακας. Αν και η διαδικασία είναι εγγενώς παραλληλοποιήσιμη, η αποδοτική υλοποίηση αποτελεί κρίσιμο παράγοντα για εφαρμογές σε πραγματικό χρόνο.

Επιπλέον, η μεθοδολογία, στη σημερινή της μορφή, επικεντρώνεται κυρίως σε δεδομένα γονιδιακής έκφρασης. Παρότι οι βασικές αρχές είναι γενικές, απαιτείται περαιτέρω μελέτη για την πλήρη αξιολόγηση της απόδοσης σε άλλα είδη ομικών δεδομένων ή σε συνδυαστικά σενάρια.

9.4 Προοπτικές Μελλοντικής Αξιοποίησης

Οι προοπτικές αξιοποίησης της προτεινόμενης μεθοδολογίας είναι ιδιαίτερα ευρείες. Ένα προφανές επόμενο βήμα είναι η επέκταση της προσέγγισης σε περιβάλλοντα πολυ-ομικής ανάλυσης, όπου δεδομένα γονιδιακής έκφρασης συνδυάζονται με άλλες πηγές πληροφορίας, όπως επιγενετικά δεδομένα ή κλινικές μεταβλητές.

Επιπλέον, η έννοια του εμπλουτισμού δεδομένων μέσω προβολών μπορεί να ενσωματωθεί σε πιο σύνθετα υπολογιστικά πλαίσια, όπως συστήματα υποστήριξης κλινικών αποφάσεων. Σε τέτοια σενάρια, το γενετικό προφίλ που παράγεται από τη μεθοδολογία μπορεί να λειτουργήσει ως εισροή για περαιτέρω ανάλυση ή ως εργαλείο προτεραιοποίησης κλινικών εξετάσεων.

Συνολικά, η προτεινόμενη προσέγγιση προσφέρει ένα ευέλικτο και επεκτάσιμο πλαίσιο, το οποίο μπορεί να προσαρμοστεί στις εξελισσόμενες ανάγκες της βιοϊατρικής έρευνας και της ιατρικής ακριβείας.

10 Συμπεράσματα

Στο παρόν παραδοτέο (Π8.2) παρουσιάστηκαν αναλυτικά οι μεθοδολογίες βαθιάς μάθησης που αναπτύχθηκαν στο πλαίσιο του έργου για την ενοποίηση και ανάλυση ομικών δεδομένων με σκοπό την πρόβλεψη φαινοτύπων και τη δημιουργία αξιόπιστων γενετικών προφίλ. Η προσέγγιση που υιοθετήθηκε αντιμετωπίζει με συστηματικό τρόπο τις βασικές προκλήσεις των δεδομένων γονιδιακής έκφρασης, όπως η υψηλή διαστασιμότητα, η αραιότητα και ο περιορισμένος αριθμός διαθέσιμων δειγμάτων.

Κεντρικό στοιχείο της προτεινόμενης μεθοδολογίας αποτελεί ο συνδυασμός τυχαίων προβολών και ανάλυσης κύριων συνιστωσών, ο οποίος επιτρέπει την ταυτόχρονη μείωση της διαστασιμότητας και τον ελεγχόμενο εμπλουτισμό των δεδομένων. Μέσω της δημιουργίας πολλαπλών, συνεπών αναπαραστάσεων κάθε δείγματος, καθίσταται δυνατή η εκπαίδευση μοντέλων βαθιάς μάθησης με αυξημένη σταθερότητα και γενικευσιμότητα. Η χρήση απλής και ελεγχόμενης αρχιτεκτονικής νευρωνικού δικτύου, σε συνδυασμό με τον μηχανισμό συνάθροισης προβλέψεων μέσω πλειοψηφικής ψήφου, οδηγεί στη διαμόρφωση προβλέψεων αυξημένης αξιοπιστίας. Η προσέγγιση αυτή επιτρέπει τη μετάβαση από την απλή ταξινόμηση φαινοτύπων στη δημιουργία πολυδιάστατων γενετικών προφίλ, τα οποία μπορούν να αξιοποιηθούν για την παροχή πρόσθετης πληροφόρησης σχετικά με την πιθανή εμφάνιση άλλων ασθενειών σε ασθενείς με δεδομένη διάγνωση.

Η πειραματική επικύρωση της μεθοδολογίας σε ετερογενή σύνολα δεδομένων γονιδιακής έκφρασης επιβεβαιώνει την αποτελεσματικότητά της και καταδεικνύει την υπεροχή της ως προς τη σταθερότητα των προβλέψεων και τη δυνατότητα γενίκευσης. Τα αποτελέσματα αυτά τεκμηριώνουν ότι η αναπτυχθείσα προσέγγιση αποτελεί μια ώριμη και πρακτικά εφαρμόσιμη λύση για προβλήματα πρόβλεψης φαινοτύπων σε περιβάλλοντα υψηλής διαστασιμότητας. Συνολικά, το Π8.2 συνεισφέρει ένα συνεκτικό και επεκτάσιμο μεθοδολογικό πλαίσιο, το οποίο μπορεί να αποτελέσει τη βάση για μελλοντικές επεκτάσεις σε πολυ-ομικά δεδομένα και εφαρμογές ιατρικής ακριβείας. Η εργασία που παρουσιάζεται στο παρόν παραδοτέο ενισχύει τη δυνατότητα αξιοποίησης τεχνικών βαθιάς μάθησης σε απαιτητικά

βιοϊατρικά προβλήματα και υποστηρίζει τη μετάβαση προς πιο ολοκληρωμένες και αξιόπιστες υπολογιστικές προσεγγίσεις στη σύγχρονη βιοϊατρική έρευνα.

Βιβλιογραφία

- [1] Panagiotis Anagnostou, Sotiris K. Tasoulis, Aristidis G. Vrahatis, Spiros V. Georgakopoulos, Vassilis P. Plagianakos, Boosting neural network performance for high dimensional data through random projections, Pattern Recognition Letters, Volume 199, 2026, Pages 149-155, ISSN 0167-8655, <https://doi.org/10.1016/j.patrec.2025.11.006>