



Πρόβλεψη αλληλεπίδρασης φαρμάκων με πρωτεΐνες χρησιμοποιώντας μεθόδους βαθιάς μάθησης

Υποψήφιος διδάκτορας Θωμάς Παπικινός,
επιβλέπωντας δρ. Θέμης Έξαρχος



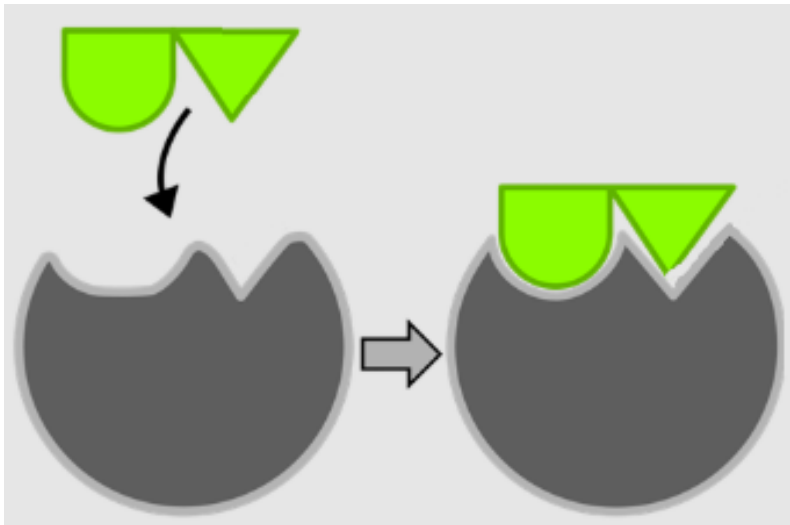
IONIAN
UNIVERSITY



Bioinformatics and Human Electrophysiology Laboratory, Department of
Informatics, Ionian University

Εισαγωγή

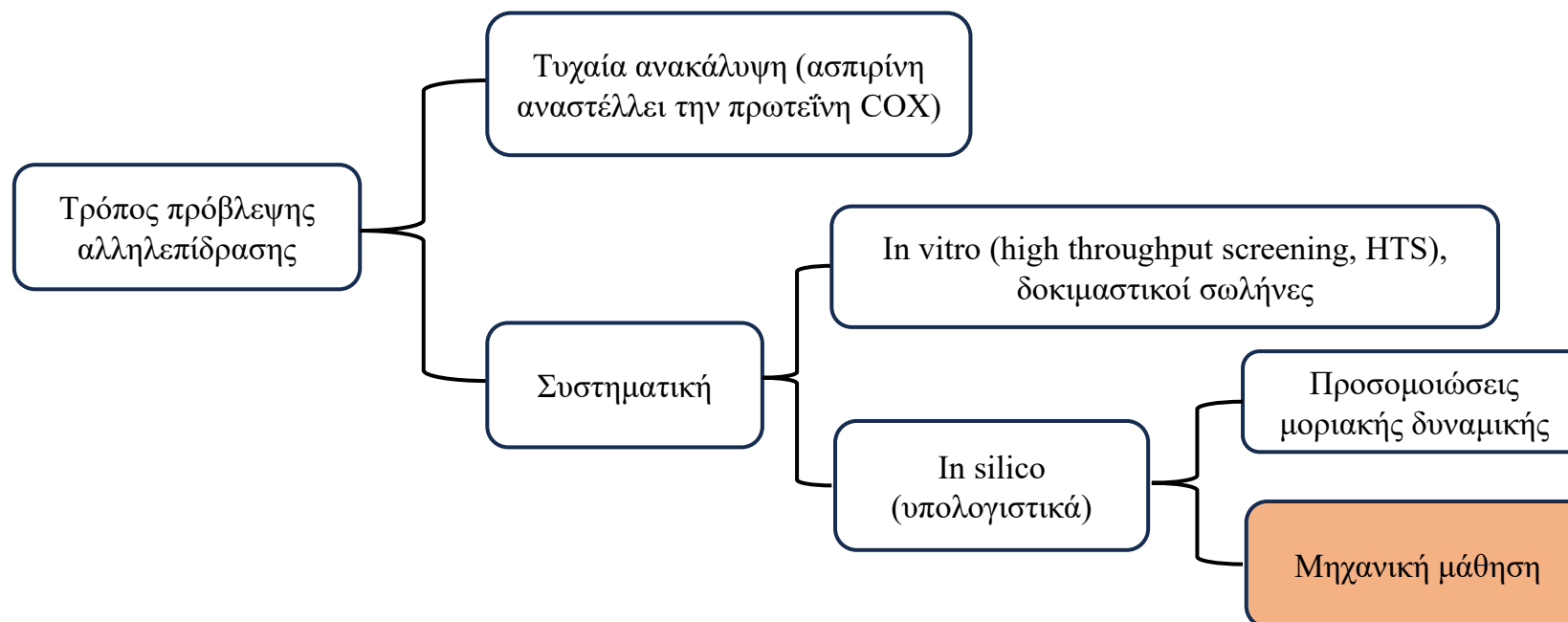
Αλληλεπίδραση φαρμάκων με πρωτεΐνες (drug-target interaction or DTI)



Σχήμα: Ένα μόριο που αλληλεπιδρά με μια πρωτεΐνη.

- Οι πρωτεΐνες επιτελούν βασικές βιολογικές λειτουργίες στον ανθρώπινο οργανισμό (π.χ. αιμοσφαιρίνη).
- Πολλές ασθένειες περιλαμβάνουν πρωτεΐνες που παίζουν καθοριστικό ρόλο στην εξέλιξή τους.
- Η εύρεση ουσιών που αλληλεπιδρούν με αυτές μπορεί να έχει θεραπευτικό χαρακτήρα.
- Μεταξύ των τύπων αλληλεπιδράσεων είναι η αναστολή και η ενεργοποίηση.
- Στον **καρκίνο του μαστού**, το trastuzumab αναστέλλει την HER2.
- Στην **Parkinson's**, το selegiline αναστέλλει τη μονοαμινοξειδάση B (MAO-B) μειώνοντας έτσι τη διάσπαση της ντοπαμίνης, αυξάνοντας έτσι τα επίπεδά της στον εγκέφαλο.
- Στην **σκλήρυνση κατά πλάκας (ΣΚΠ)**, το daclizumab αναστέλλει τη CD25, μειώνοντας τη νευροανοσολογική δραστηριότητα που προκαλείται από την ασθένεια.
- Συνεπώς η εύρεση DTIs αποτελεί πεδίο που χρίζει έρευνας.

Τρόποι εύρεσης αλληλεπιδράσεων

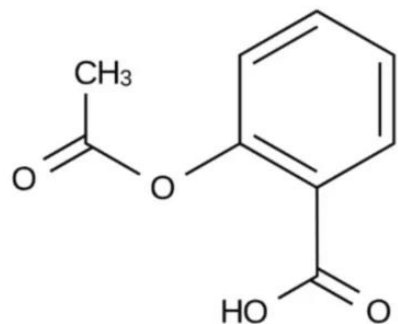


- **Τυχαίες ανακαλύψεις:** κάποια στιγμή τελειώνουν.
- **HTS κόστος (έλεγχος υψηλής απόδοσης):** της τάξης εκατομμυρίων.
- **In silico μέθοδοι:** επιτρέπουν εκμετάλλευση υπολογιστικών τεχνολογιών.
- **Προσομοιώσεις:** αργές και απαιτούν επιπλέον πληροφορίες (τοποθεσία αλληλεπίδρασης, 3D μορφολογία πρωτεΐνης κ.τ.λ.).
- **Μέθοδοι μηχανικής μάθησης:** πιο αποτελεσματικές και απαιτούν ευκολότερα προσβάσιμα δεδομένα.

Σκοπός της έρευνας:

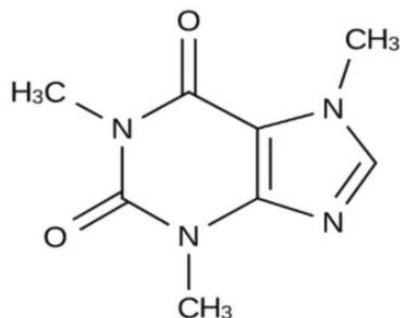
Δημιουργία μοντέλου μηχανικής μάθησης που προβλέπει DTIs (συγκεκριμένα αναστολή)

Μόρια: σε μορφή SMILES
(Simplified Molecular Input Line Entry System)



Acetylsalicylic Acid (Aspirin)

CC(=O)Oc1ccccc1C(=O)O



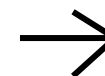
Epinephrine

CNC[C@H](O)c1ccc(O)c(O)c1

Σχήμα: Αναπαράσταση SMILES της ασπιρίνης και της επινεφρίνης.

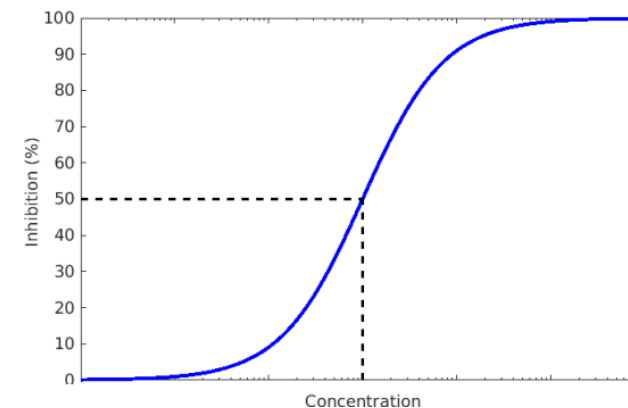
Πρωτεΐνες: αναπαράσταση ακολουθίας αμινοξέων

Amino acid	One letter
Alanine	A
Arginine	R
Asparagine	N
Aspartic acid	D
Cysteine	C
Glutamine	Q
Glutamic acid	E
Glycine	G
Histidine	H
Isoleucine	I
Leucine	L
Lysine	K
Methionine	M
Phenylalanine	F
Proline	P
Serine	S
Threonine	T
Tryptophan	W
Tyrosine	Y
Valine	V



Αποτέλεσμα: IC50

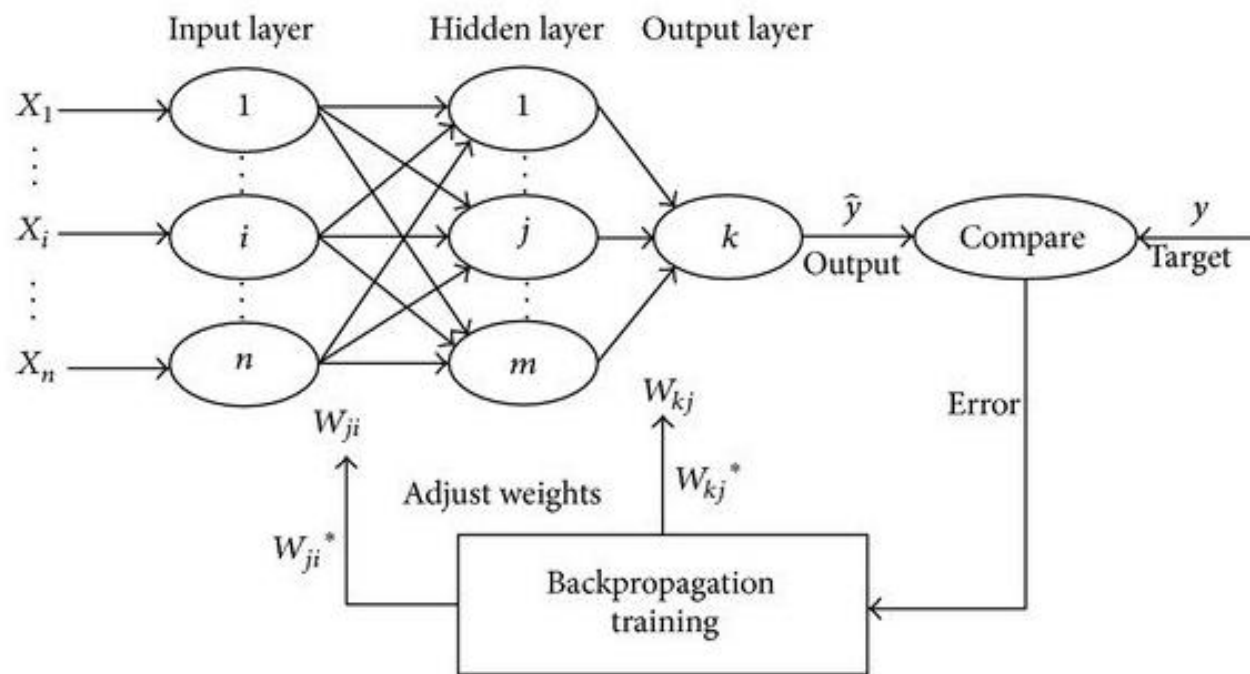
- **IC50:** απαιτούμενη ποσότητα για 50% αναστολή.
- **Μονάδα μέτρησης:** nM (nanomol).
- **Ισχυροί αναστολείς:** έχουν χαμηλό IC50.
- **Κατώφλι:** μπορεί να θεωρηθεί το 1000 nM.



Σχήμα: Η καμπύλη IC50.

Σχήμα: Τα γράμματα που συμβολίζουν τα 20 αμινοξέα.

Ανατομία ενός βασικού νευρωνικού δικτύου



Σχήμα: Ένα νευρωνικό δίκτυο με 1 επίπεδο.

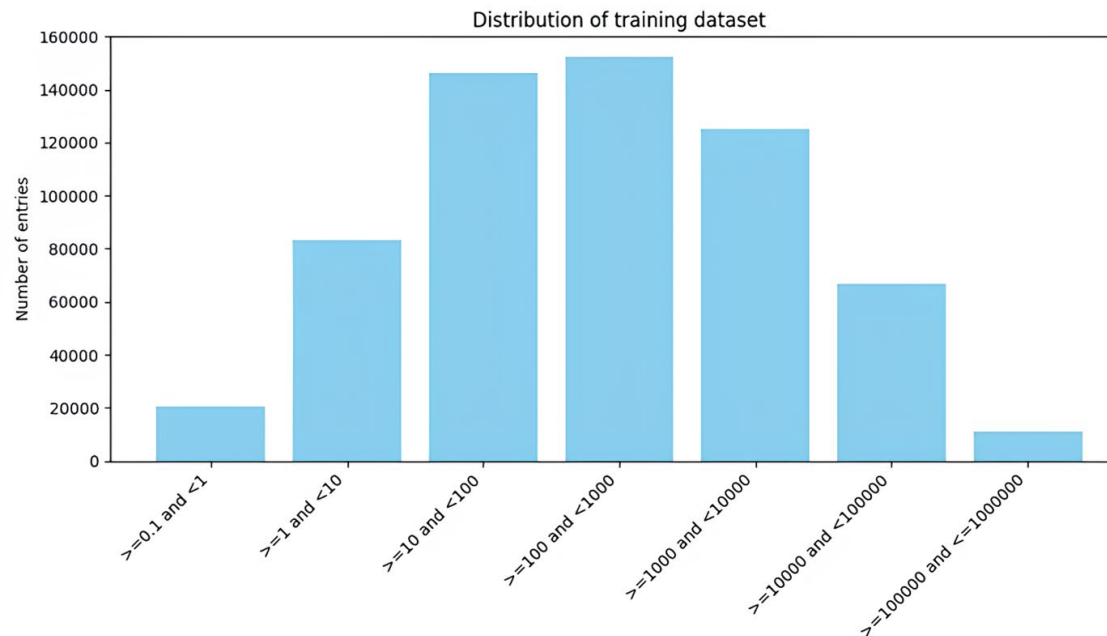
Πλεονοκρήματα της βαθιάς μάθησης

- Βαθιά μάθηση: νευρωνικά με πολλά επίπεδα.
- Αυτόματη εξαγωγή χαρακτηριστικών.
- Ικανότητα χρήσης «μεγάλων δεδομένων» (big data).
- Αξιοποίηση σύγχρονου υλικού (hardware).
- Μπορούν να διαχειριστούν πληθώρα τύπου δεδομένων.

Δεδομένα που χρησιμοποιήθηκαν

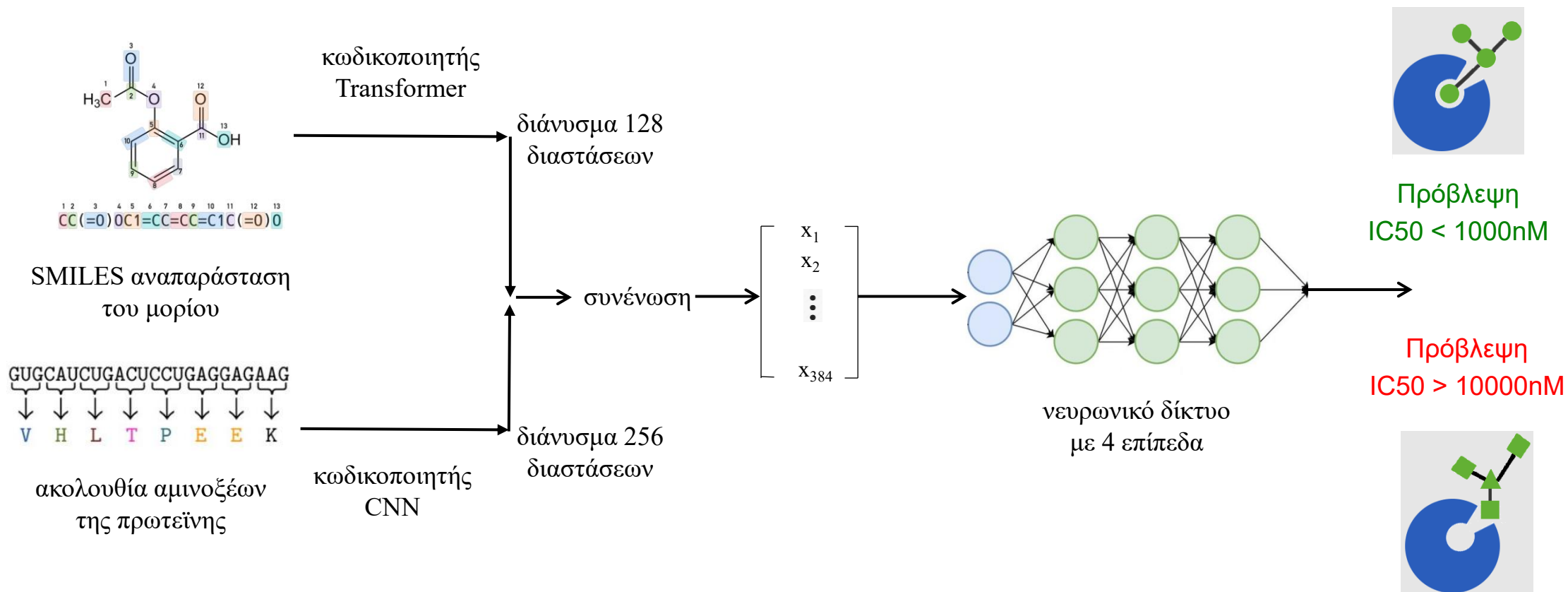


- **Δεδομένα εκπαίδευσης:** 605,180 ζευγάρια μορίων-πρωτεϊνών
- Κάθε ζευγάρι είχε ετικέτα IC50.
- 450,214 μοναδικά μόρια.
- 1,672 μοναδικές πρωτεΐνες.
- Μετατροπή σε log10 λόγω μεγάλου εύρους τάξης μεγέθους.

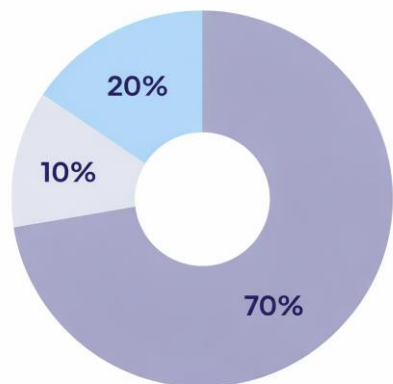


>=0.1 and <1	: 20472	(3.38%)
>=1 and <10	: 83046	(13.72%)
>=10 and <100	: 146237	(24.16%)
>=100 and <1000	: 152412	(25.18%)
>=1000 and <10000	: 125206	(20.69%)
>=10000 and <100000	: 66835	(11.04%)
>=100000 and <=1000000	: 10972	(1.81%)
Total number of entries: 605180		

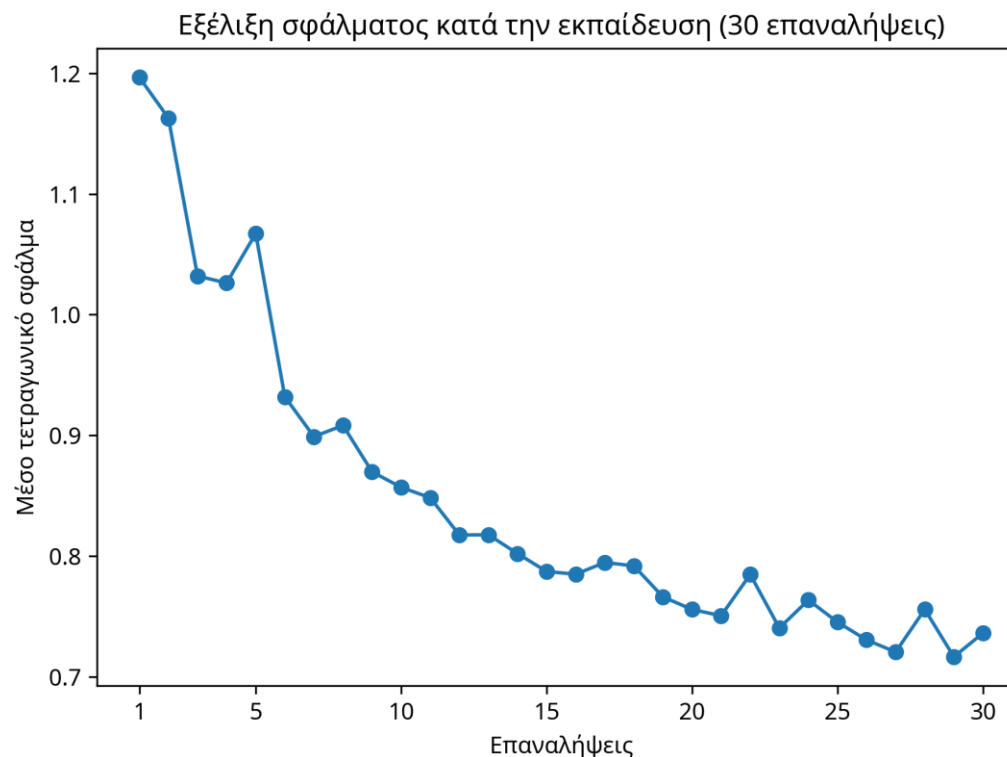
Αρχιτεκτονική του μοντέλου



Εκπαίδευση και αποτελέσματα



Σχήμα: Διαχωρισμός δεδομένων: 70-10-20. Εκπαίδευση-επιβεβαίωση-δοκιμή (training-validation-test).



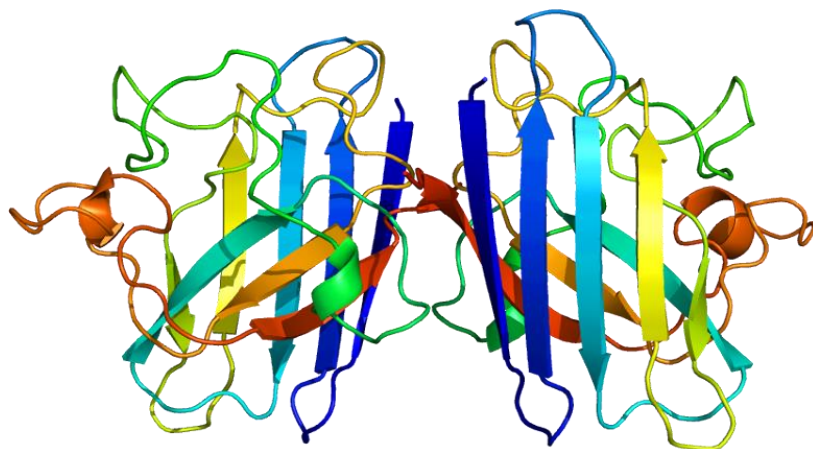
Σχήμα: Πρόοδος του ΜΤΣ κατά τη διάρκεια της εκπαίδευσης.

Μετρικές παλινδρόμησης	
Μέσο τετραγωνικό σφάλμα (ΜΤΣ)	0.71
Συντελεστής συσχέτισης Pearson	0.78
Concordance index	0.8
Μετρικές ταξινόμησης	
Accuracy	0.83
Precision	0.86
Recall	0.89
F1 Score	0.87

Πίνακας: Μετρικές του μοντέλου. Το πρόβλημα θεωρήθηκε επίσης και σαν περίπτωση ταξινόμησης (IC50<1000nM θετικό αλλιώς αρνητικό).

Συζήτηση

Μελέτη περίπτωσης: επαναστόχευση φαρμάκων για την αμυοτροφική πλάγια σκλήρυνση (ΑΠΣ).

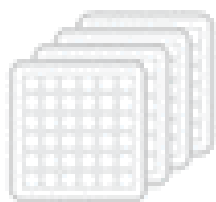


Εικόνα: Η πρωτεΐνη SOD1.

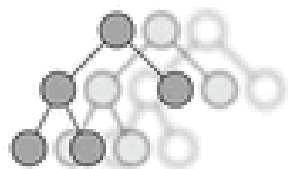
- **ΑΠΣ:** νευροεκφυλιστική ασθένεια που προσβάλλει τους κινητικούς νευρώνες και οδηγεί σε θάνατο.
- **Πρωτεΐνη SOD1:** κεντρικής σημασίας για την ΑΠΣ.
- **Εφαρμογή του μοντέλου:** στη βάση φαρμάκων Drug Repurposing Hub (3,556 φάρμακα).
- **Μεταξύ των αποτελεσμάτων** με τα μικρότερα προβλεπόμενα IC50, αρκετά από αυτά είναι ήδη βιβλιογραφικά συνδεδεμένα με την ΑΠΣ ή την SOD1, κάτι που επαληθεύει το μοντέλο.
- Ενδεικτικά:
- **Monepantel (κτηνιατρικό αντιπαρασιτικό)**, 3.9 nM προβλεπόμενο IC50: ήδη σε ανθρώπινες κλινικές δοκιμές για την ΑΠΣ.
- **Dasatinib (φάρμακο κατά της λευχαιμίας)**, 5.9 nM προβλεπόμενο IC50: βελτίωσε το προσδόκιμο ζωής σε μοντέλα ζώων της ασθένειας.
- **Fostamatinib (φάρμακο κατά της θρομβοπενίας)**, 11.4 nM προβλεπόμενο IC50: μείωσε τη νευροφλεγμονή σε μοντέλα ζώων της ασθένειας.
- **Το μοντέλο** μπορεί να εφαρμοστεί σε όποια ασθένεια έχουμε γνωστό βιολογικό στόχο (πρωτεΐνη).

Συζήτηση

Μελέτη περίπτωσης: προκαταρκτικός έλεγχος σε δοκιμές υψηλής απόδοσης αναζητώντας αναστολείς της πρωτεΐνης SOD1



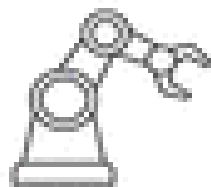
Αρχική βάση φαρμάκων Drug Repurposing Hub (3,556 ουσίες).



Εφαρμογή του μοντέλου αφαιρώντας ουσίες με προβλεπόμενο $IC_{50} > 1000$ nM.



Αριθμός ουσιών που απέμειναν: 889 (~75% μείωση).



Ουσίες μπορούν να χρησιμοποιηθούν για περαιτέρω εξέταση.

- **Precision** 0.86: δείχνει ένα επίπεδο ικανότητας ανίχνευσης πραγματικών αναστολέων.
- **Recall score** 0.89 δείχνει ότι δε χάνονται πολλοί πραγματικοί αναστολείς.
- **F1 score** 0.87 δείχνει καλή ισορροπία μεταξύ των παραπάνω.
- **Λόγω αυτών των μετρικών** το μοντέλο μπορεί να χρησιμοποιηθεί για προκαταρκτικό έλεγχο σε δοκιμές υψηλής απόδοσης.
- **Το κατώφλι** μπορεί να μεταβληθεί ανάλογα με τις απαιτήσεις μας.
- **Ο προκαταρκτικός έλεγχος** μπορεί να προκαλέσει μεγάλες μειώσεις στην HTS, η έρευνα 100,000 φαρμάκων κοστίζει 50-150k USD
- **Το μοντέλο** είναι διαθέσιμο για ελεύθερη χρήση από το κοινό.

Εικόνα: Περιγραφή του πως το μοντέλο μπορεί να χρησιμοποιηθεί στη διαδικασία ελέγχων υψηλής απόδοσης.