

Θέμα: Εφαρμογές Συμβολικής Παλινδρόμησης στη Μεταγραφωμική Ανάλυση για την Πρόβλεψη Καρδιοτοξικότητας Αναστολέων Πρωτεϊνικών Κινασών

Επιβλέπων: Μιχάλης Σαβελώνας

Στοιχεία επικοινωνίας:

msavelonas@uth.gr

Σκοπός και στόχοι Σκοπός της εργασίας είναι η μελέτη και αξιοποίηση μεθόδων συμβολικής παλινδρόμησης (symbolic regression) για την ανάλυση μεταγραφωμικών δεδομένων ανθρώπινων καρδιακών κυττάρων με στόχο την πρόβλεψη καρδιοτοξικότητας που σχετίζεται με αναστολείς πρωτεϊνικών κινασών (protein kinase inhibitors - PKIs). Έμφαση θα δοθεί στην ανάπτυξη ερμηνεύσιμων μοντέλων που θα παράγουν τόσο προβλέψεις κινδύνου όσο και αναλυτικές μαθηματικές σχέσεις μεταξύ γονιδιακής έκφρασης, βιολογικών οδών και δεικτών καρδιοτοξικότητας. Παράλληλα, θα διερευνηθεί η ικανότητα των μεθόδων αυτών να αναδεικνύουν μη γραμμικές αλληλεπιδράσεις και συμπαγείς γονιδιακές υπογραφές.

Αντικείμενο Θα διερευνηθούν τεχνικές προεπεξεργασίας και επιλογής χαρακτηριστικών σε δεδομένα μεταγραφωμικής (π.χ. κανονικοποίηση, διαφορική έκφραση, ομαδοποίηση γονιδίων σε βιολογικές οδούς), καθώς και η εφαρμογή αλγορίθμων συμβολικής παλινδρόμησης (π.χ. γενετικός προγραμματισμός, PySR) για την εξαγωγή ερμηνεύσιμων μοντέλων πρόβλεψης. Επιπλέον, θα εξεταστεί η σύγκριση με καθιερωμένες μεθόδους και η πιθανή ενσωμάτωση συμπληρωματικών πληροφοριών, όπως χαρακτηριστικά χημικής δομής φαρμάκων. Η εργασία απαιτεί καλή γνώση προγραμματισμού (Python), βασικές γνώσεις μηχανικής μάθησης και ικανότητα κατανόησης επιστημονικής βιβλιογραφίας στην αγγλική.

Η εργασία περιλαμβάνει

- X Ανάλυση δεδομένων και ερμηνεία αποτελεσμάτων
- X Υλοποίηση ή έλεγχος και αξιολόγηση λογισμικού
- ☐ Κατασκευή πειραματικής διάταξης
- X Εκτέλεση συναφούς με το θέμα ερευνητικής δραστηριότητας

Σχετιζόμενα μαθήματα

Βιοπληροφορική I/II
Λειτουργική Γονιδιωματική
Αναγνώριση Προτύπων
Τεχνητή Νοημοσύνη
Ειδικά Θέματα Επεξεργασίας και Ανάλυσης Σημάτων και Εικόνων
Εξόρυξη και Ανάλυση Δεδομένων Μεγάλου Όγκου
Βαθιά Μάθηση
Ψηφιακή Επεξεργασία Βιοσημάτων

Προτεινόμενη μεθοδολογία έρευνας

Θα μελετηθεί η σχετική βιβλιογραφία με έμφαση στη συμβολική παλινδρόμηση.
Θα διερευνηθούν δημόσια διαθέσιμα σύνολα δεδομένων.
Θα διερευνηθούν δημόσια διαθέσιμες υλοποιήσεις.
Θα υλοποιηθούν μέθοδοι εναλλακτικές μέθοδοι παλινδρόμησης.
Θα διεξαχθεί συγκριτική μελέτη που θα περιλαμβάνει προϋπάρχουσες και προτεινόμενες μεθόδους.

Προσδοκώμενα αποτελέσματα

Ανάπτυξη ερμηνεύσιμης μεθοδολογίας πρόβλεψης καρδιοτοξικότητας βασισμένης σε συμβολική παλινδρόμηση, η οποία θα οδηγεί σε μαθηματικές εκφράσεις πρόβλεψης κινδύνου και σε ανάδειξη κρίσιμων γονιδιακών υπογραφών. Η προτεινόμενη προσέγγιση αναμένεται να επιτρέπει βιολογική

ερμηνεία των αποτελεσμάτων. Ιδανικά, τα αποτελέσματα της εργασίας θα υποβληθούν σε διεθνές επιστημονικό συνέδριο ή περιοδικό με κριτές.

Ενδεικτικές πηγές

1. van Hasselt, JG Coen, et al. "Transcriptomic profiling of human cardiac cells predicts protein kinase inhibitor-associated cardiotoxicity." *Nature communications* 11.1 (2020): 4809.
2. Cranmer, Miles. "Interpretable machine learning for science with PySR and SymbolicRegression. jl." *arXiv preprint arXiv:2305.01582* (2023).