

ΘΕΜΑΤΙΚΗ ΕΝΟΤΗΤΑ: 3.ΕΦΑΡΜΟΣΜΕΝΗ ΦΥΣΙΚΗ
3.3 Πληροφορική – Επικοινωνία

ΥΠΟΛΟΓΙΣΤΙΚΗ ΜΕΘΟΔΟΛΟΓΙΑ ΕΝΤΟΠΙΣΜΟΥ ΠΡΩΤΕΪΝΩΝ ΚΑΙ
ΣΥΝΕΙΣΦΟΡΑ ΤΗΣ ΣΤΟΝ ΚΛΑΔΟ ΤΗΣ ΥΓΕΙΑΣ

Μυλωνά Ε., Σαβελώνας Μ., Μαρούλης Δ.
{emylona, msavel, dmaroulis}@di.uoa.gr
Τμήμα Πληροφορικής και Τηλεπικοινωνιών, ΕΚΠΑ, Αθήνα

Η πρωτεομική είναι η επιστήμη η οποία μελετά το σύνολο των πρωτεϊνών ενός ιστού, κυττάρου ή βιολογικού δείγματος. Σκοπός της είναι η σύγκριση δειγμάτων ιστών ανάμεσα σε υγιή και ασθενή άτομα έτσι ώστε να εντοπιστούν οι πρωτεΐνες οι οποίες έχουν υποστεί τροποποιήσεις στην έκφραση κατά την διάρκεια της νόσου. Η συστηματική ποσοτική ανάλυση των εν λόγω πρωτεϊνών οδηγεί στην ανάπτυξη εφαρμογών σε επίπεδο διάγνωσης καθώς και σε επίπεδο θεραπείας ασθενειών με την παραγωγή φαρμάκων. Η πρωτεομική χρησιμοποιεί προηγμένες τεχνικές για τον διαχωρισμό και την ανάλυση μειγμάτων πρωτεϊνών όπως είναι η δυσδιάστατη ηλεκτροφόρηση πηκτωμάτων (2D-GE). Τα αποτελέσματα που προκύπτουν από την τεχνική αυτή, εγγράφονται σε ψηφιακές εικόνες διαβαθμίσεων του γκρι οι οποίες εμπεριέχουν λευκές κηλίδες (protein spots) σε σκοτεινό υπόβαθρο και τα χρώματα του πηκτώματος μετατρέπονται σε εντάσεις εικονοστοιχείων. Η χειροκίνητη διαδικασία ανάλυσης των εικόνων από τους βιολόγους είναι αρκετά χρονοβόρα και υποκειμενική, οδηγώντας σε λανθασμένη εκτίμηση των αποτελεσμάτων. Το γεγονός αυτό δημιουργεί την ανάγκη αυτοματοποίησης αυτής της διαδικασίας με χρήση των εργαλείων της πληροφορικής. Η πληροφορική χρησιμοποιεί προηγμένες τεχνικές και αλγορίθμους που μειώνουν το ποσοστό παρέμβασης του βιολόγου και οδηγούν σε αντικειμενικά αποτελέσματα.

Παράγοντες οι οποίοι υπαισέρχονται και επηρεάζουν αρνητικά την ποιότητα των εικόνων πρωτεομικής είναι: α) το ανομοιογενές υπόβαθρο (inhomogeneous background), β) η δημιουργία συστάδων πρωτεϊνών οι οποίες περιέχουν επικαλυπτόμενες κηλίδες (overlapping spots), γ) το εύρος της συγκέντρωσης των πρωτεϊνών το οποίο οδηγεί σε κορεσμένες (saturated spots) και αχνές κηλίδες (faint spots), δ) οι οριζόντιες και κάθετες γραμμές (streaks) οι οποίες περιέχουν υπολείμματα πρωτεϊνών, ε) η παρουσία τεχνουργημάτων (artifacts) καθώς και, ζ) η παρουσία θορύβου υπό μορφή σκόνης που εισάγεται στο πηκτώμα κατά την διάρκεια του πειράματος. Η επιστήμη της πληροφορικής χρησιμοποιεί σύγχρονες μεθόδους επεξεργασίας δεδομένων και αντιμετωπίζει σε ικανοποιητικό βαθμό τα παραπάνω προβλήματα προσφέροντας στον βιολόγο-ερευνητή ανεκτίμητη βοήθεια στο έργο του. Τα υπάρχοντα εμπορικά λογισμικά πακέτα αντιμετωπίζουν αρκετά από τα προαναφερθέντα προβλήματα. Ωστόσο, τα αποτελέσματά τους είναι αμφιλεγόμενα λόγω του υψηλού ποσοστού παρέμβασης του χρήστη και απαιτείται η περαιτέρω διόρθωσή τους από τους βιολόγους μέσω μίας επίπονης και χρονοβόρας διαδικασίας.

Η παρούσα εργασία παρουσιάζει την ανάπτυξη μίας πρωτότυπης μεθοδολογίας εντοπισμού πρωτεϊνών βασισμένη στην μαθηματική μορφολογία. Η μεθοδολογία αποσκοπεί στην διάγνωση ασθενειών με την απουσία ή παρουσία πρωτεϊνών στο δείγμα κατά την διάρκεια της νόσου. Οι πρωτεΐνες αυτές μπορούν στη συνέχεια να χρησιμοποιηθούν ως στόχοι για την παραγωγή φαρμάκων. Σφάλματα που υπαισέρχονται στο στάδιο της ανίχνευσης περιλαμβάνουν: α) την συγχώνευση δύο πρωτεϊνικών κηλίδων σε μία, β) το διαχωρισμό μίας ξεχωριστής κηλίδας σε δύο, γ) την μη ανίχνευση πρωτεϊνικών κηλίδων όταν η ένταση της κηλίδας είναι χαμηλότερη από αυτή του υποβάθρου και δ) την λανθασμένη ανίχνευση αντικειμένων (artefacts) ως πρωτεϊνικές κηλίδες. Ιδιαίτερα, σε περιοχές που περιέχουν συστάδες, δηλαδή μεγάλη επικάλυψη πρωτεϊνών, η ανίχνευση όλων των κηλίδων είναι σχεδόν ανέφικτη. Η πρωτότυπη μεθοδολογία αντιμετωπίζει τα προαναφερθέντα προβλήματα συμβάλλοντας έτσι στην ιατρική σε επίπεδο διάγνωσης και επίπεδο θεραπείας ασθενειών.