

Εξατομικευμένη Ιατρική

- Εξατομικευμένη Ιατρική και μαζικά βιοϊατρικά δεδομένα
- Η γενετική σπουδαιότητα της ναυμαχίας της Σαλαμίνας
- Υπολογιστικά μοντέλα αξιολόγησης φαρμακοθεπειών





ZiTA CONGRESS ZiTA MEDICAL MANAGEMENT

ΠΕΡΙΟΔΙΚΟ
ΩΥΘ
ΔΙΑΤΡΟΦΗ ΥΓΕΙΑ ΟΜΟΡΦΙΑ

ΔΙΑΤΡΟΦΗ | ΥΓΕΙΑ | ΟΜΟΡΦΙΑ
ΩΥΘ FORUM

Στηρίζουμε κάθε σας δημιουργική σκέψη & προσπάθεια διάχυσης επιστημονικής γνώσης

- Διοργάνωση συνεδρίων, εκθέσεων, πολιτιστικών εκδηλώσεων και ταξιδίων κινήτρων
- Διαχείριση ιατρικών εταιρειών και οργανισμών
- Website και Ηλεκτρονικό Marketing
- Επιστημονικές Εκδόσεις Περιοδικών
- Χορηγίες
- Γραφιστικό - Δημιουργικό
- Γραμματειακή Υποστήριξη
- Τουρισμός Υγείας
- Νοσοκομειακό Marketing
- Γραφείο Τύπου



www.zita-group.com

Ομήρου 29, Πέτα Σαρωνικού, 190 01, Αττική, Ελλάδα

Τηλ: +30 22994 40962, Φαξ: +30 22990 66029

k.ge@zita-congress.gr, info@zita-congress.gr, info@zita-management.com

Follow us



ΣΥΝΤΑΚΤΙΚΗ ΕΠΙΤΡΟΠΗ**ΕΚΔΟΤΗΣ****Γεώργιος Π. Πατρινός***Καθηγητής, Πανεπιστήμιο Πατρών, Τμήμα Φαρμακευτικής, ΕΦΕΘ, Πάτρα***ΣΥΝΤΑΚΤΙΚΗ ΕΠΙΤΡΟΠΗ****Κλειώ Αναστασοπούλου***Αναπληρώτρια Καθηγήτρια, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών, Ιατρική Σχολή, Αθήνα***Γεράσιμος Βουτσινάς***Ερευνητής Α', Εθνικό Κέντρο Έρευνας Φυσικών Επιστημών «Δημόκριτος»
Ινστιτούτο Βιοεπιστημών και Εφαρμογών, Αθήνα***Θεodώρα Κατσιλά***Ερευνήτρια Γ', Εθνικό Ίδρυμα Ερευνών, Αθήνα***Ιωάννα Μαρουλάκου***Καθηγήτρια, Δημοκρίτειο Πανεπιστήμιο Θράκης, Τμήμα Μοριακής Βιολογίας και Γενετικής, Αλεξανδρούπολη***Σωτηρία Μπουκουβάλα***Αναπληρώτρια Καθηγήτρια, Δημοκρίτειο Πανεπιστήμιο Θράκης,
Τμήμα Μοριακής Βιολογίας και Γενετικής, Αλεξανδρούπολη***Αδαμαντία Παπαχατζοπούλου***Αναπληρώτρια Καθηγήτρια, Πανεπιστήμιο Πατρών, Τμήμα Ιατρικής, Πάτρα***Δέσποινα Σανούδου***Αναπληρώτρια Καθηγήτρια, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών, Ιατρική Σχολή, Αθήνα***Αργυρώ Σγουρού***Επίκουρη Καθηγήτρια, Ελληνικό Ανοικτό Πανεπιστήμιο, Τμήμα Βιολογίας, Πάτρα***Γεώργιος Σπυρούλιας***Καθηγητής, Πανεπιστήμιο Πατρών, Τμήμα Φαρμακευτικής, Πάτρα***Μαρία Σύρρου***Καθηγήτρια, Πανεπιστήμιο Ιωαννίνων, Τμήμα Ιατρικής, Ιωάννινα***Μαρία Τζέτη***Αναπληρώτρια Καθηγήτρια, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών, Ιατρική Σχολή, Αθήνα***Ασπασία Τσέζου***Καθηγήτρια, Πανεπιστήμιο Θεσσαλίας, Τμήμα Ιατρικής, Λάρισα***ΔΙΕΥΘΥΝΤΗΣ ΣΥΝΤΑΞΗΣ****Μανούσος Ε. Καμπούρης***Ερευνητής, Πανεπιστήμιο Πατρών, Τμήμα Φαρμακευτικής, Πάτρα*www.permed.gr

ΠΕΡΙΕΧΟΜΕΝΑ

- 85 **ΑΠΟ ΤΟΝ ΕΚΔΟΤΗ**
- 86 - 88 **ΕΠΙΣΤΗΜΟΝΙΚΑ ΝΕΑ**
- 89 - 102 **ΑΡΘΡΟ ΑΝΑΣΚΟΠΗΣΗΣ**
Εξατομικευμένη Ιατρική και μαζικά βιοϊατρικά δεδομένα
Μαργαρίτα-Ιωάννα Κουφάκη, Ιωάννης Γ. Χατζής, Γεώργιος Π. Πατρινός
- 103 - 111 **ΑΡΘΡΟ ΑΝΑΣΚΟΠΗΣΗΣ**
Υπολογιστική ανάπτυξη μοντέλων αξιολόγησης φαρμακοθεραπειών με χρήση αλγορίθμων μηχανικής εκμάθησης και ομικών δεδομένων
Ελένη Μπαρμπάνη, Μαρία Κορομηνά, Γεώργιος Π. Πατρινός
- 112 - 117 **ΠΡΩΤΟΤΥΠΟ ΑΡΘΡΟ**
Η γενετική σπουδαιότητα της ναυμαχίας της Σαλαμίνας
Κωνσταντίνος Τριανταφυλλίδης
- 118 - 122 **ΑΡΘΡΟ ΕΙΔΙΚΟΥ ΣΚΟΠΟΥ**
Η Ένωση Φοιτητών Φαρμακευτικής Ελλάδας (GPSF) συναντά τη Φαρμακογονιδιωματική και Εξατομικευμένη Ιατρική
Τατιάνα Γραμματικοπούλου, Δήμητρα Δεδούση, Μαρία Θεριανού, Ίρις Τριπολιτσιώτη
- 123 - 125 **ΟΔΗΓΙΕΣ ΠΡΟΣ ΣΥΓΓΡΑΦΕΙΣ**

Τα άρθρα που δημοσιεύονται στο περιοδικό αποτελούν πνευματική ιδιοκτησία του περιοδικού και δεν επιτρέπεται η ολική ή μερική αναδημοσίευσή τους χωρίς την έγγραφη άδεια του Εκδότη και του Διευθυντή Σύνταξης.

ΑΠΟ ΤΟΝ ΕΚΔΟΤΗ

ΜΕ ΤΟ συγκεκριμένο τεύχος του επιστημονικού περιοδικού «Εξατομικευμένη Ιατρική», κλείνει ένας χρόνος από την έκδοση του πρώτου τεύχους. Κρίνοντας από την ψηλή επισκεψιμότητα της ηλεκτρονικής έκδοσης του περιοδικού, το τιράζ της έντυπης έκδοσης του περιοδικού μέχρι τώρα, αλλά και την απήχηση που έχει η ύλη του περιοδικού μέχρι τώρα τόσο στην επιστημονική κοινότητα όσο και στους χορηγούς, τολμούμε να πούμε ότι το εγχείρημα αυτό ξεκινά με τις πλέον αισιόδοξες προβλέψεις και είναι απόλυτα εναρμονισμένο με τον αρχικό στόχο της διάδοσης της ιδέας και των αρχών, των μορφών και των μέσων της εξατομικευμένης ιατρικής, φαρμακογονιδιωματικής και εξατομικευμένης θεραπείας, και στην επαύξηση και βελτίωση του διεπιστημονικού διαλόγου.

Το 3ο τεύχος του 2020 είναι αφιερωμένο στη Γονιδιωματική πληροφορική, την κατεύθυνση που αποσκοπεί στην εφαρμογή υπολογιστικών μεθόδων και αλγορίθμων για τη συσχέτιση της γενετικής πληροφορίας και μεγάλου όγκου γενετικών δεδομένων με την κλινική εικόνα και τον εν γένει φαινότυπο. Συγκεκριμένα, το τεύχος αυτό περιλαμβάνει άρθρα ανασκόπησης σχετικά με τη χρήση υπολογιστικών εργαλείων και μηχανικής εκμάθησης για την επαναστόχευση φαρμάκων και σε συνδυασμό με ομικές τεχνολογίες, κάτι που αποτελεί την αιχμή του δόρατος στην ανάπτυξη φαρμάκων τις μέρες μας. Παράλληλα, το τεύχος αυτό φιλοξενεί άρθρο ανασκόπησης ως προς τη χρήση μαζικών βιοϊατρικών δεδομένων και τη χρήση τους στην εξατομικευμένη ιατρική, συμπεριλαμβανοντας ζητήματα όπως ο ηλεκτρονικός ιατρικός φάκελος, διαχείριση και ενσωμάτωση δεδομένων, ειδικά των γενετικών δεδομένων και άλλα.

Το τεύχος αυτό φιλοξενεί επίσης την παρουσίαση της Ένωσης Φοιτητών Φαρμακευτικής Ελλάδας (GPSF), παραθέτοντας τα αντικείμενα ενασχόλησης της Ένωσης αυτής με την Εξατομικευμένη Ιατρική, αναδεικνύοντας

το ενεργό ενδιαφέρον της φοιτητών της Φαρμακευτικής στη χώρα μας με το αντικείμενο αυτό.

Επιπρόσθετα, και με δεδομένο ότι το 2020 ήταν μια χρονιά-ορόσημο για την πατρίδα μας, σηματοδοτώντας 2500 χρόνια από την κοσμοϊστορική νίκη των Ελλήνων επί των Περσών στη Ναυμαχία της Σαλαμίνας, θεωρήσαμε αυτό ως μια εξαιρετική ευκαιρία για να φιλοξενήσουμε ένα άρθρο που να αναφέρεται στη γενετική διάσταση της ελληνικής επικράτησης κατά τη Ναυμαχία της Σαλαμίνας στο τεύχος αυτό, αφού η εξελισσόμενη πραγματικότητα της CoViD-19 απέτρεψε την δημοσίευσή του στο τεύχος που συνέπιπτε με την επέτειο (Σεπτέμβριος 2020). Συγκεκριμένα, στην εργασία αυτή διερευνήθηκε εάν υπήρξε γενετική ροή από το Ιράν στον Ευρωπαϊκό/Ελλαδικό χώρο με τη συγκέντρωση επιστημονικών στοιχείων για τη γενετική σύσταση των Ελλήνων και τη σύγκριση αυτών με αντίστοιχα δεδομένα από τον Ιράν, την Εγγύς Ανατολή, αλλά και όλη την Ευρώπη. Από αυτό προκύπτει ότι δεν υπάρχουν ενδείξεις ότι στην προαναφερόμενη χρονική περίοδο συνέβησαν μεγάλα μεταναστευτικά γεγονότα και ροή πληθυσμιακών ομάδων, άρα και γονιδίων, από το Ιράν προς τον Ελλαδικό γεωγραφικό χώρο, επιβεβαιώνοντας ότι η Ναυμαχία της Σαλαμίνας υπήρξε καθοριστική για τη γενετική σύσταση των κατοίκων των Βαλκανίων και της Ευρώπης γενικότερα. Ένα ενδιαφέρον πρωτότυπο άρθρο από πολλές απόψεις.

Ευχαριστώ θερμά για μια ακόμη φορά τον Δρ. Μανούσο Καμπούρη, διευθυντή σύνταξης για την αμέριστη βοήθειά του στην επιμέλεια των άρθρων που περιλαμβάνονται στο τεύχος αυτό και, ασφαλώς, τους συγγραφείς των άρθρων που φιλοξενούνται στο τεύχος αυτό. Εύχομαι σε όλους τους αναγνώστες μια όμορφη και δημιουργική χρονιά, μια ακόμη χρονιά ορόσημο για την πατρίδα μας, σηματοδοτώντας τα 200 χρόνια από την έναρξη του απελευθερωτικού μας αγώνα.

Γεώργιος Παν. Πατρινός

Καθηγητής

Πανεπιστήμιο Πατρών, Τμήμα Φαρμακευτικής

Η απόκρυψη των γενετικών αλλαγών προστατεύει το γενετικό απόρρητο επιτρέποντας την κοινή χρήση δεδομένων

Ερευνητές ανέπτυξαν έναν τρόπο προστασίας των ευαίσθητων προσωπικών γενετικών δεδομένων, επιτρέποντας παράλληλα την ελεύθερη πρόσβαση στα δεδομένα που θα ωφελούσαν στην ολοκλήρωση λειτουργικών μελετών του γονιδιώματος.

Για την κατανόηση της γονιδιακής έκφρασης και των ποικίλων φαινοτύπων, οι επιστήμονες στηρίζουν μέρος της μελέτης τους στην χρήση κοινόχρηστων και ελευθέρως προσβάσιμων γενετικών δεδομένων. Ωστόσο, η κοινή χρήση των γενετικών πληροφοριών εγείρει ανησυχίες σχετικά με το απόρρητο των προσωπικών δεδομένων.

Επιπλέον, λαμβάνοντας υπόψη το γεγονός ότι πολλές μελέτες του γονιδιώματος επικεντρώνονται συχνά στην διερεύνηση μίας υπό μελέτη ασθένειας κάθε φορά, υπάρχει ο κίνδυνος διαρροής προσωπικών πληροφοριών που σχετίζονται με την κατάσταση της υγείας του εκάστοτε ατόμου. Ερευνητές από το Πανεπιστήμιο Yale σημείωσαν, ωστόσο, ότι οι λειτουργικές γονιδιωματικές μελέτες δεν απαιτούν τις μεταλλαγές, υποδεικνύοντας ότι οι μεταλλαγές θα μπορούσαν να «αποκρυβούν» για να αποφευχθεί η διαρροή προσωπικών γενετικών δεδομένων.

Η ομάδα του Yale παρουσίασε μια μορφή αρχείου που επιτρέπει την ανάγνωση ακατέργαστων δεδομένων του γονιδιώματος ενώ ταυτόχρονα ελαχιστοποιούνται τυχόν διαρροές προσωπικών απόρρητων πληροφοριών, όπως ανέφεραν στο περιοδικό Cell.

Ο Gerstein και οι συνάδελφοί του στηρίχθηκαν σε ένα σενάριο κατά το οποίο κάποιος συλλέγει κρυφά DNA ενός ατόμου από ένα φλιτζάνι του καφέ, και αντιπαραβάλλει το συγκεκριμένο γονιδίωμα με πληροφορίες από μια βάση δεδομένων. Ένα τυχαίο δείγμα DNA, όταν συγκρίνεται με δεδομένα μελέτης ατόμων με διπολική διαταραχή, θα μπορούσε να δώσει για συσχέτιση του συγκεκριμένου γονιδιώματος και εμφάνιση της νόσου.

Οι ερευνητές ακολούθησαν αυτό το σενάριο - χρησιμοποιώντας δείγματα DNA που συλλέχθηκαν από τα φλιτζάνια ροφήματος από άτομα κατόπιν βέβαια της συγκατάθεσής τους - διαπιστώνοντας ότι μπορούσαν να εξαγουν συμπεράσματα για την κατάσταση της υγείας τους.

Για να αποφευχθούν τέτοιες διαρροές ευαίσθητων προσωπικών δεδομένων, οι ερευνητές επινόησαν μια διαδικασία ανάρτησης των δεδομένων που αποκρύπτει συγκεκριμένες γονιδιακές παραλλαγές από το σύνολο των γενετικών πληροφοριών - αντικαθιστώντας τις συγκεκριμένες πληροφορίες με δεδομένα από το ανθρώπινο γονιδίωμα αναφοράς.

«Μπορούμε να προστατεύσουμε το προσωπικό απόρρητο ενώ ταυτόχρονα ενθαρρύνουμε τους ανθρώπους να συμμετάσχουν σε γενετικές μελέτες προς όφελος της κοινωνίας», δήλωσε ο πρώτος συγγραφέας Mark Gerstein. ●

Η συνεργασία Ardigen και BioCollective στην μεταγονιδιωματική διερεύνηση βιοδεικτών της νόσου του Πάρκινσον

Οι εταιρείες Ardigen και BioCollective ανακοίνωσαν την συνεργασία τους στο πλαίσιο μεταγονιδιωματικής διερεύνησης βιοδεικτών για την εξέλιξη της νόσου του Πάρκινσον.

Οι εταιρείες θα χρησιμοποιήσουν την πλατφόρμα ερμηνείας του μικροβιώματος της Ardigen, που βασίζεται σε τεχνητή νοημοσύνη, για να αναλύσουν μεταγονιδιωματικά και κλινικά δεδομένα που λαμβάνονται μέσω δείγματος κοπράνων ασθενών με νόσο του Parkinson από τη BioCollective. Η προσπάθεια θα έχει στόχο τη διερεύνηση βιοδεικτών και τη σημασία αυτών στη διάγνωση και την εξέλιξη πρώιμου σταδίου της νόσου του Πάρκινσον.

«Η νόσος του Πάρκινσον είναι μια εξαιρετικά περίπλοκη ασθένεια και η μελέτη του μικροβιώματος συμβάλλει σημαντικά στην κατανόηση της νόσου», δήλωσε σε ανακοίνωσή της η CEO της BioCollective, Martha Carlin. «Πιστεύουμε ότι η συνεργασία θα βοηθήσει στην κατανόηση του ρόλου του μικροβιώματος στην υγεία του ανθρώπου».

Τον Φεβρουάριο, η BioCollective που εδρεύει στο Denver έλαβε ενίσχυση ύψους 1,2 εκατομμυρίων δολαρίων από τα Εθνικά Ινστιτούτα Υγείας των ΗΠΑ για να αναπτύξει ένα εθνικό πρότυπο αναφοράς μικροβιώματος. ●

Το πανεπιστήμιο του Manchester και η Sano Genetics στην έρευνα των γενετικών παθήσεων των οφθαλμών

Η Sano Genetics και οι ερευνητές στο Πανεπιστήμιο του Manchester στο Ηνωμένο Βασίλειο δήλωσαν ότι θα συνεργαστούν με στόχο την κλινική εφαρμογή της έρευνας για τις γενετικές παθήσεις των οφθαλμών.

Στόχος είναι η ανάπτυξη μιας βάσης δεδομένων που θα προσδιορίζει τον τρόπο με τον οποίο το γενετικό υπόβαθρο μπορεί να επηρεάσει την έκβαση της νόσου και η προσπάθεια θα βασιστεί εν μέρει στην διαδικτυακή πλατφόρμα της Sano Genetics και σε kit απομόνωσης DNA (η Sano θα παρέχει στους χρήστες kit απομόνωσης DNA από τον σίελο). Η Innovate UK, μια κυβερνητική υπηρεσία, υποστηρίζει την προσπάθεια με 80.000 £, ενώ ο συνολικός προϋπολογισμός για το έργο υπολογίζεται περίπου 153.000 £. Η διάρκεια του εγχειρήματος υπολογίζεται στα δύο χρόνια.

Ο Patrick Short, διευθύνων σύμβουλος της Sano Genetics, δήλωσε ότι η συνεργασία στοχεύει στην εξαγωγή αποτελεσμάτων που θα μπορούν να χρησιμοποιηθούν άμεσα στην πρόβλεψη του κινδύνου ανάπτυξης γενετικών παθήσεων των οφθαλμών και στη θεραπεία αυτών.

Ο Jamie Ellingford, ερευνητής στο Πανεπιστήμιο του Μάντσεστερ και επικεφαλής ερευνητής του εγχειρήματος, δήλωσε ότι η προσπάθεια θα μπορούσε να οδηγήσει σε ταχύτερη διάγνωση των γενετικών παθήσεων των οφθαλμών και την ανάπτυξη νέων θεραπευτικών προσεγγίσεων. ●

Ταξινόμηση γενετικών παραλλαγών με κριτήρια ACMG/AMP και γνώμονα την Τεχνητή Νοημοσύνη

Η Fabric Classification Engine (ACE) είναι η νέα μηχανή εξαγωγής συμπερασμάτων βασισμένη σε Τεχνητή Νοημοσύνη (Artificial Intelligence, AI), η οποία αξιοποιώντας πληροφορίες γονιδίων και γονιδιακών παραλλαγών επιτρέπει την ακριβή χαρακτηρισμό γενετικών παραλλαγών με βάση τα κριτήρια του American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG / AMP).

Με τον όρο ACMG/AMP, αναφερόμαστε σε ένα σύνολο οδηγιών για την εκτίμηση της παθογένειας κληρονομικών νόσων και της διάγνωσης αυτών με βάση τις μεταλλαγές. Οι πολυάριθμες πληροφορίες που απαιτούνται για τον σωστό χαρακτηρισμό των αλλαγών, είναι δύσκολο να ενσωματωθούν αποτελεσματικά χειρονακτικά, με αποτέλεσμα να προκύπτουν πολλές ασυμφωνίες στην ερμηνεία των αποτελεσμάτων και υψηλό κόστος.

Η νέα μηχανή ταξινόμησης AI (ACE) του Fabric είναι μια μηχανή τεχνητής νοημοσύνης που συνάγει αυτόματα την ταξινόμηση των μεταλλαγών, επικυρώνοντας τα αποτελέσματα ACMG-AMP έναντι άλλων βάσεων δεδομένων, όπως ClinVar και VariSNP. Το Fabric Classification Engine είναι ενσωματωμένο στο Fabric Enterprise, μια πλατφόρμα λογισμικού που βασίζεται σε νέφωμα για κλινική γονιδιωματική ανάλυση, για να επιτρέπει σε εργαστήρια γενετικών αναλύσεων εύκολη πρόσβαση. ●

Εφαρμογή κινητού NGS αλληλούχισης νέας γενιάς, η νέα διάσταση στις αναλύσεις πεδίου

Στα πλαίσια της συνεχιζόμενης τάσης για αποκεντρωμένες γονιδιωματικές αναλύσεις με φορητά μέσα και με τη λιγότερη δυνατή εξάρτηση από δικτύωση και επικοινωνίες, το εργαστήριο Cold Spring Harbor (CSHL) ανακοίνωσε ότι ερευνητές του ανέπτυξαν το πρώτο λογισμικό ανάλυσης γονιδιώματος, που τρέχει σε κινητό τηλέφωνο, το iGenomics, μια νέα εφαρμογή για iPhone που συνδυάζεται με φορητή συσκευή αλληλούχισης DNA ώστε να προκύπτει μια συσκευή που θυμίζει το ανάλογο σαρωτή χειρός “tricorder” στην παλαιά σειρά επιστημονικής φαντασίας Star Trek. Η εφαρμογή iGenomics τρέχει στο iOS καταργώντας την ανάγκη για εξελιγμένα-και ακριβά- υπολογιστικά συστήματα, φορητά ή όχι, που έπρεπε να αναπτύσσονται επί του πεδίου, ιδίως σε περιπτώσεις οικολογικών μελετών, όπου η ανάγκη είναι πάγια, ή σε έκτακτες περιπτώσεις, όπως η απόκριση σε επιδημικό κύμα. Ο Asryn Palatnick δημιούργησε το λογισμικό iGenomics στο CSHL στο εργαστήριο του Αναπληρωτή Καθηγητή Schatz, εντός οκτώ ετών, αρχίζοντας από δεκατεσσάρων, ως μαθητής. Σήμερα μηχανικός λογισμικού για τη Facebook, ο Palatnick είχε ήδη εμπειρία στην ανάπτυξη εφαρμογών για το iPhone όταν συνεργάστηκε με το εργαστήριο του Schatz. Το iGenomics αναπτύχθηκε για να συνδυάζεται με τα διάφορα φορητά μοντέλα του αλληλοχοποποιητή Oxford Nanopore, καθώς διαπιστώθηκε ότι ενώ οι αλληλοχοποποιητές σμικρύνονταν, δεν συνοδεύονταν από ανάλογες τεχνολογικές εξελίξεις στην υπολογιστική υποδομή που τους υποστήριζε με αποτέλεσμα αδυναμία διεξαγωγής της μελέτης σε περιβάλλον πλήρως φορητής συσκευής. Οι κομβικές υπολογιστικές διεργασίες, όπως σύγκριση, παράθεση και ανάλυση των δεδομένων αλληλουχίας απαιτούσαν διασυνδεδεμένους εξυπηρετητές ή εξειδικευμένους, εξαιρετικά ακριβούς φορητούς υπολογιστές. Το προσωπικό αμέσου επεμβάσεως σε επιδημίες μετακινείται με βαλίτσες γεμάτες από ανταλλακτικά συσκευών Nanopore και φορητούς υπολογιστές προκειμένου

να αναπτυχθεί και να διεξάγει αναλύσεις πεδίου σε απομακρυσμένες περιοχές. Με το iGenomics αυτές οι επιχειρήσεις γίνονται περισσότερο προσιτές σε τιμή και τεχνογνωσία και εύκαμπτες σε σχεδιασμό και διεξαγωγή. Οι χρήστες μπορούν να ανταλλάξουν δεδομένα αλληλουχιών μεταξύ τους ακόμη και στις πλέον απρόσιτες περιοχές όπου δεν υπάρχει Διαδίκτυο, με αποτέλεσμα να εξετάζεται η χρήση της εφαρμογής σε διαστημικά πειράματα αλληλούχισης DNA, καθώς είναι γνωστό ότι υπάρχει μεγάλο ενδιαφέρον για τη διεξαγωγή αναλύσεων μοριακής βιολογίας σε τέτοιες συνθήκες και περιβάλλοντα, για ποικίλους λόγους, κάτι αδύνατο με τυπική υπολογιστική υποδομή και ανάγκες συνεχούς ανταλλαγής δεδομένων μέσω δικτύων εξυπηρετητών και βάσεων δεδομένων.

Στην επιθεώρηση Gigascience, οι Palatnick και Schatz αναφέρουν ότι ο αλγόριθμος iGenomics μπορεί να χαρτογραφήσει ταχύτατα αλληλουχίες ιικών παθογόνων, όπως οι ιοί Zika και οι γριποϊοί, ώστε να εντοπίσει μεταλλάξεις και παραλλαγές σημασίας για την διάγνωση και θεραπεία. Επίσης προσφέρεται διαδικτυακή εκπαίδευση για ανάπτυξη άλλων εφαρμογών, όπως η γονιδιωματική ανάλυση του ιού SARS-CoV-2. Με τον τρόπο αυτό οι αποκεντρωμένες αναλύσεις, που μέχρι τώρα γίνονταν με αναλώσιμα δοκίμια ή με φορητά εργαστήρια με δυνατότητα εξέτασης ορισμένων μόνο τόπων περνά σε άλλο επίπεδο, όπου η πλήρης φορητότητα και η μαζική ανάλυση συνυπάρχουν. Ο Schatz πιθανολογεί ότι η νέα εφεύρεση θα είναι χρήσιμη σε προσωπικό ανεπτυγμένο στο πεδίο, αλλά και σε αστικό περιβάλλον. Θεωρεί ότι όπως σήμερα όλοι κυκλοφορούν με μια κάμερα σχετικά υψηλής ευκρίνειας στην τσέπη, έτσι και στο μέλλον ο καθένας θα μπορεί να κάνει μια γενετική ανάλυση του αμέσου περιβάλλοντός του για να προσδιορίσει την βιοασφάλειά του, πχ από παθογόνους παράγοντες, ή και του εαυτού του, χωρίς εξάρτηση από προσωπικό εταιρικής ή άλλης υπαγωγής και ως εκ τούτου με μέγιστη διακριτικότητα. ●

ΑΡΘΡΟ ΑΝΑΣΚΟΠΗΣΗΣ

Εξατομικευμένη Ιατρική και μαζικά βιοϊατρικά δεδομένα

Μαργαρίτα-Ιωάννα Κουφάκη*, Ιωάννης Γ. Χατζής², Γεώργιος Π. Πατρινός¹

¹Πανεπιστήμιο Πατρών, Σχολή Επιστημών Υγείας, Τμήμα Φαρμακευτικής, Πάτρα

²Δημόσιο Ινστιτούτο Επαγγελματικής Κατάρτισης Μεσολογγίου, Μεσολόγγι

*Πανεπιστήμιο Πατρών, Σχολή Επιστημών Υγείας, Τμήμα Φαρμακευτικής, Εργαστήριο Φαρμακογονιδιωματικής και Εξατομικευμένης Θεραπείας, Πανεπιστημιούπολη, Ρίο, 265 04, Πάτρα. Τηλέφωνο: 2610-962339, Email: gpatrinos@upatras.gr

ΠΕΡΙΛΗΨΗ

Η εξατομικευμένη ιατρική αναπτύχθηκε τα τελευταία χρόνια ως μια νέα θεραπευτική προσέγγιση που αφορά τη θεραπεία και περίθαλψη ενός ασθενή με γνώμονα τις βιολογικές και περιβαλλοντικές πληροφορίες που έχουν συγκεντρωθεί για αυτόν. Κάθε μέρα ένας μεγάλος όγκος βιοϊατρικών δεδομένων συγκεντρώνεται και αποθηκεύεται σε ηλεκτρονικούς φακέλους υγείας, δημιουργώντας έναν τεράστιο όγκο δεδομένων, τα οποία παρουσιάζουν διάφορες προκλήσεις για τη διαχείριση και την ανάλυσή τους. Διάφοροι τομείς γνώσεων είναι απαραίτητοι για την πρόσκτηση των δεξιοτήτων και ικανοτήτων που απαιτούνται για να αξιοποιούνται πλήρως οι δυνα-

τότητες που προσφέρει αυτός ο μεγάλος όγκος των βιοϊατρικών δεδομένων. Αρχίζουμε με μια εισαγωγή στην έννοια «μαζικά δεδομένα» καθώς και στην αποθήκευση και στη διαχείριση τους. Εξετάζουμε τις στατιστικές μεθόδους και την επιστήμη των δεδομένων, που αποτελούν σημαντικά θεμέλια για την τεχνητή νοημοσύνη, τη μηχανική εκμάθηση και την επεξεργασία φυσικής γλώσσας, απαραίτητες προϋποθέσεις για την ανάπτυξη προγνωστικών μοντέλων με σκοπό τη λήψη κλινικών αποφάσεων. Εν κατακλείδι, προτείνουμε ειδική κατάρτιση για την προετοιμασία της νέας γενιάς επιστημόνων για τον χειρισμό των μαζικών βιοϊατρικών δεδομένων.

ΛΕΞΕΙΣ ΕΥΡΕΤΗΡΙΟΥ: Εξατομικευμένη Ιατρική, μαζικά βιοϊατρικά δεδομένα, τεχνητή νοημοσύνη, μηχανική εκμάθηση, ηλεκτρονικός φάκελος υγείας

Εξατομικευμένη ιατρική και Ιατρική Ακριβείας

Ιατρική ακριβείας ονομάζεται η διαδικασία προσαρμογής της ιατρικής πράξης στα ατομικά χαρακτηριστικά (γενετικά και μη) ενός συγκεκριμένου ασθενούς [1-3].

Κάτι τέτοιο δεν συνεπάγεται τη δημιουργία στοχευμένων φαρμακευτικών προϊόντων προσαρμοσμένων στο γονότυπο ενός συγκεκριμένου ασθενούς, μιας και αυτός είναι στόχος της εξατομικευμένης ιατρικής. Στην πραγματικότητα, η διάκριση μεταξύ της ιατρικής ακρι-

* Αντεπιστέλλων Συγγραφέας

Πανεπιστήμιο Πατρών, Σχολή Επιστημών Υγείας, Τμήμα Φαρμακευτικής, Πανεπιστημιούπολη, Ρίο, 265 04, Πάτρα.
Τηλεφωνο: 2610-962339, Email: koufakimargianna@gmail.com

βείας και της εξατομικευμένης ιατρικής έγκειται στο ότι η πρώτη επικεντρώνεται στη δημιουργία ενός νέου τρόπου ταξινόμησης μιας ασθένειας, ο οποίος λαμβάνει υπόψη του τα μοναδικά χαρακτηριστικά ενός ασθενούς, ενώ η εξατομικευμένη ιατρική επικεντρώνεται στην ανακάλυψη/ανάπτυξη εξειδικευμένων μεθόδων θεραπείας (π.χ. φάρμακα, ανοσοθεραπείες) που στοχεύουν στις ατομικές γονιδιωματικές υπογραφές του ασθενούς [2]. Αυτή η ειδοποιός διαφορά οδήγησε στη συχνότερη εφαρμογή της «εξατομικευμένης ιατρικής» από τη φαρμακοβιομηχανία, ενώ την «ιατρική ακριβείας» μελετούν ιδιαίτερα οι κλινικοί ιατροί που ενδιαφέρονται να σχεδιάσουν κατάλληλες κλινικές οδηγίες για διαφορετικές υποομάδες ασθενών.

Προκειμένου να προσαρμόσουν τις ιατρικές θεραπείες στα χαρακτηριστικά μεμονωμένων ασθενών, οι ερευνητές έχουν επικεντρωθεί στη διάκριση ασθενών σε «υποομάδες ιδιαίτερης σημασίας». Η δημιουργία αυτών των υποομάδων «ιδιαίτερης σημασίας» γίνεται με γνώμονα ένα συγκεκριμένο χαρακτηριστικό του ασθενούς. Για παράδειγμα, αυτό το χαρακτηριστικό θα μπορούσε να είναι η εθνικότητα, το φύλο ή η σεξουαλική ταυτότητα, η κοινωνικοοικονομική ομάδα (χαμηλό ή υψηλό εισόδημα), ή ακόμη ο τύπος της ασθένειας (π.χ. άσθμα).

Οι ασθενείς μπορεί επίσης να συσχετισθούν μεταξύ τους έχοντας ως κριτήριο τον τύπο της αλλεργίας τους. Για παράδειγμα, όλοι οι ασθενείς με αλλεργίες στα αυγά είναι μια υποομάδα ασθενών ιδιαίτερης σημασίας που απαιτεί προσαρμοσμένα θεραπευτικά σχήματα που είναι ιδιαίτερα σημαντικό να γνωρίζουν οι κλινικοί ιατροί. Π.χ. Αυτοί οι ασθενείς θα πρέπει να λάβουν το εμβόλιο κατά της γρίπης σε διαιρεμένη δόση [4] και όχι σε εφάπαξ.

Ενώ κάποια χαρακτηριστικά είναι ήδη γνωστά για τη δημιουργία υποομάδων ασθενών, όπως όσων πάσχουν από τροφικές αλλεργίες, όπως αναφέρθηκε προηγουμένως, ωστόσο, δεν είναι γνωστά χαρακτηριστικά με τα οποία θα μπορούσε να δημιουργηθούν νέες υποομάδες ασθενών. Για το σκοπό αυτό έχουν σχεδιαστεί μέθοδοι πληροφορικής για τη δημιουργία υποομάδων ασθενών, οι οποίες αναζητούν κοινά χαρακτηριστικά α) στα δεδομένα που έχουν καταγραφεί στον ηλεκτρονικό φάκελο υγείας [5], β) στις χρονικές μεταβολές στις καταστάσεις εργαστηριακών τιμών (π.χ. ελεγχόμενος έναντι ανεξέλεγκτου διαβήτη) [6], γ) στις ανεπιθύμητες αντιδράσεις σε φάρμακα που έχουν αναφερθεί λόγω γενετικών παραγόντων όπως μεταλλάξεις του CYP [7] αλλά και δ) στους τύπους καρκίνου [8]. Απαιτούνται εξελιγμένες μέθοδοι πληροφορικής για τον εντοπισμό συγκεκρι-

κριμένων πληθυσμών ασθενών και τη διαστρωμάτωση σε υποπληθυσμούς ιδιαίτερης κλινικής σημασίας, ώστε να μπορεί να εφαρμοστεί σ' αυτούς ιατρική ακριβείας. Αυτό παραμένει ένα ερευνητικό πεδίο με σημαντικές προκλήσεις, καθώς κάθε μεμονωμένος ασθενής εμφανίζει ένα σύνθετο συνδυασμό φαινοτύπων της νόσου και συμπτωματολογίας [9].

Μαζικά βιοϊατρικά δεδομένα

Με τον όρο «μαζικά δεδομένα -Big Data» ορίζουμε τον τεράστιο όγκο δεδομένων ο οποίος με τις παραδοσιακές υπολογιστικές μεθόδους και τεχνικές είναι σχεδόν αδύνατο να αναλυθεί και να εξαχθούν αποτελέσματα.

Οι υπολογιστικές προκλήσεις αφορούν τη χωρητικότητα των μέσων αποθήκευσης δεδομένων, το εύρος ζώνης του δικτύου για τη μετακίνηση δεδομένων από μια συσκευή αποθήκευσης σε μια άλλη ή το πλήθος και το είδος των υπολογισμών που είναι απαραίτητοι για την επεξεργασία και την ανάλυση των δεδομένων. Επίσης, πολλές φορές τα υπολογιστικά αποτελέσματα που παράγονται συμβαίνει να υπερβαίνουν τον όγκο των ίδιων των δεδομένων.

Τα μαζικά δεδομένα χαρακτηρίζονται συχνά από τα τέσσερα V.

Το πρώτο V είναι ο όγκος (Volume) των δεδομένων. Αυτό το χαρακτηριστικό αναφέρεται στη χωρητικότητα των μέσων που απαιτείται για την αποθήκευση των δεδομένων και είναι ένα από τα πιο σημαντικά προβλήματα που αντιμετωπίζει κάποιος με τα μαζικά δεδομένα. Το δεύτερο V είναι η ταχύτητα (Velocity). Τα δεδομένα ίσως να παράγονται από κάποιο σύστημα ηλεκτρονικών μετρήσεων όπως για παράδειγμα μια φορητή συσκευή η οποία δημιουργεί τα δεδομένα πιο γρήγορα από ότι η υπολογιστική δομή που κάποιος διαθέτει μπορεί να διαχειριστεί και να αποθηκεύσει. Αυτό καθιστά αδύνατη τη μεταφορά και αποθήκευση δεδομένων από τη μια συσκευή στην άλλη. Το τρίτο V είναι η ποικιλία (Variety). Τα μαζικά δεδομένα δεν είναι πάντα ομοιογενή αλλά είναι συχνά ένας ετερογενής συνδυασμός τύπων δεδομένων που προέρχονται από διαφορετικές πηγές μέτρησης. Ένα σχετικό παράδειγμα αποτελούν οι ηλεκτρονικοί φάκελοι υγείας, γιατί περιέχουν δεδομένα διαφορετικής μορφής, όπως θα αναλύσουμε παρακάτω. Το τέταρτο V είναι η αξιοπιστία (Veracity), δηλαδή πόσο ακριβή και σαφή είναι τα δεδομένα ώστε να μπορεί να τα εμπιστευτεί κανείς ως αξιόπιστα. Τα μαζικά δεδομένα περιέχουν συχνά ασάφειες, λανθασμένες ή ελλιπείς τιμές. Αυτό δημιουργεί τεράστιες προκλήσεις στην προσπάθεια να υπάρξει τελικά ένα σύνολο δεδο-

μένων έτοιμο για ανάλυση.

Όλα τα παραπάνω χαρακτηριστικά θεωρούνται τα πιο συνηθισμένα στη βιβλιογραφία. Όπως, υπάρχουν τουλάχιστον δύο ακόμη που θα μπορούσαν να συμπεριληφθούν σε αυτά τα τέσσερα. Το πρώτο είναι η πολυπλοκότητα (Vexedness) ή μεταβλητότητα. Τα δεδομένα που είναι ιεραρχικά ή διαχρονικά προσθέτουν σύνθετες διαστάσεις όταν μάλιστα συνδυάζονται με όλα τα παραπάνω V.

Το δεύτερο, είναι η αξία (Value) των δεδομένων. Υπάρχει όπως προβληματισμός ορισμένες φορές για το αν τα δεδομένα που συλλέγονται είναι αξιόλογα δικιολογείται να τεθεί σε εφαρμογή μια διαδικασία για τη μαζική αποθήκευση δεδομένων και επεξεργασίας τους με υπολογιστές υψηλής απόδοσης. Τα μαζικά δεδομένα δεν είναι πάντα η βέλτιστη λύση, και όπως ορισμένοι ισχυρίζονται οι στοχευμένες προσεγγίσεις με ανάλυση μικρότερου όγκου δεδομένων ίσως να έχουν καλύτερα αποτελέσματα στην απάντηση ορισμένων επιστημονικών ερωτημάτων [10].

Ηλεκτρονικός Φάκελος Υγείας – ΗΦΥ

Ο ηλεκτρονικός φάκελος υγείας αποτελεί ένα σύγχρονο παράδειγμα σημαντικής πηγής μαζικών δεδομένων που έχει ευρεία εφαρμογή σε χιλιάδες ακαδημαϊκά και ιατρικά κέντρα παγκοσμίως. Λόγω των χαρακτηριστικών του έχει αποκτήσει ένα σπουδαίο ρόλο στην παρακολούθηση των δεδομένων και πληροφοριών των ασθενών αλλά και στη βελτίωση της τιμολόγησης των ιατρικών υπηρεσιών. Η εξέλιξη της τεχνολογίας της πληροφορικής επέτρεψε την υιοθέτηση του ΗΦΥ κυρίως λόγω της ανέξοδης αποθήκευσης δεδομένων και τις διαθέσιμες βάσεις δεδομένων που μπορούν να χειριστούν αποτελεσματικά μαζικά δεδομένα. Οι βάσεις δεδομένων που χρησιμοποιούνται για τους ΗΦΥ έχουν τεράστιες δυνατότητες. Πιο συγκεκριμένα, μπορούν να ενσωματώσουν ένα εντυπωσιακό εύρος διαφορετικών δεδομένων ασθενών, συμπεριλαμβανομένων δημογραφικών στοιχείων, εργαστηριακών εξετάσεων, απεικονίσεων, ιατρικού ιστορικού και της χρήσης οποιασδήποτε φαρμακευτικής αγωγής, καθώς και κλινικών σημειώσεων που περιλαμβάνουν σχόλια τόσο από τον ιατρό όσο και από τον ασθενή σε ελεύθερο κείμενο. Το τελευταίο διάστημα προστίθενται στα παραπάνω όλο και περισσότερα γονιδιωματικά δεδομένα αλλά και δεδομένα επιτήρησης υγείας που καταγράφονται από ενδεδειγμένες ή έξυπνες συσκευές.

Αυτές οι νέες πηγές μαζικών δεδομένων δημιουργούν προκλήσεις στη διαχείριση των δεδομένων αλλά και στη χρήση τους για τη λήψη κλινικών αποφάσεων. Με άλλα

λόγια, η συσσώρευση πρωτογενών δεδομένων οδηγεί στην ανάγκη για ανάπτυξη μεθόδων διαχείρισης των ροών δεδομένων αλλά και προγραμμάτων μετατροπής τους σε μορφή αξιοποιήσιμη από την εξατομικευμένη ιατρική. Επιπλέον για κάθε ασθενή προκύπτουν νέα δεδομένα μέσω της χρήσης ιατρικών εφαρμογών αλλά και έξυπνων φορητών συσκευών. Σύμφωνα με μελέτες, εκτιμάται ότι κάθε ασθενής στο εγγύς μέλλον θα έχει πολλά terabytes ή ακόμα και petabytes δεδομένων και πληροφοριών που θα πρέπει να αποθηκευτούν και να επεξεργαστούν καταλλήλως ως μέρος της φροντίδας του. Αυτά τα δεδομένα θα χρησιμεύσουν ως πρώτη ύλη τόσο για εξατομικευμένη ιατρική όσο και για ιατρική ακρίβειας οι οποίες θα αναπτυχθούν με τη βοήθεια των μεθόδων που περιγράφονται παρακάτω.

Διαχείριση και Ενσωμάτωση Δεδομένων

Εάν υποθέσουμε ότι τα δεδομένα είναι οι δομικοί λίθοι για την έρευνα και εφαρμογή της εξατομικευμένης ιατρικής, τότε οι βάσεις δεδομένων είναι το κονίαμα που εξασφαλίζει την συνοχή τους. Οι βάσεις δεδομένων παρέχουν τη δομή στην οποία διατηρούνται και διατίθενται τα δεδομένα για μελλοντική χρήση. Τα συστήματα βάσεων δεδομένων παρέχουν τους υπολογιστικούς μηχανισμούς, που χρειάζονται για την αποθήκευση, επεξεργασία και ανάκτηση δεδομένων, συνήθως μέσω γραφικών διεπαφών και γλωσσών επερωτήσεων. Η κυρίαρχη αρχιτεκτονική των βιοϊατρικών βάσεων δεδομένων είναι σχεσιακή. Τα δεδομένα αποθηκεύονται σε πίνακες που αντιπροσωπεύουν μια συγκεκριμένη οντότητα, όπως τα δημογραφικά στοιχεία.

Το τελευταίο διάστημα παρατηρείται ένα αυξανόμενο ενδιαφέρον για τις βάσεις δεδομένων στις οποίες τα δεδομένα παρουσιάζονται ως κόμβοι σε ένα μη κατευθυνόμενο γράφημα και οι σχέσεις μεταξύ των κόμβων ως σύνδεσμοι [11,12]. Όποια αρχιτεκτονική και να έχει η βάση δεδομένων, ο χρήστης μπορεί να χειριστεί τα δεδομένα μέσω ερωτημάτων χρησιμοποιώντας μια συγκεκριμένη γλώσσα προσαρμοσμένη για τη βάση αυτή. Ανεξάρτητα από την αρχιτεκτονική της βάσης δεδομένων, όλες οι βάσεις δεδομένων σχεδιάζονται και υλοποιούνται χρησιμοποιώντας ένα μοντέλο δεδομένων, το οποίο καθορίζει τα χαρακτηριστικά της βάσης δεδομένων και παρέχει οδηγίες για τους εμπλεκόμενους.

Είναι αλήθεια ότι οι ειδικοί της επιστήμης δεδομένων διστάζουν να αναπτύξουν εξελιγμένες βάσεις δεδομένων για την αποθήκευση και ενσωμάτωση βιοϊατρικών δεδομένων λόγω της πολύπλοκης φύσης αυτών των δεδομένων. Μολαταύτα, οι ειδικοί διαθέτουν κάποιες υπολογιστικές δυνατότητες ώστε να ενσωματώνουν ή

να συνδέουν φαινομενικά ανόμοιες βάσεις δεδομένων για να προκύπτει μια ολοκληρωμένη εικόνα ενός κλινικού προβλήματος. Για παράδειγμα, στην εκτίμηση της αιτίας για την οποία το HbA1C (glycated haemoglobin) του διαβητικού ασθενούς είναι εκτός ελέγχου, τα δεδομένα που λαμβάνονται από τη μόνιμη παρακολούθηση της σωματικής δραστηριότητας με τη βοήθεια μιας φορητής προσωπικής συσκευής, όπως το Fitbit, θα μπορούσαν να προστεθούν στα κλινικά δεδομένα του ασθενούς. Αυτό θα βοηθούσε τον ιατρό να ποσοτικοποιήσει τη σωματική δραστηριότητα ενός συγκεκριμένου ασθενούς για να βελτιστοποιήσει το συνολικό πρόγραμμα διαβητικού ελέγχου αυτού. Η άνθιση νέων πηγών δεδομένων οδήγησε σε ένα αυξημένο ενδιαφέρον για ανάπτυξη νέων μεθόδων σύνδεσης και συσχέτισης των διαφόρων αρχείων, αναγνωρίζοντας παράλληλα τις δυσκολίες και προκλήσεις που αυτό συνεπάγεται [13,14].

Προκλήσεις

Πέρα από την ποικιλία τύπων και πηγών δεδομένων που περιέχει ο ΗΦΥ, υπάρχουν και άλλα στοιχεία ποικιλομορφίας στα βιοϊατρικά δεδομένα που δυσχεραίνουν την ενσωμάτωσή τους σε βάσεις δεδομένων. Η διαφορά μονάδων μέτρησης ανάμεσα στις βάσεις δεδομένων είναι, για παράδειγμα, ένα συχνό πρόβλημα που εμποδίζει την καθολική επεξεργασία των δεδομένων και την κανονικοποίησή τους. Αυτή η ποικιλομορφία στην παρουσίαση των δεδομένων αποτελεί μια σημαντική πρόκληση για την αποτελεσματική και ακριβή συσχέτιση και σύνδεση των δεδομένων καθώς και για την τελική ενσωμάτωσή τους στη βάση.

Για να αντιμετωπιστεί αυτή η πρόκληση, πρέπει να στραφούμε στις επιστημολογικές διαστάσεις των βιοϊατρικών δεδομένων με τον τρόπο που αυτές παρουσιάζονται στο συντακτικό τους μέρος και στη σημασιολογία τους. Αυτό είναι απαραίτητο για την εναρμόνιση των δεδομένων αλλά και για την ενσωμάτωσή τους στη συνέχεια. Όλο και περισσότερο, οι επαγγελματίες της πληροφορικής στρέφονται στις οντολογίες για να πραγματοποιήσουν τη διαδικασία εναρμόνισης των δεδομένων, προκειμένου να αποδώσουν τις έννοιες και τις σχέσεις μεταξύ τους σε μορφή γραφημάτων- πχ η ενσωμάτωση των multi-omics δεδομένων στην υπάρχουσα βιολογική γνώση [15]. Ως εκ τούτου, οι οντολογίες μπορούν να θεωρηθούν ως τύπος μοντέλου δεδομένων.

Ένα άλλο πιεστικό ζήτημα στη διαχείριση και ενσωμάτωση δεδομένων είναι η διασφάλιση της ποιότητας αυτών. Ένας σοβαρός λόγος για την κακή ποιότητα των

δεδομένων είναι ότι υπάρχουν ελλιπή στοιχεία, αφού είναι πιθανό εργαστηριακές αναφορές να μην αποθηκευτούν στον ιατρικό φάκελο ή τμήμα μιας φυσικής εξέτασης να μην ολοκληρωθεί. Προκειμένου να προκύψουν αποτελεσματικά συμπεράσματα από τα δεδομένα, πρέπει να υπάρχει ένα προκαθορισμένο πρωτόκολλο για τη διαχείριση των δεδομένων που λείπουν, τα οποία μπορεί να περιλαμβάνουν διάφορες μεθόδους καταλογισμού, όπως το deep learning [16]. Οι Kim και συνεργάτες εφάρμοσαν ένα νέο πλαίσιο ενσωμάτωσης για την πρόβλεψη τιμών ίδιας μορφής με τα δεδομένα που λείπουν [17].

Μια πιθανή λύση στα προβλήματα που προκύπτουν στη διαχείριση μαζικών κλινικών δεδομένων είναι οι αποθήκες κλινικών δεδομένων μεγάλης κλίμακας. Αυτές αποτελούν σημαντικούς πόρους δεδομένων, η δημιουργία και συντήρηση των οποίων εξυπηρετούν μια μεγάλη ποικιλία χρηστών και ενδιαφερομένων στον τομέα της εξατομικευμένης ιατρικής. Κατά τη διαδικασία εξαγωγής-μεταφοράς-φόρτωσης που είναι ο κύριος ρόλος του παραδείγματος χρήσης μιας αποθήκης δεδομένων, μπορούν να εφαρμοστούν διαδικασίες που εξασφαλίζουν την ποιότητα των δεδομένων και ενσωματώνουν δεδομένα από διάφορες πηγές με αποτέλεσμα την παροχή μιας ασφαλούς πλατφόρμας για την ανάλυση των δεδομένων που δεν έχουν ταυτοποιηθεί. Αυτές οι αποθήκες θα μπορούσαν να προσφέρουν στους ερευνητές τη δυνατότητα ταυτοποίησης και εξαγωγής επιθυμητού κούρτιου και στους ειδικούς της ιατρικής ακριβείας τα μέσα για την αντιμετώπιση ενός ασθενούς με ένα συγκεκριμένο φαινότυπο την ώρα της περίθαλψης.

Στατιστική Ανάλυση

Η στατιστική ανάλυση είναι ο κλάδος των μαθηματικών που χρησιμοποιεί μοντέλα για να συνοψίσει τα δεδομένα και να εξαχθούν συμπεράσματα. Το όριο μεταξύ της στατιστικής ανάλυσης και της μηχανικής εκμάθησης (machine learning, το πεδίο της τεχνητής νοημοσύνης που επιτρέπει στα συστήματα ηλεκτρονικών υπολογιστών να μαθαίνουν από τα δεδομένα) είναι απροσδιόριστο και αποτελεί κοινό θέμα συζήτησης. Κάθε ένας από τους δύο αυτούς κλάδους επικεντρώνεται σε μια διαφορετική πτυχή της διαδικασίας εξαγωγής συμπερασμάτων από τα δεδομένα. Η μηχανική εκμάθηση επικεντρώνεται στην πραγματοποίηση προβλέψεων ακριβείας από τα δεδομένα, ενώ η στατιστική ανάλυση επικεντρώνεται στην αξιολόγηση της εγκυρότητας ενός μοντέλου για τα δεδομένα για την εξαγωγή συμπερασμάτων. Έτσι, η στατιστική ανάλυση και η εκμη-

χανική μάθηση αλληλοκαλύπτουν τα μεταξύ τους κενά ξεπερνώντας τους υπάρχοντες περιορισμούς. Αρκετοί μάλιστα πιστεύουν ότι, καθώς τα επιστημονικά πεδία των εφαρμοσμένων μαθηματικών και της επιστήμης των υπολογιστών συνεχίζουν να αλληλεπιδρούν, οι δύο κλάδοι θα συγχωνευθούν σε ένα μόνο πεδίο, που μερικές φορές αποκαλείται στατιστική εκμάθηση [18].

Η σημασία της στατιστικής ανάλυσης στις βιοϊατρικές εφαρμογές είναι αξιοσημείωτη. Εστιάζοντας το αντικείμενο της μελέτης στο μοντέλο των δεδομένων και καθιστώντας σαφείς τις παραδοχές της μοντελοποίησης, η στατιστική ανάλυση επιτρέπει ερμηνευτικές και αιτιολογημένες υποθέσεις χάρη στη χρήση των πιθανοτήτων και του στατιστικού συμπεράσματος. Αυτά τα δυο αποτελούν θεμελιώδη στοιχεία της στατιστικής ανάλυσης, γιατί επιτρέπουν στους βιοϊατρικούς ερευνητές να ποσοτικοποιούν με ακρίβεια υποθέσεις σχετικά με τα δεδομένα, όσον αφορά τις πιθανότητες ή τις συχνότητες εμφάνισης σε σχέση με το στατιστικό μοντέλο που χρησιμοποιούν. Αυτό θεωρείται κρίσιμο σε κλινικές εφαρμογές, όπου οι αποφάσεις που επηρεάζουν την υγεία των ασθενών πρέπει να δικαιολογούνται με ακριβείς, ορθολογικούς και εύχρηστους τρόπους για την εκπλήρωση των κοινών δεοντολογικών και νομικών απαιτήσεων.

Χρησιμοποιώντας τα εργαλεία των στατιστικών συμπερασμάτων, οι ερευνητές μπορούν να δοκιμάσουν πολλαπλά μοντέλα ή υποθέσεις και να προσδιορίσουν εκείνα που ερμηνεύουν καλύτερα τα δεδομένα. Με τη διαμόρφωση της διαδικασίας συμπερασμάτων όσον αφορά τις κατανομές πιθανοτήτων, μπορούν να ποσοτικοποιήσουν την αβεβαιότητα στα συμπεράσματά τους ως συνέπεια των στοχαστικών παραγόντων, όπως τα σφάλματα μέτρησης και τα ελλειπή δεδομένα.

Στην εποχή των μαζικών δεδομένων, αυτές οι πτυχές αποκτούν πρωταρχική σημασία, καθώς οι δυσκολίες που απαντώνται στον έλεγχο μεγάλου αριθμού υποθέσεων ή στη βελτιστοποίηση σύνθετων μοντέλων μπορούν να οδηγήσουν σε εσφαλμένη ερμηνεία των δεδομένων. Η στατιστική ανάλυση συμπληρώνει με αυτόν τον τρόπο τις προσεγγίσεις της μηχανικής εκμάθησης και της τεχνητής νοημοσύνης στην ανάλυση βιοϊατρικών δεδομένων με ουσιαστικούς τρόπους, όπως θα δούμε παρακάτω. Επιπρόσθετα, είναι κρίσιμο η επόμενη γενιά επιστημόνων να είναι εφοδιασμένη με ένα στιβαρό υπόβαθρο στη στατιστική ανάλυση ώστε να είναι ικανή να προβεί σε αξιόπιστες προβλέψεις από μεγάλα σύνολα βιοϊατρικών δεδομένων.

Επιστήμη Δεδομένων

Η επιστήμη των δεδομένων αναφέρεται γενικότερα

στην ενσωμάτωση στατιστικών και υπολογιστικών τεχνικών, για την εξόρυξη γνώσεων από μαζικά δεδομένα. Ως ένας κλάδος που βασίζεται στα δεδομένα, η επιστήμη αυτή είναι σε θέση να απευθύνει προκαθορισμένες ερωτήσεις, καθώς και να διατυπώνει νέες υποθέσεις αμερόληπτα. Στην περίπτωση των βιοϊατρικών δεδομένων, η επιστήμη των δεδομένων μπορεί να εφαρμοστεί για απόκτηση νέων γνώσεων, βιολογικά χρηστικών ως προς τη βελτίωση της διάγνωσης, θεραπείας και πρόληψης μιας ασθένειας.

Η στατιστική, όπως περιγράφεται στην προηγούμενη ενότητα, αποτελεί θεμέλιο της επιστήμης των δεδομένων. Ωστόσο, η γνώση μόνο της στατιστικής θεωρίας δεν επαρκεί για την ανάλυση μεγάλων και πραγματικών συνόλων δεδομένων. Οι δεξιότητες πληροφορικής είναι ζωτικής σημασίας για την επιστήμη των δεδομένων, καθώς ο όγκος των συνόλων δεδομένων έχει αυξηθεί, κάτι το οποίο απαιτεί τη χρήση υπολογιστών για την αποτελεσματική αποθήκευση, αναζήτηση και ανάλυση τους. Η καλή γνώση των υπολογιστών που χρειάζεται για να είναι αποτελεσματική η επιστήμη των δεδομένων απαιτείται στην πράξη και όχι θεωρητικά.

Για το λόγο αυτό, ο όρος "γνώση των υπολογιστών" συχνά αντικαθίσταται από τον όρο "δεξιότητες χακαρίσματος" ή στα αγγλικά "hacking skills". Είναι αλήθεια ότι για κάποιον ειδικευμένο στην επιστήμη των δεδομένων είναι απαραίτητο ένα υπόβαθρο στην επιστήμη των υπολογιστών με την προϋπόθεση να διαθέτει ικανότητες στον προγραμματισμό για να αντιληφθεί πλήρως και να αναλύσει τα δεδομένα. Υπάρχουν πολλές γλώσσες προγραμματισμού για ανάλυση δεδομένων αλλά οι επιστήμονες χρησιμοποιούν πιο συχνά τις Python και R. Στην περίπτωση των πιο μεγάλων συνόλων δεδομένων, χρησιμοποιούνται διάφορα άλλα προγράμματα όπου η χρήση της C είναι απαραίτητη, ενώ η αποθήκευσή τους απαιτεί επιπλέον τη χρήση βάσεων δεδομένων όπως SQL.

Για να διασφαλίσει κάποιος ότι τα ερωτήματα που τίθενται σχετικά με τα δεδομένα είναι λογικά και οδηγούν στην ερμηνεία των αποτελεσμάτων πρέπει να διαθέτει μια εξειδίκευση στο πεδίο της επιστήμης δεδομένων. Οι επιστήμονες δεδομένων δεν αναλύουν παθητικά και αυτόματα τα δεδομένα. Αντίθετα, καλούνται να κάνουν διαλογή μεταβλητών κατά τη διάρκεια της διαδικασίας επιλογής ή / και μετατροπής αυτών, να επιλέξουν τις καταλληλότερες μεθόδους για την απάντηση συγκεκριμένων ερωτήσεων και κατά συνέπεια τον καλύτερο τρόπο επικοινωνίας και ερμηνείας των ευρημάτων χρησιμοποιώντας αποτελεσματικές τεχνικές απεικόνισης.

Η σύζευξη υψηλής τεχνικής κατάρτιση στη στατιστι-

κή, την επιστήμη των υπολογιστών, την πληροφορική παρέχει πρακτικές δεξιότητες που είναι απαραίτητες για την επιστήμη των δεδομένων. Σε αυτές περιλαμβάνονται οι γνώσεις μεθόδων ανάκτησης και επεξεργασίας δεδομένων, διεξαγωγής διερευνητικών αναλύσεων, δημιουργίας μοντέλων για την απάντηση σε επιστημονικά ερωτήματα και ενημερωτικών και οπτικά ελκυστικών παρουσιάσεων αποτελεσμάτων. Η διαδικασία της ανάλυσης δεδομένων δεν είναι γραμμική. Πολλοί επαναληπτικοί κύκλοι επεξεργασίας πρέπει γίνονται πριν ληφθούν τα "τελικά" αποτελέσματα. Δηλαδή, μετά τη διεξαγωγή διερευνητικών αναλύσεων ή κατασκευής αρχικών μοντέλων, ενδέχεται να χρειαστεί προσαρμογή των χαρακτηριστικών των δεδομένων ή / και μοντέλων που αξιοποιούνται σύμφωνα με το υπό εξέταση θέμα. Σε διάφορες περιπτώσεις, οι επιστήμονες δεδομένων αναπτύσσουν τα δικά τους εργαλεία και μεθόδους, προσαρμοσμένα στις ανάγκες που προκύπτουν κατά τις αναλύσεις πραγματικών συνόλων δεδομένων.

Για να διασφαλιστεί ότι τα αποτελέσματα της ερευνητικής διαδικασίας θα αποφέρουν το μέγιστο όφελος, πολλοί επιστήμονες δεδομένων έχουν αναπτύξει ερευνητικά πρωτόκολλα με υψηλή αναπαραγωγικότητα. Αυτά περιλαμβάνουν τη δημιουργία πακέτων λογισμικού ανοιχτού κώδικα που διατίθενται δωρεάν, τις δημοσιεύσεις προτεινόμενων βημάτων για τη λήψη αποτελεσμάτων και την ανταλλαγή δεδομένων που είναι απαραίτητα για την αναπαραγωγή των ευρημάτων [19-22]. Με αυτό τον τρόπο διατίθενται διάφορες τεχνολογίες με ταυτόχρονη προώθηση τόσο της διαφάνειας όσο και της αναπαραγωγικότητας των μεθόδων που εφαρμόζονται σε μεγάλα σύνολα δεδομένων. Το RStudio είναι ένα ολοκληρωμένο περιβάλλον ανάπτυξης που ενίσχυσε σημαντικά την χρηστικότητα και τη δημοτικότητα της γλώσσας προγραμματισμού R από τους επιστήμονες δεδομένων, κυρίως για λόγους που αφορούν στη βελτίωση της ροής εργασιών και στη διευκόλυνση της δημιουργίας εγγράφων R Markdown τα οποία μπορούν εύκολα να μετατραπούν σε διάφορες άλλες μορφές (π.χ. HTML, PDF) ώστε να παρουσιαστούν τα τελικά αποτελέσματα [22].

Μια άλλη χρήσιμη κατηγορία εφαρμογών είναι τα εργαστηριακά σημειωματάρια, όπως το Jupyter και το Apache Zeppelin. Αυτά παρέχουν διαδραστικά περιβάλλοντα υπολογιστών που βασίζονται σε τεχνολογίες διαδικτύου και υποστηρίζουν τη χρήση λογισμικού ανοιχτού κώδικα και γλωσσών προγραμματισμού, όπως οι Python, R, Scala, Groovy και SQL. Οι επιστήμονες δεδομένων μπορούν να εκμεταλλευτούν αυτά τα λογισμικά για την ανταλλαγή, την ανάλυση και την απεικόνιση των

δεδομένων, κάτι που επιτρέπει την εύκολη συνεργασία μεταξύ επιστημόνων. Για παράδειγμα, ένας επιστήμονας δεδομένων που χρησιμοποιεί Python μπορεί να επεξεργαστεί τα δεδομένα με τη χρήση της βιβλιοθήκης «pandas», να αναλύσει δεδομένα με το scikit-learn και να απεικονίσει δεδομένα χρησιμοποιώντας τη βιβλιοθήκη Altair. Επίσης, τα παραπάνω λογισμικά υποστηρίζουν επεξεργασία μαζικών δεδομένων και τεχνολογίες υπολογιστών όπως οι Hadoop, Spark και Hive. Τα συστήματα παρακολούθησης των διαδοχικών εκδόσεων λογισμικών, όπως το Git, παρέχουν αποτελεσματικά μέσα για την διαχρονική παρακολούθηση της εξέλιξης μεγάλων έργων.

Γι' αυτό το σκοπό έχει αναπτυχθεί η GitHub, μια πλατφόρμα που φιλοξενεί έργα που χρησιμοποιούν το Git για έλεγχο της έκδοσης, η οποία έχει γίνει ένα ευρέως χρησιμοποιούμενο αποθετήριο όπου μπορεί κάποιος να μοιραστεί κώδικες, μικρά σύνολα δεδομένων και αποτελέσματα αναλύσεων. Πιο πρόσφατα, οι πλατφόρμες όπως το Docker και το Singularity παρέχουν ένα φιλικό προς το χρήστη μέσο διανομής κώδικα με προεγκατεστημένα εργαλεία λογισμικού και διαδικασίες ελεγχόμενες από τους χρήστες, οι οποίες μπορούν επίσης να βοηθήσουν στην αναπαραγωγικότητα [23]. Παράλληλα έχει αναπτυχθεί εξειδικευμένο λογισμικό (π.χ. Kubernetes, OpenShift) το οποίο υποστηρίζει τη δημιουργία, τη διαχείριση και την ανάπτυξη επεκτάσιμων εφαρμογών σε κατανομημένα συστήματα. Η αύξηση της πρόσβασης σε cloud computing έχει επιτρέψει πολλά από αυτά τα εργαλεία και τεχνολογικές προσεγγίσεις να αξιοποιηθούν και για μαζικά δεδομένα [24]. Η ακαδημαϊκή κοινότητα αλλά και η βιομηχανία αξιοποιούν αυτές τις τεχνολογίες για να στηρίξουν τις προσπάθειες της επιστήμης των δεδομένων που έχουν ως στόχο τη βελτίωση της υγείας.

Τεχνητή Νοημοσύνη

Βασικές Αρχές

Ο όρος «τεχνητή νοημοσύνη» (AI) έχει εξελιχθεί και έχει πλέον αποκτήσει ένα πιο γενικό, διεπιστημονικό και περιεκτικό νόημα σε σχέση με την πρώτη χρήση του. Ως μια υποκατηγορία της επιστήμης των υπολογιστών, η τεχνητή νοημοσύνη χρησιμοποιείται συχνά ως εναλλακτικός όρος της "μηχανικής εκμάθησης". Ορθότερα όμως η «μηχανική εκμάθηση» είναι ουσιαστικά ένα παρακλάδι της AI το οποίο ασχολείται με την ευρύτερη έννοια της επαγωγικής λογικής. Ωστόσο, ένα μεγάλο μέρος προαπαιτούμενων κομβικών ζητημάτων, που επικεντρώνονται στην αφαιρετική λογική,

ισχύουν στο μεγαλύτερο μέρος των εφαρμογών βιοϊατρικής πληροφορικής, που χρησιμοποιούνται σήμερα ενεργά.

Οι θεμελιώδεις αρχές της τεχνητής νοημοσύνης σε συνδυασμό με βιοϊατρικές εφαρμογές πληροφορικής [25] είναι απαραίτητες για όσους επιθυμούν να εκμεταλλευτούν πλήρως τα μαζικά δεδομένα για εξατομικευμένη ιατρική αλλά και για άλλες εφαρμογές στην υγειονομική περιθαλψη. Αυτές οι αρχές είναι πολύ χρήσιμες και για τη βαθύτερη κατανόηση της μηχανικής εκμάθησης αλλά και για το μέλλον της έρευνας στην ΑΙ. Συνοπτικά, οι βασικές αρχές της τεχνητής νοημοσύνης επικεντρώνονται στο πώς μπορούν να οργανώνονται, να παρουσιάζονται, να ερμηνεύονται, να διερευνώνται και να εφαρμόζονται βιοϊατρικά δεδομένα με σκοπό να αντλούνται γνώσεις, να λαμβάνονται αποφάσεις και, τελικά, να γίνονται προβλέψεις.

Η κατάρτιση θα πρέπει να ξεκινά με μια ιστορική ανασκόπηση της εξέλιξης της τεχνητής νοημοσύνης, με τον προσδιορισμό των ορισμών, με αναφορά σε σημαντικές εξελίξεις σε εφαρμογές και ηθικά ζητήματα [26]. Αυτό θα πρέπει να συνδυάζεται με θέματα που σχετίζονται με τη λογική (δηλ. Πρόταση και λογική πρώτης τάξης) που περιγράφουν την κοινή επίσημη γλώσσα για δεδομένα και γνώσεις που επιτρέπουν την διεπαφή μεταξύ προσώπου και μηχανής. Υπάρχουν συγκεκριμένες προδιαγραφές για την αναπαράσταση των δεδομένων, συμπεριλαμβανομένων των πλαισίων, των κανόνων, των δεντρογραμμάτων, των οντολογιών και των σημασιολογικών δικτύων ενώ η παρουσίαση είναι ένα καίριο θέμα που συνδέει τόσο την αφαιρετική όσο και την επαγωγική λογική.

Είναι επίσης σημαντικό να κατανοηθεί ο ρόλος ενός παράγοντα ως παραδοσιακού δομικού στοιχείου ενός συστήματος ΑΙ, το οποίο αντιλαμβάνεται το περιβάλλον του μέσω αισθητήρων και επενεργεί σε αυτό με ενεργοποιητές. Ένα άλλο κρίσιμο θέμα περιλαμβάνει την εισαγωγή στα βασικά της επίλυσης προβλημάτων μέσω αλγορίθμων αναζήτησης, συμπεριλαμβανομένων των ομοιόμορφων αναζητήσεων (π.χ. εύρους ή βάθους κατά προτεραιότητα) και της ευρετικής αναζήτησης (π.χ. άπληστη μέθοδος ή A^* αναζήτηση). Η αναζήτηση είναι συναφής με τις συνήθειες προκλήσεις στην πρόσβαση στις βιοϊατρικές πληροφορίες και είναι απαραίτητη για τη βελτιστοποίηση των περιορισμών. Πολλές φορές, πρέπει να ικανοποιηθούν κάποιοι περιορισμοί ορισμένων μεταβλητών για να επιτευχθεί η λύση.

Η κατάρτιση στις βασικές αρχές της τεχνητής νοημοσύνης επεκτείνεται επίσης στην κατανόηση της αιτιολόγησης με αβεβαιότητα και του τρόπου που αυτή

συνδέεται με την πιθανοτική βιοϊατρική γνώση. Αυτό οδηγεί σε άλλα υπολογιστικά θέματα των υπό προϋποθέσεις πιθανοτήτων, της εντροπίας, της συμπερασματολογίας κατά Bayes. Η ενσωμάτωσή τους σε συστήματα βασισμένα στη γνώση, συμπεριλαμβανομένων των συμπερασμάτων βάσει κανόνων, των εξειδικευμένων συστημάτων και των σύγχρονων συστημάτων υποστήριξης των κλινικών αποφάσεων είναι απόρροια των παραπάνω και έχει ιδιαίτερη σημασία για την υγειονομική περιθαλψη [25]. Υπάρχουν ακόμη αρκετά προηγμένα θέματα τεχνητής νοημοσύνης τα οποία θα μπορούσαν να βρουν εφαρμογή στο πεδίο της βιοϊατρικής πληροφορικής στο μέλλον.

Μηχανική εκμάθηση

Όπως προαναφέρθηκε, η μηχανική εκμάθηση αποτελεί υποπεδίο της τεχνητής νοημοσύνης που ασχολείται με την ευρύτερη έννοια της επαγωγικής λογικής. Πιο συγκεκριμένα, αποτελείται από ένα σύνολο μεθόδων που μπορούν να αναγνωρίζουν και εξαγάγουν μοτίβα από ακατέργαστα δεδομένα ώστε να τα χρησιμοποιήσουν για την πρόβλεψη μελλοντικών δεδομένων ή για τη λήψη αποφάσεων [25,27,28]. Χωρίζεται δε σε δυο μεγάλες κατηγορίες: την επιτηρούμενη και μη επιτηρούμενη εκμάθηση.

Στην επιτηρούμενη εκμάθηση, χρησιμοποιείται μια συνάρτηση που αντιστοιχίζει μια εισακτέα τιμή που διαθέτει ένα σύνολο χαρακτηριστικών, σε ένα αποτέλεσμα. Αν η τιμή του αποτελέσματος είναι κατηγορική τότε η παραπάνω διαδικασία ονομάζεται ταξινόμηση ή κατηγοριοποίηση ενώ αν είναι συνεχής ονομάζεται παλινδρόμηση.

Στη μη επιτηρούμενη ή περιγραφική εκμάθηση, το μόνο που είναι γνωστό είναι οι εισακτέες τιμές βάσει των οποίων εντοπίζονται ενδιαφέροντα μοτίβα όπως ομάδες, ανωμαλίες και λανθάνοντες παράγοντες. Η ανάλυση κατά ομάδες στοχεύει στην ομαδοποίηση παρόμοιων αντικειμένων σε ομάδες ενώ η ανίχνευση ανωμαλιών στοχεύει στον εντοπισμό των αποκλίσεων στα δεδομένα. Τέλος, η εύρεση των λανθανόντων παραγόντων είναι πολύ σημαντική αφού συμβάλλει στην εξαγωγή συμπαγών αναπαραστάσεων των δεδομένων ή άλλων ενημερωτικών χαρακτηριστικών. Δεδομένου ότι πολλά βιοϊατρικά προβλήματα μπορούν να διατυπωθούν ή να αποδοθούν μέσω των παραπάνω διεργασιών φαίνεται ότι η μηχανική εκμάθηση προσφέρει μια σειρά ισχυρών εργαλείων για την επίλυση των προβλημάτων της επιστήμης των δεδομένων στη βιοϊατρική.

Για το λόγο αυτό έχουν αναπτυχθεί πολλές επιτυχημένες εφαρμογές που αφορούν τη διάγνωση ασθενειών,

την ανακάλυψη βιολογικών δεικτών και φαρμάκων, τη μελέτη ομικών πεδίων, την πρόβλεψη κλινικών αποτελεσμάτων και την παρακολούθηση των ασθενών, την εξατομικευμένη θεραπεία αλλά και την έγκαιρη πρόβλεψη επιδημιών. Επιπρόσθετα, αυτές οι εφαρμογές συμβάλλουν στη δημιουργία ηλεκτρονικών φακέλων υγείας, στην ανάπτυξη έξυπνων φορητών συσκευών που καταγράφουν την κατάσταση της υγείας του ατόμου ενώ καθοριστικό ρόλο διαδραματίζουν στην υποστήριξη της λήψης αποφάσεων στους τομείς της ακτινολογίας, δερματολογίας, οφθαλμολογίας και παθολογίας. Οι περισσότερες από αυτές στηρίζονται στις πιο γνωστές γλώσσες προγραμματισμού που χρησιμοποιούνται ευρέως σε εφαρμογές μηχανικής εκμάθησης, όπως Python, Java, R, C ++, C, JavaScript, Scala και Julia.

Μέθοδοι και αλγόριθμοι

Στον πεδίο της μηχανικής εκμάθησης αξιοποιείται ένα ευρύ φάσμα μεθόδων και αλγορίθμων το οποίο διαφοροποιείται ανάλογα με την κατηγορία εκμάθησης. Μια κλασική μέθοδος της επιτηρούμενης εκμάθησης ονομάζεται εκμάθηση του δέντρου αποφάσεων. Σε αυτήν κάθε εσωτερικός κόμβος περιγράφει μια δοκιμή σε ένα χαρακτηριστικό, κάθε κλάδος αντιστοιχεί σε ένα αποτέλεσμα και κάθε κόμβος φύλλου αντιπροσωπεύει μια τελική τιμή. Αν και τα δέντρα αποφάσεων θεωρούνται μια καλή και εύκολα ερμηνεύσιμη μέθοδος ωστόσο αποτελούν εκτιμητές μεγάλων διακυμάνσεων, αφού ελαφρώς διαφορετικές εισακτές τιμές μπορούν να οδηγήσουν σε πολύ διαφορετικές δομές δέντρων.

Για να ξεπεραστεί αυτή η αστάθεια, προτάθηκε η έννοια του τυχαίου δάσους το οποίο είναι πολύ πιο ισχυρό από το δέντρο αποφάσεων. Αυτό προκύπτει όταν συγκεντρώνονται πολλά δέντρα αποφάσεων βασισμένα σε τυχαία υποσύνολα δεδομένων και χαρακτηριστικών.

Άλλα παραδείγματα κλασικών μεθόδων επιτηρούμενης μάθησης περιλαμβάνουν Μηχανές Διανυσμάτων Υποστήριξης (SVM), γραμμική παλινδρόμηση (linear regression), μοντέλο λογιστικής παλινδρόμησης/logit (logistic regression), αφελείς Μπαεσιανοί Ταξινομητές (naive Bayes classifiers), γραμμική διακρίνουσα ανάλυση (linear discriminant analysis, LDA) και αλγόριθμος των k-πλησιέστερων γειτόνων (k-nearest neighbors algorithm, k-NN). Σε αυτές τις κλασικές μεθόδους, τα χαρακτηριστικά που αντιπροσωπεύουν ένα αντικείμενο είναι σχεδιασμένα με το χέρι και δεν είναι βελτιστοποιημένα για τη μαθησιακή εργασία.

Το τελευταίο διάστημα η μηχανική εκμάθηση έχει αρχίσει πλέον να εφαρμόζεται όχι μόνο για την χαρτογράφηση αντικειμένων αλλά για την ανακάλυψη των

αντικειμένων και την ίδια την αναπαράστασή τους. Ένα χαρακτηριστικό παράδειγμα αυτής της κατηγορίας είναι οι μέθοδοι εκμάθησης των νευρικών δικτύων [27] που έχουν αποδειχθεί εξαιρετικά επιτυχείς σε πολλούς τομείς εφαρμογών της μηχανικής εκμάθησης συμπεριλαμβανομένης της επιστήμης των βιοϊατρικών δεδομένων [29-33].

Προκλήσεις

Δεδομένης της πολύ μεγάλης κλίμακας και πολυπλοκότητας των μαζικών δεδομένων της βιοϊατρικής, η μηχανική εκμάθηση αντιμετωπίζει σημαντικές υπολογιστικές και μεθοδολογικές προκλήσεις, ώστε η τελική επιλογή μιας μεθόδου να είναι αρκετά δύσκολη και προβληματική. Αυτές είναι:

α. το ζήτημα της υπερεκπαίδευσης που προκύπτει όταν προσθέτουμε σε ένα μαθησιακό μοντέλο πολλές παραμέτρους που δεν μπορούν να δικαιολογηθούν από τα δεδομένα

β. η αδυναμία επιλογής του καταλληλότερου μοντέλου μεταξύ πολλών μοντέλων με διαφορετικές πολυπλοκότητες,

γ. η αναζήτηση της βέλτιστης στρατηγικής όταν δεν έχουμε μια κλειστού τύπου λύση,

δ. η βελτιστοποίηση των υπερπαραμέτρων όταν υπάρχουν πάρα πολλές παράμετροι

ε. η δυσκολία της βιοϊατρικής ερμηνείας των αποτελεσμάτων όταν προβλέπονται πολλά υποσχόμενα αποτελέσματα από πολύπλοκα μοντέλα

στ. η έλλειψη μιας μεθόδου μηχανικής εκμάθησης που να είναι κατάλληλη για όλους τους τύπους δεδομένων [34,35].

Για να αντιμετωπιστούν αυτές οι προκλήσεις και να καταστεί η μηχανική εκμάθηση πιο φιλική στους χρήστες, ειδικά στους μη επαγγελματίες, έχουν γίνει σοβαρές προσπάθειες στην ανάπτυξη του τομέα της αυτοματοποιημένης μηχανικής εκμάθησης (AutoML). Αυτός προβλέπει την αυτοματοποίηση της διαδικασίας εφαρμογής μηχανικής εκμάθησης σε πραγματικά προβλήματα. Τα υπάρχοντα συστήματα AutoML (π.χ. AutoWeka [36], AutoSklearn [37], TPOT [38] και PennAI [39]) έχουν σχεδιαστεί για την αυτοματοποίηση ενός ή περισσότερων βημάτων της διαδικασίας, όπως η προετοιμασία των δεδομένων, η αναζήτηση καθυκόντων, η τροποποίηση των χαρακτηριστικών, η επιλογή μοντέλου, η βελτιστοποίηση υπερπαραμέτρων, κ.ά.

Συμπερασματικά, γίνεται αντιληπτό ότι η μηχανική εκμάθηση είναι ένα αξιοσημείωτο και αρκετά αποτελεσματικό εργαλείο για τη διαχείριση μαζικών δεδομένων στον κλάδο της υγείας. Κατά συνέπεια, θα ήταν απαραί-

τητο να συμπεριληφθεί σε ένα πρόγραμμα σπουδών για την κατάρτιση της επόμενης γενιάς επιστημόνων στους κλάδους της βιοϊατρικής πληροφορικής και επιστήμης δεδομένων.

Επεξεργασία φυσικής γλώσσας και εξόρυξη κειμένου

Στόχοι

Κάθε χρόνο εκατοντάδες χιλιάδες νέες επιστημονικές δημοσιεύσεις και άρθρα εκδίδονται και αποθηκεύονται σε διάφορα αποθετήρια βιβλιογραφίας όπως η PubMed. Παράλληλα, εκατοντάδες χιλιάδες γραπτά κείμενα και ιατρικές σημειώσεις για ασθενείς καταγράφονται κάθε χρόνο και ενσωματώνονται στον ΗΦΥ. Όλα αυτά τα γραπτά αρχεία έχουν τεράστια αξία και σημασία αφού αποτελούν πρωτογενή δεδομένα. Συνεπώς, υπάρχει ανάγκη για ενσωμάτωση όλων αυτών σε μια βάση δεδομένων και εναρμόνισής τους με άλλα σετ δεδομένων.

Αυτό μπορεί να επιτευχθεί μέσω της επεξεργασίας φυσικής γλώσσας (NLP) που αποτελεί ένα ακόμη υποπεδίο της τεχνητής νοημοσύνης. Ο στόχος αυτής της λειτουργίας είναι να αυτοματοποιεί την επιμέλεια των εγγράφων τόσο από την επιστημονική βιβλιογραφία όσο και από τις κλινικές σημειώσεις για να δώσει τελικά μια κατανόηση του περιεχομένου τους. Με άλλα λόγια, η μέθοδος αυτή επιτρέπει την αυτόματη εξαγωγή λέξεων-κλειδιών και φράσεων από έγγραφα με σκοπό το σχολιασμό του νοήματός τους. Το περιεχόμενο που εξάγεται μπορεί στη συνέχεια να μετατραπεί σε δομημένα δεδομένα που μπορούν να ενσωματωθούν με άλλα δεδομένα είτε σε μια βάση δεδομένων.

Η αυτοματοποίηση της εξαγωγής δεδομένων από γραπτά κείμενα δεν είναι μια εύκολη διαδικασία αλλά είναι ένας εξελισσόμενος και πολλά υποσχόμενος ερευνητικός τομέας. Η δυσκολία του έγκειται στο γεγονός ότι ο υπολογιστής πρέπει να είναι σωστά “εκπαιδευμένος” στο να αναγνωρίζει, να καταγράφει και να αναδεικνύει τις σχέσεις σε λέξεις, όρους και ονόματα οντοτήτων των γραπτών κειμένων.

Όσον αφορά την αυτοματοποίηση στην αναγνώριση λέξεων και όρων, μια ενδιαφέρουσα εφαρμογή είναι η εύρεση και αναγνώριση των φαρμακευτικών σκευασμάτων που αναφέρονται σε μια σειρά κλινικών σημειώσεων. Στην συγκεκριμένη περίπτωση, ο υπολογιστής πρέπει να γνωρίζει τις διαφορετικές ονομασίες φαρμάκων, τις συντομογραφίες τους, τα ακρωνύμιά τους αλλά και την ανθρώπινη στενογραφία. Ένας άνθρωπος εγκέφαλος μπορεί να το κάνει αυτό αλλά ένας υπολο-

γιστής πρέπει να “εκπαιδευτεί” πάρα πολύ σκληρά για να το επιτύχει. Σύμφωνα με τον Hobbs, ο υπολογιστής μπορεί να επιτύχει σε ένα επίπεδο 60% τη σωστή και έγκυρη εξαγωγή δεδομένων χωρίς δυσκολία [40] αλλά η επίτευξη ενός μεγαλύτερου ποσοστού σε επίπεδο 90% απαιτεί την αναγνώριση σπάνιων όρων, συντμήσεων κλπ. Η βελτίωση του αλγορίθμου θα μπορούσε να αντιμετωπίσει το πρόβλημα αλλά αυτό χρειάζεται πολύ χρόνο και μελέτη.

- Στον πυρήνα της αναζήτησης μιας αυτοματοποιημένης διαδικασίας, υπάρχει μια μέθοδος γνωστή ως «αναγνώριση γνωστών οντοτήτων». Το συχνότερο πρόβλημα στην εφαρμογή της είναι η εύρεση αναφορών σε οντότητες όπως γονίδια, πρωτεΐνες, ασθένειες, φάρμακα, ονόματα οργανισμών, σε κείμενο φυσικής γλώσσας και οι αντιστοιχισί τους με τη θέση και τον τύπο τους. Η εύρεση και αντιστοίχιση των οντοτήτων είναι το δομικό χαρακτηριστικό για όλες τις εργασίες εξόρυξης κειμένου. Έχουν αναπτυχθεί διάφορα προγράμματα ανοιχτού κώδικα για το σκοπό αυτό, όπως το BANNER που είναι ένα σύστημα για την επισήμανση ονομάτων γονιδίων στη βιβλιογραφία [43].

Τα τελευταία χρόνια έχει παρατηρηθεί μια εξέλιξη στις μεθόδους αυτοματοποίησης για την εξαγωγή κειμένου στον τομέα της βιοϊατρικής αλλά οι προκλήσεις παραμένουν-. Αν και η ταχεία αλλαγή και ασυνέπεια των όρων και συντομογραφιών που χρησιμοποιούνται σε κείμενα βιοϊατρικής μειώνει σημαντικά την ακρίβεια της διαδικασίας, η συχνή χρήση μακροσκελών ονομάτων σε αυτά τα επιστημονικά κείμενα διευκολύνει τον υπολογιστή στο να προσδιορίσει αν υπάρχει όνομα οντότητας χωρίς απαραίτητα να καθορίσει τα ακριβή όριά του [41,42].

Τέλος, το επόμενο βήμα για την ολοκλήρωση της διαδικασίας εξαγωγής κειμένου, είναι η ανάδειξη των σχέσεων ανάμεσα σε δυο ή περισσότερες οντότητες. Έχουν σχεδιαστεί διάφορα συστήματα εξαγωγής σχέσεων γι’ αυτό το σκοπό. Παράδειγμα αυτών αποτελεί το Pharmspresso το οποίο βρίσκει τις αναφορές σε γονίδια και φάρμακα αλλά και τη σχέση τους σε επιστημονικές δημοσιεύσεις [44]. Ο στόχος του είναι να βοηθήσει τον υπολογιστή να καταλάβει τη σχέση γονιδίου-φαρμάκου όπως τη θέτει ο συγγραφέας του άρθρου. Η πληροφορία που προκύπτει έχει τεράστια σημασία για τη βιοϊατρική έρευνα, διότι χρησιμεύει για τον επαναπροσδιορισμό φαρμάκων και την εύρεση νέων ενδείξεων σε φαρμακευτικά σκευάσματα που είναι ήδη σε κυκλοφορία [45].

Ρόλος και εφαρμογές

Η γεφύρωση της επεξεργασίας φυσικής γλώσσας

(NLP) με τις υπόλοιπες μεθόδους που περιγράφηκαν παραπάνω είναι το τελικό βασικό στοιχείο για την ενσωμάτωση δεδομένων σε ένα βιολογικό πλαίσιο. Η διαδικασία αυτή είναι γνωστή ως κανονικοποίηση και σχετίζεται με την δημιουργία μιας μοναδικής ονομασίας ταυτοποίησης μιας οντότητας [46] ώστε να αναγνωρίζεται άμεσα σε όλα τα είδη κειμένου. Ένα απλό παράδειγμα είναι η ονομασία των γονιδίων και η εύρεσή τους. Όταν ένα γονίδιο αναφέρεται σε ένα κείμενο, θα πρέπει να είναι γραμμένο με κατάλληλη ονομασία ώστε να μπορεί να αντιστοιχιστεί με το αναγνωριστικό που έχει στην πλατφόρμα Entrez Gene. Η κανονικοποίηση σαν διεργασία αποτελείται από τέσσερα βασικά βήματα. Το πρώτο είναι επιλογή ενός λεξικού στο οποίο θα αντιστοιχίζονται οι αναφορές του κειμένου. Το δεύτερο είναι εύρεση και η ταυτοποίηση των αναφορών του κειμένου που μας ενδιαφέρουν. Αυτό ίσως να περιλαμβάνει και τη διαχείριση των προθεμάτων, επιθεμάτων και λίστες των οντοτήτων. Το τρίτο είναι αντιστοίχιση των αναφορών του κειμένου με αυτές που περιλαμβάνει το λεξικό ενώ το τελευταίο περιλαμβάνει την μετα-επεξεργασία των αναφορών με σκοπό να απομακρυνθούν τα ψευδώς θετικά αποτελέσματα λόγω ασάφειας.

Αν και φαίνεται ότι τα βήματα είναι καλά μελετημένα και διορθώνουν τυχόν λάθη ή ασάφειες, στα βιοϊατρικά και κλινικά κείμενα απαιτούνται επιπρόσθετα μέτρα για την μείωση των ψευδώς θετικών αποτελεσμάτων, όπως για παράδειγμα η εφαρμογή μιας μεθόδου χαρακτηρισμού του περιεχομένου του κειμένου στο οποίο αναφέρεται η οντότητα. Αυτή η μέθοδος ονομάζεται ισχυρισμός και βασίζεται στην ιδέα ότι οι οντότητες και τα περιεχόμενά τους περιγράφονται με διαφορετικό τρόπο στα βιοϊατρικά και στα κλινικά κείμενα [48,49]. Φυσικά, είναι δυνατόν το περιεχόμενο μιας οντότητας να είναι γενικό ή ειδικό αναλόγως το κείμενο. Για παράδειγμα, σε ένα κλινικό κείμενο τα συμπτώματα αναφέρονται ως επιβεβαιωμένα ή μη.

Συνοψίζοντας, η αυτοματοποίηση της εξόρυξης κειμένου αλλά και της επεξεργασίας φυσικής γλώσσας παίζει καθοριστικό ρόλο στη σωστή ενσωμάτωση των στοιχείων σε βάσεις δεδομένων καθώς επίσης και στα συστήματα εξαγωγής πληροφοριών και επεξεργασίας βάσεων δεδομένων. Η ακριβής αναπαράσταση των οντοτήτων, οι σχέσεις τους και οι ισχυρισμοί τους μπορεί να έχουν κρίσιμες συνέπειες για την κατανόηση του τρόπου με τον οποίο εξηγούνται τα βιολογικά μονοπάτια για τα κλινικά προφίλ των ασθενών και τελικά για την παραγωγή γνώσης από πληθυσμιακές μελέτες [49].

Μελλοντικές προοπτικές

Η πρόοδος στην εξατομικευμένη ιατρική και την ιατρική ακριβείας εξαρτάται από την ικανότητά μας να καθορίσουμε τα μοναδικά χαρακτηριστικά των ατόμων ή μικρών ομάδων ατόμων όπου απαιτούνται συγκεκριμένες στρατηγικές πρόληψης και θεραπείας ασθενειών. Πριν εφαρμοστεί κάτι τέτοιο στην κλινική πρακτική, είναι αναγκαίο να καθοριστούν ποια είναι τα σημαντικά χαρακτηριστικά για κάθε ασθένεια. Για να γίνει αυτό εφικτό, είναι απαραίτητη η μέτρηση ενός μεγάλου αριθμού εσωτερικών και εξωτερικών βιολογικών διεργασιών.

Αυτό έχει σαν αποτέλεσμα, τη συγκέντρωση όλο και μεγαλύτερου όγκου βιοϊατρικών δεδομένων η οποία απαιτεί την ανάπτυξη και εφαρμογή ειδικών μεθόδων τεχνολογίας και πληροφορικής για την αποθήκευση, διαχείριση, ανάλυση και ερμηνεία τους.

Στην παρούσα εργασία έχουμε κάνει μια σύντομη ανασκόπηση στους βασικότερους επιστημονικούς τομείς που είναι απαραίτητο να έχει εκπαιδευτεί η νέα γενιά επιστημόνων ώστε να είναι εξοικειωμένοι με την εξατομικευμένη ιατρική, δίνοντας μεγαλύτερη έμφαση στην τεχνητή νοημοσύνη συμπεριλαμβανομένης της μηχανικής εκμάθησης και την επεξεργασία φυσικής γλώσσας στα μαζικά δεδομένα. Όλες αυτές οι υπολογιστικές μέθοδοι διαδραματίζουν ένα κομβικό ρόλο στην εξαγωγή χρήσιμων πληροφοριών από περίπλοκα μοτίβα σε μαζικά δεδομένα ενώ απαιτείται συμπληρωματική εκπαίδευση στη διαχείριση και ενσωμάτωση δεδομένων, στη στατιστική και στην επιστήμη των δεδομένων.

Για τους λόγους αυτούς προτείνεται η δημιουργία ενός προγράμματος σπουδών για την εκπαίδευση της νέας γενιάς επιστημόνων οι οποίοι προέρχονται είτε από τον κλάδο της υγείας είτε της πληροφορικής στα μαζικά βιοϊατρικά δεδομένα. Αυτό το πρόγραμμα θα απευθύνεται σε φοιτητές που επιθυμούν να αποκτήσουν μια περαιτέρω εξειδίκευση και θα περιλαμβάνει θεματικές ενότητες ή κύκλους κατάρτισης σχετικά με:

α. Τις διαδικασίες και τις μεθόδους διαχείρισης και ενσωμάτωσης δεδομένων σε βάσεις, καθώς και τις γνώσεις πλήρους χειρισμού των βάσεων αυτών.

β. Τις βασικές έννοιες και τις μεθόδους στατιστικής ανάλυσης και των πιθανοτήτων, που απαιτούνται για την εξαγωγή αποτελεσμάτων και συμπερασμάτων.

γ. Την επιστήμη των δεδομένων, συμπεριλαμβανομένου του προγραμματισμού ηλεκτρονικών υπολογιστών και των μεθόδων για τη βελτίωση της αναπαραγωγιμότητας

δ. Την εισαγωγή στις θεμελιώδεις αρχές της τεχνητής νοημοσύνης

ε. Τη σχέση μηχανικής εκμάθησης και τεχνητής νοημοσύνης για την επεξεργασία φυσικής γλώσσας σε μη δομημένα δεδομένα.

Τέλος, στις παραπάνω θεματικές ενότητες θα ήταν ωφέλιμο να συμπεριληφθεί και μια ενότητα η οποία θα αναφέρεται στην εξατομικευμένη ιατρική και στην ιατρική ακριβείας ώστε να δώσει κίνητρο για τη χρήση μαζικών δεδομένων στις βιοϊατρικές επιστήμες.

Οι παραπάνω θεματικές ενότητες θα μπορούσε να προσφέρονται ως επίσημα μαθήματα στο πλαίσιο ενός μεταπτυχιακού προγράμματος είτε δια ζώσης είτε ακόμη εξ αποστάσεως. Θα μπορούσε ακόμη να περιλαμβάνονται σε ένα εντατικό βραχυχρόνιο πρόγραμμα κατάρτισης (σεμινάριο) το οποίο θα προσέφερε μια θεμελιώδη κατανόηση των θεματικών εννοιών και στοιχειώδη πρακτική εμπειρία. Αδιαμφισβήτητα, η βέλτιστη πρόταση είναι η δημιουργία μεταπτυχιακού κύκλου μαθημάτων με δια ζώσης διδασκαλία σε φυσική αίθουσα, ώστε αφενός να μπορεί να καλυφθεί σε βάθος η ύλη σε όλες οι ενότητες αλλά και να υπάρχει άμεση αλληλεπίδραση φοιτητών και εκπαιδευτών.

Επιπλέον, για την πλήρη κάλυψη των θεματικών εννοιών και την καλύτερη κατάρτιση των φοιτητών, προτείνεται η δημιουργία ενός προγράμματος σπουδών που θα περιλαμβάνει μαθήματα που θα συμπληρώνει το ένα το άλλο, μέσω ορολογίας και παραδειγμάτων. Ακόμη η ύπαρξη μιας ολοκληρωμένης υπολογιστικής πλατφόρμας όπου οι φοιτητές θα μπορούν να εργαστούν με μαζικά δεδομένα, να πραγματοποιούν αναλύσεις, να αξιολογούν τα αποτελέσματα και να σχεδιάζουν εργαλεία υποστήριξης εικονικών κλινικών αποφάσεων για εξατομικευμένη φροντίδα θα ήταν ιδανική.

Όλα τα παραπάνω θα συνέβαλαν στην εμπέδωση της ύλης, στην εξοικείωση όλων των φοιτητών με τέτοιου τύπου δεδομένα αλλά και θα οδηγούσαν, πιθανώς, στο σχεδιασμό νέων εικονικών εργαλείων και προγραμμάτων για τη λήψη κλινικών αποφάσεων στην εξατομικευμένη ιατρική. Οι φοιτητές θα έχουν ελεύθερη πρόσβαση σε όλα τα λογισμικά προγράμματα που απαιτούνται όπως για παράδειγμα το OpenMRS [50], ένα πρόγραμμα διαχείρισης δεδομένων ασθενών. Οι εργασίες τους μπορεί να είναι διαθέσιμες αρχικά σε ένα τοπικό επίπεδο και στη συνέχεια να ενσωματωθούν σε μεγαλύτερες βάσεις δεδομένων συμβάλλοντας έτσι στην επιστημονική κοινότητα. Η πρόταση αυτή είναι αρκετά προχωρημένη και απαιτεί την αναθεώρηση πολλών ζητημάτων αλλά η υλοποίηση της δεν είναι ανέφικτη.

Οι συμμετέχοντες στο παραπάνω πρόγραμμα θα αποκτήσουν ένα επιπρόσθετο επαγγελματικό εφόδιο στον

τομέα της βιοϊατρικής επιστήμης που θα τους δώσει ένα ανταγωνιστικό πλεονέκτημα στην αγορά εργασίας. Πριν από μερικά χρόνια οι ειδικοί των υπολογιστών, οι επιστήμονες δεδομένων, και οι στατιστικολόγοι συμμετείχαν ως σύμβουλοι ή συνεργάτες σε έργα όπου ήταν απαραίτητη η διαχείριση και ανάλυση δεδομένων με άμεση συμμετοχή στο σχεδιασμό της μελέτης.

Σήμερα, είναι ανάγκη να ισχυροποιηθεί η θέση πολλών νέων επιστημόνων οι οποίοι θα μπορούν να διατυπώνουν επιστημονικά ερωτήματα και να πραγματοποιούν τη δική τους έρευνα χωρίς να εμπλέκονται σ' αυτή τη διαδικασία άλλοι αναλυτές δεδομένων [51]. Ο συνδυασμός των σπουδών τους και των δεξιοτήτων που θα έχουν αναπτύξει κατά τη διάρκεια αυτών είναι ιδιαίτερα χρήσιμη στη σημερινή εποχή. Αυτή η νέα προσέγγιση στη βιοϊατρική έρευνα έχει ονομαστεί «σκέψη χωρίς όρια».

Συμπεράσματα

Ζούμε πλέον στην εποχή των μαζικών βιοϊατρικών δεδομένων. Κάθε μέρα ο αριθμός των βιοϊατρικών δεδομένων όλο και αυξάνεται αφού αποκτούμε νέες πληροφορίες για τον οργανισμό μας. Η συνεχής αύξηση των διαθέσιμων βιοϊατρικών δεδομένων αποτελεί μια κινητήρια δύναμη για την ενίσχυση του ρόλου και των εφαρμογών της εξατομικευμένης ιατρικής στην περίθαλψη. Για να εδραιωθεί όμως, αυτή η προσέγγιση περιθάλψης απαιτείται κατάλληλα εκπαιδευμένο και καταρτισμένο επιστημονικό και ερευνητικό προσωπικό.

Σε αυτήν την εργασία, προτείνεται η δημιουργία ενός προγράμματος κατάρτισης νέων επιστημόνων από τους κλάδους της πληροφορικής και των βιοϊατρικών επιστημών ώστε να αποκτήσουν κάποιες βασικές αλλά σημαντικές γνώσεις σε υπολογιστικές εφαρμογές που αξιοποιούνται στην εξαγωγή και διαχείριση μαζικών βιοϊατρικών δεδομένων. Καθώς οι κλάδοι των εφαρμοσμένων μαθηματικών και της επιστήμης των υπολογιστών συνεχίζουν να αλληλεπιδρούν, στο μέλλον αυτοί θα συγχωνευθούν σε ένα μόνο πεδίο, που ήδη μερικοί το αποκαλούν στατιστική μάθηση [18]. Είναι ευθύνη της ακαδημαϊκής κοινότητας να εκπαιδεύει και να προετοιμάζει τη νέα γενιά επιστημόνων οι οποίοι θα οδηγήσουν σ' αυτή την αλλαγή, διδάσκοντάς τους τη χρήση σύγχρονων εργαλείων και μεθόδων της στατιστικής ανάλυσης με τρόπους που να την ενθαρρύνουν. Η συμμετοχή σε ένα τέτοιο πρόγραμμα θα συνέβαλε στην επαγγελματική και ακαδημαϊκή ανέλιξη του νέου επιστήμονα ενώ θα εδραίωνε ακόμη περισσότερο το ρόλο της εξατομικευμένης ιατρικής.

Συνοψίζοντας, η χρήση βιοϊατρικών δεδομένων σε

συνδυασμό με την εκπαίδευση επιστημόνων πληροφορικής αλλά και επαγγελματιών υγείας, όπως περιεγράφηκε σε αυτήν την εργασία θα αποφέρει

σημαντικά αποτελέσματα στην εξέλιξη και επίδραση της εξατομικευμένης ιατρικής στον τομέα της υγείας. ●

ABSTRACT

Personalized Medicine and big biomedical data

Margarita-Ioanna Koufaki¹, Ioannis G. Hatzis², George P. Patrinos^{1*}

¹University of Patras School of Health Sciences, Department of Pharmacy, Patras, Greece

²Public Institute of Vocational Training of Messolonghi, Messolonghi, Greece

Personalized medicine depends on our ability to measure and process biological and environmental information about patients so as to enact individually optimized treatment. Much of this data is being stored in electronic health records yielding big data that pose challenges for management and analysis. We review here several areas of knowledge that are necessary for next-generation scientists to fully realize the potential of biomedical

big data. We begin with an overview of big data and its storage and management. We then review statistics and data science, the very foundations for artificial intelligence, machine learning, and natural language processing needed to develop predictive models for clinical decision making. We conclude with some specific training recommendations for preparing next-generation scientists for biomedical big data.

KEY WORDS: Personalized Medicine, biomedical big data, artificial intelligence, machine learning, electronic healthcare record

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Collins FS, Varmus H. A new initiative on precision medicine. *N. Engl. J. Med.* 372(9), 793–795 (2015).
- Council NR. Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease [Internet]. Available from: <https://www.nap.edu/catalog/13284/toward-precision-medicine-building-a-knowledge-network-for-biomedical-research>.
- Nimmegern E, Benediktsson I, Norstedt I. Personalized Medicine in Europe. *Clin Transl Sci.* 10(2), 61–63 (2017).
- Erlewyn-Lajeunesse M, Brathwaite N, Lucas JSA, Warner JO. Recommendations for the administration of influenza vaccine in children allergic to egg. *BMJ.* 339, b3680 (2009).
- Tran T, Luo W, Phung D, et al. A framework for feature extraction from hospital medical data with applications in risk prediction. *BMC Bioinformatics.* 15(1), 425 (2014).
- Albers DJ, Elhadad N, Tabak E, Perotte A, Hripcsak G. Dynamical Phenotyping: Using Temporal Analysis of Clinically Collected Physiologic Data to Stratify Populations. *PLOS ONE.* 9(6), e96443 (2014).
- Westphal JF. Macrolide – induced clinically relevant drug interactions with cytochrome P-450A (CYP) 3A4: an update focused on clarithromycin, azithromycin and dirithromycin. *British Journal of Clinical Pharmacology.* 50(4), 285–295 (2000).
- Phillips HS, Kharbanda S, Chen R, et al. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell.* 9(3), 157–173 (2006).
- Warner JL, Denny JC, Kreda DA, Alterovitz G. Seeing the forest through the trees: uncovering phe-

ΒΙΒΛΙΟΓΡΑΦΙΑ

- conomic complexity through interactive network visualization. *J Am Med Inform Assoc.* 22(2), 324–329 (2015).
10. Huang X, Jennings SF, Bruce B, et al. Big data - a 21st century science Maginot Line? No-boundary thinking: shifting from the big data paradigm. *BioData Min.* 8, 7 (2015).
 11. Angles R, Gutierrez C. Survey of Graph Database Models. *ACM Comput. Surv.* 40(1), 1:1–1:39 (2008).
 12. Robinson I, Webber J, Eifrem E. *Graph Databases.* 1 edition. O'Reilly Media, Beijing ; Sebastopol, CA.
 13. Harron K, Dibben C, Boyd J, et al. Challenges in administrative data linkage for research. *Big Data Soc.* 4(2), 2053951717745678 (2017).
 14. Mamun A-A, Aseltine R, Rajasekaran S. Efficient Record Linkage Algorithms Using Complete Linkage Clustering. *PLoS ONE.* 11(4), e0154446 (2016).
 15. Kim D, Joung J-G, Sohn K-A, et al. Knowledge boosting: a graph-based integration approach with multi-omics data and genomic knowledge for cancer clinical outcome prediction. *J Am Med Inform Assoc.* 22(1), 109–120 (2015).
 16. Beaulieu-Jones BK, Moore JH. MISSING DATA IMPUTATION IN THE ELECTRONIC HEALTH RECORD USING DEEPLY LEARNED AUTOENCODERS. *Pac Symp Biocomput.* 22, 207–218 (2017).
 17. Kim D, Li R, Dudek SM, Ritchie MD. Predicting censored survival data based on the interactions between meta-dimensional omics data in breast cancer. *J Biomed Inform.* 56, 220–228 (2015).
 18. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition* [Internet]. 2nd ed. Springer-Verlag, New York Available from: //www.springer.com/us/book/9780387848570.
 19. Greene CS, Garmire LX, Gilbert JA, Ritchie MD, Hunter LE. Celebrating parasites. *Nat. Genet.* 49(4), 483–484 (2017).
 20. Huber W, Carey VJ, Gentleman R, et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods.* 12(2), 115–121 (2015).
 21. Peng RD. Reproducible research in computational science. *Science.* 334(6060), 1226–1227 (2011).
 22. Wickham H, Grolemund G. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data.* 1 edition. O'Reilly Media, Sebastopol, CA.
 23. Beaulieu-Jones BK, Greene CS. Reproducibility of computational workflows is automated using continuous analysis. *Nat. Biotechnol.* 35(4), 342–346 (2017).
 24. Cole BS, Moore JH. Eleven quick tips for architecting biomedical informatics workflows with cloud computing. *PLoS Comput. Biol.* 14(3), e1005994 (2018).
 25. Russell S, Norvig P. *Artificial Intelligence: A Modern Approach.* 3rd ed. Prentice Hall Press, Upper Saddle River, NJ, USA.
 26. Ertel W. *Introduction to Artificial Intelligence* [Internet]. 2nd ed. Springer International Publishing Available from: //www.springer.com/us/book/9783319584867.
 27. Goodfellow I, Bengio Y, Courville A. *Deep Learning.* The MIT Press.
 28. Murphy KP. *Machine Learning: A Probabilistic Perspective.* The MIT Press.
 29. Baldi P. Deep Learning in Biomedical Data Science. *Annu. Rev. Biomed. Data Sci.* 1(1), 181–205 (2018).
 30. Cao C, Liu F, Tan H, et al. Deep Learning and Its Applications in Biomedicine. *Genomics Proteomics Bioinformatics.* 16(1), 17–32 (2018).
 31. Ching T, Himmelstein DS, Beaulieu-Jones BK, et al. Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface.* 15(141) (2018).
 32. Ravi D, Wong C, Deligianni F, et al. Deep Learning for Health Informatics. *IEEE J Biomed Health Inform.* 21(1), 4–21 (2017).
 33. Wainberg M, Merico D, DeLong A, Frey BJ. Deep learning in biomedicine. *Nat. Biotechnol.* 36(9), 829–838 (2018).
 34. Olson RS, La Cava W, Orzechowski P, Urbanowicz RJ, Moore JH. PMLB: a large benchmark suite for machine learning evaluation and comparison. *BioData Min.* 10, 36 (2017).
 35. Olson RS, Cava WL, Mustahsan Z, Varik A, Moore JH. Data-driven advice for applying machine learning to bioinformatics problems. *Pac Symp Biocomput.* 23, 192–203 (2018).
 36. Nantasenamat C, Worachartcheewan A, Jamsak S, et al. AutoWeka: toward an automated data

ΒΙΒΛΙΟΓΡΑΦΙΑ

- mining software for QSAR and QSPR studies. *Methods Mol. Biol.* 1260, 119–147 (2015).
37. Feuer M, Klein A, Eggensperger K, Springenberg J, Blum M, Hutter F. Efficient and Robust Automated Machine Learning [Internet]. In: *Advances in Neural Information Processing Systems 28*. Cortes C, Lawrence ND, Lee DD, Sugiyama M, Garnett R (Eds.). . Curran Associates, Inc., 2962–2970 (2015) [cited 2018 Mar 2]. Available from: <http://papers.nips.cc/paper/5872-efficient-and-robust-automated-machine-learning.pdf>.
 38. Olson RS, Bartley N, Urbanowicz RJ, Moore JH. Evaluation of a Tree-based Pipeline Optimization Tool for Automating Data Science [Internet]. In: *Proceedings of the Genetic and Evolutionary Computation Conference 2016*. ACM, New York, NY, USA, 485–492 (2016) [cited 2018 Mar 2]. Available from: <http://doi.acm.org/10.1145/2908812.2908918>.
 39. Olson RS, Sipper M, Cava WL, et al. A System for Accessible Artificial Intelligence. In: *Genetic Programming Theory and Practice XV*. Banzhaf W, Olson RS, Tozier W, Riolo R (Eds.). . Springer International Publishing, 121–134 (2018).
 40. Hobbs JR, Appelt D, Bear J, et al. FASTUS: A Cascaded Finite-State Transducer for Extracting Information from Natural-Language Text. *arXiv:cmp-lg/9705013* [Internet]. (1997). Available from: <http://arxiv.org/abs/cmp-lg/9705013>.
 41. Leser U, Hakenberg J. What makes a gene name? Named entity recognition in the biomedical literature. *Brief. Bioinformatics.* 6(4), 357–369 (2005).
 42. Yeh A, Morgan A, Colosimo M, Hirschman L. BioCreAtIvE task 1A: gene mention finding evaluation. *BMC Bioinformatics.* 6 Suppl 1, S2 (2005).
 43. Leaman R, Gonzalez G. BANNER: an executable survey of advances in biomedical named entity recognition. *Pac Symp Biocomput.* , 652–663 (2008).
 44. Garten Y, Altman RB. Pharmspresso: a text mining tool for extraction of pharmacogenomic concepts and relationships from full text. *BMC Bioinformatics.* 10 Suppl 2, S6 (2009).
 45. Yang H-T, Ju J-H, Wong Y-T, Shmulevich I, Chiang J-H. Literature-based discovery of new candidates for drug repurposing. *Brief. Bioinformatics.* 18(3), 488–497 (2017).
 46. Crim J, McDonald R, Pereira F. Automatically annotating documents with normalized gene lists. *BMC Bioinformatics.* 6 Suppl 1, S13 (2005).
 47. Chen L, Liu H, Friedman C. Gene name ambiguity of eukaryotic nomenclatures. *Bioinformatics.* 21(2), 248–256 (2005).
 48. Miwa M, Thompson P, McNaught J, Kell DB, Ananiadou S. Extracting semantically enriched events from biomedical literature. *BMC Bioinformatics.* 13, 108 (2012).
 49. Velupillai S, Mowery D, South BR, Kvist M, Dalanis H. Recent Advances in Clinical Natural Language Processing in Support of Semantic Analysis. *Yearb Med Inform.* 10(1), 183–193 (2015).
 50. Wolfe BA, Mamlin BW, Biondich PG, et al. The OpenMRS system: collaborating toward an open source EMR for developing countries. *AMIA Annu Symp Proc.* , 1146 (2006).
 51. Huang X, Bruce B, Buchan A, et al. No-boundary thinking in bioinformatics research. *BioData Min.* 6(1), 19 (2013).

ΑΡΘΡΟ ΑΝΑΣΚΟΠΗΣΗΣ

Υπολογιστική ανάπτυξη μοντέλων αξιολόγησης φαρμακοθεραπειών με χρήση αλγορίθμων μηχανικής εκμάθησης και ομικών δεδομένων

Ελένη Μπαρμπάνη, Μαρία Κορομηνά*, Γεώργιος Π. Πατρινός

Πανεπιστήμιο Πατρών, Σχολή Επιστημών Υγείας, Τμήμα Φαρμακευτικής, Εργαστήριο Φαρμακογονιδιωματικής και Εξατομικευμένης Θεραπείας, Πάτρα

ΠΕΡΙΛΗΨΗ

Στην εποχή της ψηφιακής υγείας και της τεχνητής νοημοσύνης, επιτακτική καθίσταται η ανάγκη της φαρμακευτικής βιομηχανίας για καινοτόμες και μετασχηματιστικές τεχνολογίες ανάπτυξης φαρμάκων. Οι αλγόριθμοι τεχνητής νοημοσύνης και μηχανικής εκμάθησης, αν και αργά, έχουν αναμφισβήτητα αρχίσει να επιφέρουν επανάσταση στον τομέα ανάπτυξης φαρμάκων τα τελευταία πέντε έτη. Στη συγκεκριμένη ανασκόπηση, περιγράφουμε τις πιο συχνά χρησιμοποιούμενες προσεγγίσεις μηχανικής εκμάθησης στα εργαλεία ανάπτυξης φαρμάκων και τις βάσεις ομικών δεδομένων. Αναλύουμε τις νέες υπολογιστικές προ-

σεγγίσεις στο πεδίο της ανακάλυψης φαρμάκων στο πλαίσιο της ανάπτυξης και επαναστόχευσης φαρμάκων, αλλά και τις συνέργειες μεταξύ των ομικών επιστημών, της τεχνητής νοημοσύνης και της μηχανικής εκμάθησης. Επιπρόσθετα, παραθέτουμε μια μελλοντική προοπτική σχετικά με τους τρόπους με τους οποίους οι προσεγγίσεις της μηχανικής εκμάθησης θα είναι εφικτό να εφαρμοστούν προκειμένου όχι απλώς να επισπεύσουν την ανακάλυψη φαρμάκων αλλά και να ενισχύσουν την Ιατρική Ακριβείας με κύριο γνώμονα το όφελος των ασθενών και της δημόσιας υγείας.

ΛΕΞΕΙΣ ΚΥΡΙΑΡΧΙΑΣ: τεχνητή νοημοσύνη, μηχανική μάθηση, ανάπτυξη φαρμάκων, επανατοποθέτηση φαρμάκων, φαρμακευτική βιομηχανία

* Αντεπιστέλλων Συγγραφέας

Μαρία Κορομηνά, Εργαστήριο Φαρμακογονιδιωματικής και Εξατομικευμένης Θεραπείας, Τμήμα Φαρμακευτικής, Σχολή Επιστημών Υγείας, Πανεπιστήμιο Πατρών, Πάτρα, Email: m.koromina@upnet.gr

Εισαγωγή

Η φαρμακευτική βιομηχανία, ειδικά στα πλαίσια της έρευνας και ανάπτυξης φαρμάκων, απαιτεί τη χρήση νέων τεχνολογιών προσαρμοσμένων στη σύγχρονη εποχή της ψηφιακής υγείας και της τεχνητής νοημοσύνης [1].

Μια σύντομη περιγραφή της ιστορίας της τεχνητής νοημοσύνης πραγματοποιήθηκε πρόσφατα από τον Garvey (2018) και διέκρινε τρία πρότυπα: το «GOFAI» (1950-60), το «Expert Systems» (τέλη 1970-80) και τη «machine learning» (2010-σήμερα). Το πρώτο πρότυπο GOFAI (Good-Old-Fashioned Artificial Intelligence), συντομογραφία της «παλιομοδίτικης τεχνητής νοημοσύνης», επικεντρώθηκε στη δημιουργία συστημάτων κοινής λογικής και οδήγησε στην ανάπτυξη θεμελιωδών τεχνικών. Το πρότυπο των «Expert systems» περιόρισε την απήχηση και κατανόησή του από το ευρύ κοινό σε ανθρώπους με εξειδίκευση σε συγκεκριμένους τομείς όπως η χημεία, η ιατρική και προσπάθησε να αναπαράγει τις γνώσεις και τις διαδικασίες λήψης των αποφάσεών τους. Το γεγονός αυτό οδήγησε στο εξειδικευμένο σύστημα τεχνητής νοημοσύνης, το MYCIN και τελικά σε πιο γνωστό λογισμικό όπως το TurboTax. Ενώ απέφεραν κάποια πρακτικά αλλά περιορισμένα αποτελέσματα, και τα δύο αυτά πρότυπα τεχνητής νοημοσύνης απέτυχαν να δημιουργήσουν τις «σκεπτόμενες μηχανές» που είχαν οραματιστεί οι εφευρέτες και πρωτοπόροι. Το τρέχον πρότυπο, η μηχανική εκμάθηση / «machine learning», έχει ξεπεράσει μερικούς από τους φραγμούς σχετικά με την ανταπόκριση και εφαρμογή της στον πραγματικό κόσμο, χάρη στην ολοένα και αυξανόμενη πληθώρα ανθρώπινων δεδομένων, στην τεράστια αύξηση της υπολογιστικής ισχύος και στην αναβίωση τόσο νευρωνικών δικτύων όσο και άλλων αλγορίθμων μηχανικής εκμάθησης. Αυτοί οι αλγόριθμοι εκμάθησης μπορούν να «εκπαιδευτούν» με σκοπό να αναγνωρίζουν και να επεκτείνουν πρότυπα προερχόμενα από ανθρώπινα δεδομένα και, επομένως, δεν απαιτούν ιδιαίτερα απαιτητικές γνώσεις προγραμματισμού [2].

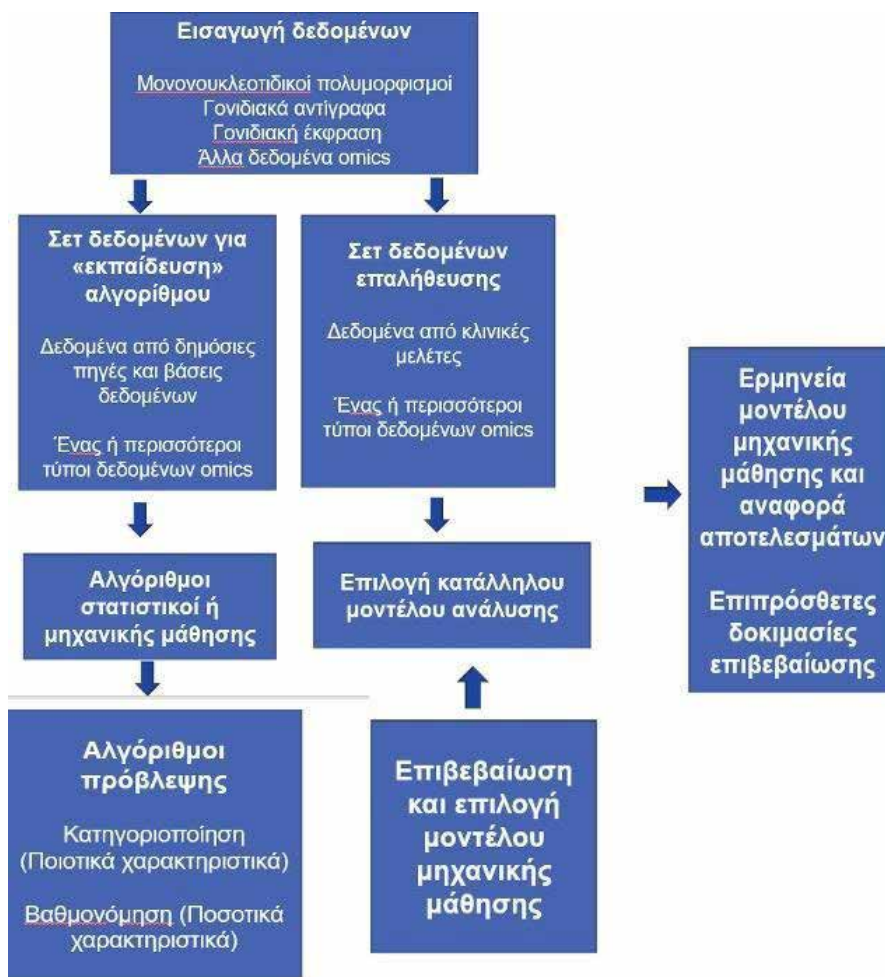
Η τεχνητή νοημοσύνη και οι υπολογιστικοί αλγόριθμοι απαιτούν επανεξέταση των διαδικασιών ανάπτυξης και ανακάλυψης φαρμάκων, λαμβάνοντας παράλληλα υπόψη τις προοπτικές και τις προκλήσεις που τις συνοδεύουν. Υπάρχει η δυνατότητα ωστόσο, να αξιοποιηθούν υπό το πρίσμα της συναρπαστικής εμπειρίας των ομικών επιστημών, ώστε να συμπεριληφθεί μια ποικιλία μοριακών διαφορών μεταξύ ατόμων και πληθυσμών, η οποία μπορεί να είναι ενδεικτική της ανταπόκρισης στη θεραπεία ή των παρενεργειών, και συνεπώς να καθοδη-

γήσει τα εργαλεία ανακάλυψης και ανάπτυξης φαρμάκων. Αυτές οι μοριακές διαφορές συχνά περιλαμβάνουν διάφορους τύπους γονιδιωματικών παραλλαγών, όπως σημειακές μεταλλάξεις, διαγραφές, παρεμβολές και μετατοπίσεις γονιδιακών αλληλουχιών, οι οποίες μπορεί να αποτελέσουν ή να υποδείξουν άμεσους μοριακούς στόχους για ανάπτυξη φαρμακευτικών θεραπευτικών προσεγγίσεων. Τέτοια παραδείγματα περιλαμβάνουν τις κλινικά ενεργές παραλλαγές, που εντοπίστηκαν στα γονίδια EGFR και ALK και μπορεί να αποτελούν στόχο των φαρμάκων της κατηγορίας αναστολέων κινάσης [3]. Οι κλινικές πληροφορίες σχετικά με τους διαφορετικούς τύπους γενετικών παραλλαγών μπορούν να εξαχθούν από κάποιες βάσεις δεδομένων όπως η ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>), η COSMIC (<https://cancer.sanger.ac.uk/cosmic>) και η OMIM (<https://www.omim.org>).

Παρά την ραγδαία εξέλιξη των αλγορίθμων in silico πρόβλεψης, η υπολογιστική πρόβλεψη της φαρμακευτικής απόκρισης ειδικά σε σύνθετες και πολυπαραγοντικές ασθένειες, παραμένει μία πρόκληση. Ο μεγάλος όγκος καθώς και η ετερογένεια των δεδομένων συχνά αποτελούν τροχοπέδη στη βελτίωση της προβλεπτικής ισχύος των υπολογιστικών μοντέλων. Άλλοι κρίσιμης σημασίας προβληματισμοί κατά την ανάπτυξή τους είναι: η επιλογή των κατάλληλων συνόλων δεδομένων με στόχο την εκπαίδευση και τις δοκιμές των μοντέλων, η επιλογή των πιο ενδεδειγμένων υπολογιστικών προσεγγίσεων για εφαρμογή, καθώς και η επικύρωση και αξιολόγηση των συγκεκριμένων υπολογιστικών μοντέλων.

Η χρήση γονιδιωματικών πληροφοριών εξαγόμενων από εφαρμογές της Αλληλούχισης Επόμενης Γενιάς (Next-Generation Sequencing - NGS) χιλιάδων ασθενών σε συνδυασμό με κλινικές πληροφορίες που αφορούν τα χαρακτηριστικά γνωρίσματα της νόσου και τα θεραπευτικά αποτελέσματα, μπορεί ενδεχομένως να οδηγήσει στον εντοπισμό δεικτών σχετικών με την ανταπόκριση στη θεραπεία μέσω μιας διαδικασίας μοντελοποίησης πολλαπλών παραλλαγών. Για το σκοπό αυτό, οι εποπτευόμενοι (supervised) αλγόριθμοι μηχανικής εκμάθησης επιτρέπουν την πρόβλεψη πολλαπλών δεικτών που αφορούν τη φαρμακευτική απόκριση εφαρμόζοντας πολυ-ομικές και πολλαπλών καθηκόντων (multi-task) μεθοδολογίες εκμάθησης οι οποίες εξαγουν πληροφορίες από δείγματα ασθενών καθώς και από ομοιότητες μεταξύ των φαρμάκων [4].

Εδώ, περιγράφουμε τις πιο συχνά χρησιμοποιούμενες υπολογιστικές τεχνικές, που χρησιμοποιούν ομικά



Εικόνα 1: Διάγραμμα ροής της ανάπτυξης υπολογιστικών μοντέλων πρόβλεψης της φαρμακευτικής απόκρισης.

δεδομένα μεγάλης κλίμακας στις περισσότερες των περιπτώσεων. Παρέχουμε επίσης λεπτομέρειες σχετικά με τις μεθόδους της μηχανικής εκμάθησης που χρησιμοποιούνται κατά τη διαδικασία ανάπτυξης και διαλογής φαρμάκων. Τέλος, παρατίθεται και μία περιγραφή των διαθέσιμων προσεγγίσεων μηχανικής εκμάθησης που αφορούν την επαναστόχευση φαρμάκων (Drug repositioning).

Επισκόπηση της στρατηγικής σχεδιασμού υπολογιστικών μοντέλων πρόβλεψης

Στις περισσότερες περιπτώσεις, η ανάπτυξη υπολογιστικών μοντέλων πρόβλεψης της φαρμακευτικής απόκρισης βασίζεται σε τέσσερα διαφορετικά στάδια. Κατά το πρώτο στάδιο, αρχικά επιλέγονται τα κατάλληλα σύνολα δεδομένων και προ-επεξεργάζονται μέσω επιλογής των σχετικών υποσυνόλων δεδομένων. Στη συνέχεια πραγματοποιείται “κανονικοποίησή” των δεδομένων

αυτών και φιλτράρισμα, εξάλειψη των δεδομένων που αντιστοιχούν σε θόρυβο καθώς και εκείνων που περιέχουν πληροφορίες μη σχετικές με την εκάστοτε μελέτη (Εικόνα 1) [5]. Αυτά τα σύνολα δεδομένων μπορεί να αποτελούνται από μονονουκλεοτιδικούς πολυμορφισμούς, παραλλαγές του πλήθους των αντιγράφων των γονιδίων (Copy Number Variants, CNVs) και από δεδομένα που προκύπτουν από τη γονιδιακή έκφραση. Ενδιαφέρον παρουσιάζει το εύρημα συγκριτικών μελετών ανάλυσης ότι τα δεδομένα γονιδιακής έκφρασης έχουν και την πιο ισχυρή προγνωστική αξία, ενώ τα υπολογιστικά μοντέλα αύξησαν οριακά την ακρίβεια πρόβλεψης της φαρμακευτικής απόκρισης [6].

Η πιθανότητα εμφάνισης μη αναμενόμενων αποκρίσεων κατά το πλαίσιο των θεραπειών είναι απαραίτητο να ληφθεί υπόψη κατά την εφαρμογή αλγορίθμων με στόχο την εκτίμηση της ευαισθησίας σε ορισμένη φαρμακευτική αγωγή. Για το λόγο αυτό, τα σύγχρονα μοντέ-

| Πίνακας 1. | | | | | |
|------------------------|----------------------|-----------------------|----------------------|------------------------|------------------------|
| Δημόσια Βάση δεδομένων | NCI-DREAM7 | CCLC | NCI-60 | GDSC | TCGA |
| Καρκινικοί τύποι | Καρκίνος μαστού | 36 τύποι καρκίνου | 9 τύποι καρκίνου | 29 τύποι καρκίνου | 33 τύποι καρκίνου |
| Αριθμός δειγμάτων | 53 κυτταρικές σειρές | 947 κυτταρικές σειρές | 59 κυτταρικές σειρές | 1124 κυτταρικές σειρές | 11000 δείγματα ασθενών |

Πίνακας 1: Συνοπτική καταγραφή των δημοσίων βάσεων ποικίλων ομικών δεδομένων μαζί με τον αντίστοιχο καρκινικό τύπο και τον αριθμό δειγμάτων κάθε βάσης (CCLC: Cancer Cell Line Encyclopedia, DREAM7: Dialogue on Reverse Engineering Assessment and Methods, GDSC: Genomics of Drug Sensitivity in Cancer, NCI: National Cancer Institute, TCGA: The Cancer Genome Atlas).

λα ενσωματώνουν δεδομένα από ένα ευρύτερο φάσμα κυτταρικών σειρών ή/και κλινικών δειγμάτων ώστε να επιτευχθεί ακριβέστερη εκτίμηση της κλινικής ετερογένειας ή/και των αποκρίσεων στη θεραπεία [4].

Το δεύτερο στάδιο σχεδιασμού του αλγορίθμου περιλαμβάνει τη φάση εκπαίδευσης του μοντέλου που έχει επιλεγεί (**Εικόνα 1**). Μπορεί έτσι να εφαρμοστεί μια ποικιλία τεχνικών μηχανικής εκμάθησης ενώ με τα δεδομένα μπορούν να τροφοδοτηθούν διαφορετικά, εναλλακτικά μοντέλα ώστε να επιλεγεί το ακριβέστερο σε πρόβλεψη και να εξελιχθεί. Ο τύπος των δεδομένων καταχώρησης και τα ειδικά χαρακτηριστικά γνωρίσματα του ζητήματος πρόβλεψης της φαρμακευτικής απόκρισης απαιτούν προσεκτική εξέταση. Το τρίτο βήμα σχεδιασμού του υπολογιστικού μοντέλου πρόβλεψης αναφέρεται συχνά ως ανεξάρτητη αξιολόγηση. Κατά τη διάρκεια αυτού, διεξάγονται πολλαπλές δοκιμές σε ανεξάρτητα μεταξύ τους σύνολα δεδομένων από το επιλεγμένο και εξελιγμένο μοντέλο. Αυτό αποσκοπεί να επαληθεύσει ότι το υποψήφιο μοντέλο πρόβλεψης μπορεί με ακρίβεια να προβλέψει τις φαρμακευτικές αποκρίσεις πάνω σε νεοφανή, μη προηγούμενως μελετημένα δεδομένα που προέρχονται από διαφορετικά εργαστήρια και πλατφόρμες μέτρησης.

Τελικό στάδιο αποτελεί η εφαρμογή του μοντέλου σε δεδομένα που σε κλινικό επίπεδο μοιάζουν με τα χαρακτηριστικά της υπό εξέταση ασθένειας. Για παράδειγμα, μοντέλα που εκπαιδεύονται με τη χρήση δεδομένων προερχόμενων από κυτταρικές σειρές μπορούν να δοκιμαστούν σε στερεές ή υγρές βιοψίες ασθενών. Και πάλι, η ερμηνεία και η εφαρμογή των κατάλληλων μεθοδολογιών για την αποσαφήνιση των προβλέψεων καθίστανται κρίσιμες παράμετροι (**Εικόνα 1**).

Αξιολόγηση του μοντέλου πρόβλεψης

Η αξιολόγηση του μοντέλου αποτελεί κρίσιμο και αναπόσπαστο κομμάτι κατά τη διαδικασία δημιουργίας ενός ισχυρού και ακριβούς μοντέλου *in silico* πρόβλεψης, καθώς αυτό θα πρέπει να έχει ικανοποιητική απόδοση ακόμη, και ιδίως, σε δεδομένα που δεν έχει 'δει' ποτέ. Η αξιολόγηση του μοντέλου πραγματοποιείται αφού 'εκπαιδευτεί' ο αλγόριθμος με ένα σύνολο δεδομένων εκπαίδευσης (training set), ενώ έπεται η επαλήθευση του μοντέλου με τη χρήση ανεξάρτητων συνόλων δεδομένων (test set), ώστε να εκτιμηθεί το τελικό μοντέλο. Μεταξύ των πιο σημαντικών παραμέτρων απόδοσης που προορίζονται για εφαρμογή σε ζητήματα κατηγοριοποίησης δεδομένων είναι: (α) η συνολική ευαισθησία (Sensitivity) του μοντέλου, (β) η ακρίβεια (Precision), (γ) η ανάκληση (Recall), (δ) η περιοχή κάτω από την καμπύλη (Area Under Curve). Επιπλέον αξιολογούνται και οι καμπύλες ακριβείας-ανάκλησης (Precision-Recall curves) που αποτελούν παράγωγο μέγεθος των (β) και (γ).

Η συνήθης διαδικασία αξιολόγησης του προσαρμοζόμενου μοντέλου είναι γνωστή ως Διασταυρούμενη Επικύρωση (Cross-Validation, CV). Κατά τη διάρκεια αυτής, το αρχικό σύνολο δεδομένων χωρίζεται σε δύο ξεχωριστά υποσύνολα, εκ των οποίων το ένα θα χρησιμοποιηθεί για την εκπαίδευση του αλγορίθμου και το άλλο για τον έλεγχο της απόδοσης του μοντέλου. Τα πιο συχνά χρησιμοποιούμενα σχήματα διασταυρούμενης επικύρωσης είναι η 'K-fold' (K-Fold Cross Validation, KF-CV) και η 'Leave-one-out' (Leave-One-Out Cross Validation, LOO-CV) διασταυρούμενη επικύρωση. Κατά τη μέθοδο KF-CV, το αρχικό σύνολο δεδομένων κατανέμεται σε μέρη K, με τα μέρη K-1 να χρησιμοποιούνται για εκπαι-

δευση και το τελευταίο για δοκιμές επαλήθευσης του εκάστοτε μοντέλου. Η αρχή λειτουργίας της μεθόδου LOO-CV προϋποθέτει ότι ένα μεμονωμένο υποσύνολο από το αρχικό σύνολο δεδομένων θα περιλαμβάνει το σύνολο των εναπομείναντων δεδομένων και το οποίο θα προορίζεται για επαλήθευση. Στη συνέχεια, η διαδικασία επαναλαμβάνεται τόσες φορές όσα και τα εξεταζόμενα δεδομένα (data points) [7], [8].

Στην εικόνα 2 αναπαρίστανται συνοπτικά τα στάδια ανάπτυξης υπολογιστικών μοντέλων πρόβλεψης της φαρμακευτικής απόκρισης:

Δημόσιες πηγές ομικών δεδομένων για τη δημιουργία αλγορίθμων πρόβλεψης της απόκρισης σε φαρμακευτική θεραπεία

Όπως αναλύθηκε στην προηγούμενη ενότητα, η πλειοψηφία των μοντέλων πρόβλεψης της φαρμακευτικής απόκρισης εκπαιδεύεται με τη χρήση συνόλων δεδομένων που δημιουργούνται από διαφορετικά ερευνητικά προγράμματα. Παρόλο που αυτή η προσέγγιση μπορεί να έχει βιολογική ακρίβεια, υπόκειται σε ένα σύνολο περιορισμών, όπως είναι ο αριθμός των ασθενών που υπόκεινται σε ανάλυση ή οι ψευδείς αναγνώσεις ομικών και κλινικών δεδομένων τα οποία μπορούν ενδεχομένως να ενσωματωθούν στην ανάπτυξη μοντέλων. Μία εναλλακτική προσέγγιση, η οποία κερδίζει ολοένα και μεγαλύτερη αποδοχή, είναι η εκπαίδευση (ή/και δοκιμή) μοντέλων σε δεδομένα δημοσίων βάσεων δεδομένων προερχόμενα από μεγάλα ερευνητικά προγράμματα (**Πίνακας 1**).

Κλασικές προσεγγίσεις μηχανικής εκμάθησης στην πρόβλεψη της φαρμακευτικής απόκρισης

Γενικά, οι τεχνικές μηχανικής εκμάθησης κατηγοριοποιούνται στις εποπτευόμενες (supervised), οι οποίες λαμβάνουν υπόψη πληροφορίες σχετικά με τις κατηγορίες των δεδομένων εκπαίδευσης (training data) και στις μη εποπτευόμενες (unsupervised), οι οποίες στοχεύουν στη δημιουργία ομάδων μεταξύ των δεδομένων εκπαίδευσης. Όσον αφορά τις τελευταίες, η ένταξη των δεδομένων (data points) σε μία ομάδα γίνεται με τέτοιο τρόπο ώστε τα στοιχεία μιας ομάδας να είναι όσο το δυνατόν πιο όμοια μεταξύ τους και να διαφέρουν όσο το δυνατόν περισσότερο από εκείνα άλλων ομάδων. Συνήθως, ανάλυση που βασίζεται στις μη εποπτευόμενες μεθόδους προηγείται και είναι δυνατό να χρησιμοποιηθεί ώστε να αποκτήσουμε μία πρώτη εκτίμηση των δεδομένων [9], [10], [11].

Οι εποπτευόμενες μέθοδοι μηχανικής εκμάθησης περιλαμβάνουν ένα ευρύ φάσμα τεχνικών, το οποίο χωρίζεται σε δύο κύριες κατηγορίες. Η κύρια κατηγορία περιλαμβάνει τους αλγόριθμους κατηγοριοποίησης (classification algorithms) και παλινδρόμησης (regression-based algorithms) με τους πρώτους να χρησιμοποιούνται για τον προσδιορισμό της κατηγορίας ένταξης μιας νέας καταχώρησης, για παράδειγμα αν μια καρκινική κυτταρική σειρά αναμένεται να ανταποκριθεί με τον επιθυμητό ή με ανεπιθύμητο τρόπο μετά από τη χορήγηση ενός συγκεκριμένου φαρμάκου [12], [13]. Η δεύτερη κατηγορία τεχνικών αποσκοπεί στην εκτίμηση της αξίας μιας εξεταζόμενης μεταβλητής [14], [15]. Από τους πιο ευρέως χρησιμοποιούμενους αλγόριθμους κατηγοριοποίησης είναι οι επονομαζόμενοι ως Support Vector Machines (SVMs) και τα Random Forests [16], [17], [18], [19], [20]. Ωστόσο, βελτιωμένη προγνωστική απόδοση είναι δυνατό να επιτευχθεί υπό διάφορες συνθήκες [13] χρησιμοποιώντας μοντέλα που βασίζονται στη μέθοδο της παλινδρόμησης, όπως τα elastic net ή ridge regression [21].

Προσεγγίσεις μηχανικής εκμάθησης στην επαναστόχευση φαρμάκων

Η επαναστόχευση φαρμάκων (Drug repositioning) ορίζεται ως η διαδικασία επιλογής ενός ήδη γνωστού φαρμάκου για μια εναλλακτική φαρμακολογική χρήση. Λόγω της ραγδαίας ανάπτυξης του τομέα της βιοπληροφορικής και της ανάλυσης ομικών δεδομένων μεγάλης κλίμακας, η επαναστόχευση φαρμάκων έχει μειώσει σημαντικά τον απαιτούμενο χρόνο κατά τη διαδικασία ανάπτυξης φαρμάκων. Στις μέρες μας, οι ερευνητές χρειάζονται περίπου 1 με 2 χρόνια μόνο για να εντοπίσουν νέους πιθανούς στόχους φαρμάκων και περίπου 8 έτη ώστε να αναπτύξουν κατά προσέγγιση ένα επαναστοχευμένο φάρμακο (**Εικόνα 2**).

Συνοπτικά, οι προσεγγίσεις επαναστόχευσης φαρμάκων μπορούν να ταξινομηθούν σε τρεις κατηγορίες: προσεγγίσεις βάσει δικτύου (network based approaches), προσεγγίσεις εξόρυξης κειμένου (text-mining approaches) και σημασιολογικές προσεγγίσεις (semantic approaches). Οι μέθοδοι της πρώτης κατηγορίας (network-based methods) διαιρούνται περαιτέρω στις μεθόδους ομαδοποίησης (clustering) και διάδοσης (propagation). Οι πρώτες αναζητούν τυχόν συσχετίσεις μεταξύ φαρμάκων και νόσων ή πιθανών στόχων σε μικρότερες επιμέρους "συστάδες" ενός μεγαλύτερου δικτύου, ενώ οι τελευταίες βασίζονται σε πρωτότερα



Εικόνα 2: Διάγραμμα ροής της σταδιακής διαδικασίας που ακολουθείται κατά την επαναστόχευση φαρμάκων.

ανακτημένες συσχετίσεις μέσω των αντίστοιχα χρησιμοποιούμενων δικτύων.

Τέλος, περαιτέρω διάκριση των ανωτέρω μεθόδων μπορεί να γίνει με κριτήριο τη σύνθεση των δημιουργούμενων δικτύων. Πιο συγκεκριμένα, αυτά που σχηματίζονται χρησιμοποιώντας έναν τύπο πληροφοριών, όπως οι πρωτεϊνικές αλληλεπιδράσεις (Protein-Protein Interactions-PPIs) χαρακτηρίζονται ως ομοιογενείς, ενώ δίκτυα που ενσωματώνουν διάφορους τύπους δεδομένων, όπως αυτά που προέρχονται από αναλύσεις ομικών δεδομένων είναι γνωστά ως ετερογενή [22].

Αντίθετα, η μέθοδος εξόρυξης κειμένου (text-mining approach) εκμεταλλεύεται το τεράστιο εύρος της διαθέσιμης βιβλιογραφίας, το οποίο στη συνέχεια φιλτράρεται για τη διατήρηση μόνο εκείνων των πηγών δεδομένων που θα επιτρέψουν την εξαγωγή πληροφοριών για τους υπό μελέτη βιολογικούς όρους. Τέλος, στις σημασιολογικές προσεγγίσεις (semantic approaches), η δημιουργία δικτύων καθοδηγείται από προηγούμενες βιοϊατρικές γνώσεις, όπως εξάγονται από βάσεις δεδομένων και στη συνέχεια χρησιμοποιούνται αλγόριθμοι μηχανικής εκμάθησης για τον εντοπισμό νέων αλληλεπιδράσεων και σχέσεων, οι οποίες υπάρχουν σε αυτά τα δίκτυα [23].

Συζήτηση και μελλοντικές προοπτικές

Στις μέρες μας, όλα τα στάδια της ανακάλυψης και ανάπτυξης φαρμάκων, από την επιλογή και επικύρωση του στόχου μέχρι και τις κλινικές δοκιμές μπορούν να υπόκεινται και να επωφελούνται από την εφαρμογή αλγορίθμων και εργαλείων μηχανικής εκμάθησης. Με-

ταξύ των διαφορετικών προκλήσεων που μπορούν να αντιμετωπιστούν μέσω αυτών των μεθοδολογιών περιλαμβάνονται τόσο η απαιτητική διαδικασία αναγνώρισης νέων μοριακών στόχων όσο και η απόκτηση μιας βαθύτερης γνώσης τόσο των μηχανισμών της νόσου, αλλά και των πολύπλοκων και πολυπαραγοντικών φαινοτύπων πολλών ασθενειών [24],[25].

Οι προσεγγίσεις της τεχνητής νοημοσύνης (που συμπεριλαμβάνει και τη μηχανική εκμάθηση) μπορούν να ωφελήσουν ουσιαστικά κάθε βήμα της διαδικασίας σχεδιασμού και ανάπτυξης φαρμάκων. Ξεκινώντας από τη διαδικασία διαλογής (screening) χημικών ενώσεων μέσα από τεράστιες βιβλιοθήκες και συνεχίζοντας με την ανακάλυψη της πρόδρομης ένωσης και την ταυτοποίηση της ένωσης οδηγού [26], [27], έως και τη διερεύνηση των πιθανών τροποποιήσεων που θα οδηγήσουν σε ενώσεις με βελτιωμένα χαρακτηριστικά φαρμακοκινητικής, φαρμακοδυναμικής και τοξικολογίας. Επιπροσθέτως, βελτιστοποιείται και το στάδιο σχεδιασμού της συνθετικής τους πορείας συνήθως μέσω αναδρομικής σύνθεσης [28], [29]. Οι αλγόριθμοι μηχανικής εκμάθησης μπορούν να χρησιμοποιηθούν επίσης και για τον εντοπισμό νέων βιοδεικτών με στόχο τη βελτίωση της αποτελεσματικότητας των φαρμάκων, ενισχύοντας έτσι τον τομέα της Ιατρικής Ακριβείας [30].

Επιπλέον, μια πληθώρα μοντέλων μηχανικής εκμάθησης για την ανακάλυψη βιοδεικτών και για την πρόβλεψη της φαρμακευτικής ευαισθησίας υπόσχεται να βελτιώσει σημαντικά τα ποσοστά της κλινικής επιτυχίας, τόσο μέσω της συμβολής τους στην αποκάλυψη μοριακών μηχανισμών δράσεων, αλλά και μέσω της

παροχής των απαιτούμενων γνώσεων που προϋποθέτει η επίτευξη της εξατομίκευσης της θεραπείας [31], [32]. Αυτό επιτυγχάνεται πρωτίστως μέσω της ανάλυσης, της διερεύνησης και της ερμηνείας των ομικών δεδομένων με τη χρήση μεθοδολογιών τεχνητής νοημοσύνης και μηχανικής εκμάθησης. Περαιτέρω εφαρμογή των μεθόδων μηχανικής εκμάθησης πραγματοποιείται στους τομείς της φαρμακευτικής βιομηχανίας όπως στον τομέα της χημειο-πληροφορικής, της γονιδιωματικής πληροφορικής και της βιοϊατρικής απεικόνισης [33], [34], [35].

Η εισαγωγή αλγορίθμων και εργαλείων μηχανικής εκμάθησης καθίσταται εφικτή λόγω της αυξημένης υπολογιστικής ισχύος σε συνδυασμό με την αυξημένη διαθεσιμότητα σε δεδομένα μεγάλου όγκου, με διάφορες φαρμακευτικές εταιρείες να επενδύουν τελευταία σε αυτή. Ωστόσο, παρά τις υψηλές προσδοκίες των αλγορίθμων/εργαλείων αυτών στο πεδίο της ανακάλυψης φαρμάκων και της ιατρικής ακριβείας, υπάρχουν μόνο ορισμένες περιπτώσεις όπου βιοδείκτες και μοντέλα πρόβλεψης έχουν εφαρμοστεί σε κλινικές δοκιμές. Σημαντικοί παράγοντες, που επηρεάζουν την υιοθέτηση τέτοιων προγνωστικών μοντέλων, είναι η επιλογή των κατάλληλων μοντέλων, η δυνατότητα αναπαραγωγής μοντέλων που δημιουργούνται με τη χρήση νευρωνικών δικτύων, η πρόσβαση σε επιμελημένα δεδομένα καθώς και ο σχεδιασμός κατάλληλων δοκιμών για κλινικό περιβάλλον [29].

Μεταξύ των πιο σημαντικών περιορισμών που πρέπει να εξεταστούν προσεκτικά στον τομέα εφαρμογής της μηχανικής εκμάθησης στην ανακάλυψη και στην επαναστόχευση φαρμάκων, είναι η ποιότητα των πειραματικών δεδομένων. Ένα μοντέλο μπορεί να είναι τόσο ακριβές όσο και τα δεδομένα εκπαίδευσης αυτού. Εάν τα δεδομένα είναι εσφαλμένα, εάν δημιουργούν 'θόρυβο' και εάν η συλλογή και η ενσωμάτωσή τους δεν ακολουθεί μια συστηματική και καλά χαρακτηρισμένη διαδικασία αξιολόγησης αυτών, τότε είναι πολύ πιθανό ότι η πρόβλεψη του αλγορίθμου θα είναι ανακριβής. Εάν ο αλγόριθμος και οι χρήστες του δεν λάβουν υπόψη αυτές τις παρατηρήσεις, τότε τα αποτελέσματα της πρόβλεψης μέσω μηχανικής εκμάθησης θα είναι επίσης εσφαλμένα [29].

Χωρίς αμφιβολία, η διαδικασία μηχανικής εκμάθησης παραμένει ένα «μαύρο κουτί» [36]. Από την στιγμή που οι λειτουργίες του μοντέλου δεν προσδιορίζονται ρητά, ο δημιουργός του αλγορίθμου μπορεί να μην γνωρίζει τι ελέγχεται κατά τη διάρκεια των ενδιάμεσων σταδίων

ή και την ακριβή διαδικασία που οδηγεί στο συγκεκριμένο αποτέλεσμα/πρόβλεψη του αλγορίθμου. Για το σκοπό αυτό, η πρόσληψη του κατάλληλα εκπαιδευμένου προσωπικού που θα αποσκοπεί στην γεφύρωση του χάσματος μεταξύ των ομικών επιστημών και της τεχνητής νοημοσύνης, είναι ένας άλλος παράγοντας που πρέπει να εξεταστεί προσεκτικά.

Ανεξάρτητα από τους παραπάνω περιορισμούς, η επίτευξη συνεργασίας μεταξύ φαρμακευτικών εταιρειών και εταιρειών που επικεντρώνονται στην ανάπτυξη προσεγγίσεων μηχανικής εκμάθησης, θα μπορούσε να βοηθήσει στην ταυτοποίηση νέων θεραπευτικών μορίων, αλλά και στην ανακάλυψη νέων θεραπευτικών μεθόδων μέσω της χρήσης δεδομένων μεγάλου όγκου, με απώτερο σκοπό τη βελτιστοποίηση της υγείας. Λαμβάνοντας υπόψη το υψηλό χρονικό αντίτιμο, τους πόρους και την προσπάθεια σχεδιασμού και ανάπτυξης νέων φαρμάκων, η οποία συνοδεύεται από μειωμένα ποσοστά έγκρισης, εύλογα στηρίζομαστε σε υπολογιστικές προσεγγίσεις που αποσκοπούν στην περαιτέρω βελτίωση της αποτελεσματικότητας της διαδικασίας ανάπτυξης φαρμάκων. Αυτές οι εξελίξεις θα συμβάλλουν στη βελτίωση των υπηρεσιών υγειονομικής περίθαλψης, αλλά και στην επίτευξη της ιατρικής ακριβείας [28].

Συμπεράσματα

Συνοπτικά, περιγράψαμε τη λογική και τα βήματα σχεδιασμού αλγορίθμων μηχανικής εκμάθησης, οι οποίοι μπορούν να χρησιμοποιηθούν κατά την μέθοδο της *in silico* διαλογής φαρμάκων (*in silico* drug screening). Οι αλγόριθμοι αυτοί επιτρέπουν τον εντοπισμό δραστικών ουσιών, των οποίων η αναστολή ή η ενίσχυση είναι δυνατό να βελτιώσει την αποτελεσματικότητα των θεραπευτικών προσεγγίσεων. Επιπλέον, όταν εφαρμοστεί η μέθοδος εξόρυξης δεδομένων (*data-mining*), τότε είναι δυνατό να αποσαφηνιστούν σημαντικές σχέσεις μεταξύ γονιδίων και φαρμακευτικής απόκρισης. Για το σκοπό αυτό, παρέχουμε επίσης μία σύνοψη χρήσιμων προσεγγίσεων μηχανικής εκμάθησης που χρησιμοποιούνται στην επαναστόχευση φαρμάκων, η οποία αποτελεί μια εναλλακτική προσέγγιση της 'παραδοσιακής' μεθόδου ανακάλυψης φαρμάκων. Εν κατακλείδι, η τρέχουσα ανασκόπηση αποδεικνύει ότι το πεδίο της Τεχνητής Νοημοσύνης, που περιλαμβάνει τη μηχανική εκμάθηση, μπορεί να εφαρμοστεί στην κλινική πράξη με κύριο σκοπό τη βελτίωση της Ιατρικής Ακριβείας και της διαδικασίας ανακάλυψης φαρμάκων/διαλογής φαρμάκων. ●

ABSTRACT

Computational development of drug response prediction models via machine learning approaches and omics data

Eleni Barbani, Maria Koromina, George P. Patrinos

*University of Patras, School of Health Sciences, Department of Pharmacy,
Laboratory of Pharmacogenomics and Individualized Therapy, Patras, Greece*

In the age of digital health and artificial intelligence, the pharmaceutical industry needs innovative and transformative drug development technologies. Undoubtedly, artificial intelligence and machine learning algorithms have begun to revolutionize the drug development industry over the past five years, although at a slow pace. In this review, we describe the most commonly used “machine learning” approaches in drug development tools and databases. We also analyse new

computational approaches in the field of drug discovery and namely in the context of drug repurposing, as well as the synergies between group sciences, artificial intelligence and machine learning. In addition, we present a future perspective on the ways in which machine learning approaches can be applied, in order not only to accelerate the discovery of drugs, but also to enhance Precision Medicine with the most benefit for patients and public health.

KEY WORDS: artificial intelligence, machine learning, drug development, drug repurposing, pharmaceutical industry

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Kohane, I. S. Ten things we have to do to achieve precision medicine. *Science*, 2015;349, 37.
2. Garvey, C. Interview with Colin Garvey, Rensselaer Polytechnic Institute. *Artificial Intelligence and Systems Medicine Convergence. OMICS*. 2018; 22, 130-132.
3. Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., JR. & Kinzler, K. W. Cancer genome landscapes. *Science*. 2013; 339, 1546-1558.
4. Azuaje, F. Computational models for predicting drug responses in cancer research. *Brief Bioinform*. 2017; 18, 820-829.
5. Libbrecht, M. W. & Noble, W. S. Machine learning applications in genetics and genomics. *Nat Rev Genet*. 2015;16, 321-332.
6. Costello, J. C., Heiser, L. M., Georgii, E., et al. A community effort to assess and improve drug sensitivity prediction algorithms. *Nat Biotechnol*. 2014;32, 1202-1212.
7. Baek, S., Tsai, C. A. & Chen, J. J. Development of biomarker classifiers from high-dimensional data. *Brief Bioinform*. 2019;10, 537-546.
8. Varma, S. & Simon, R. Bias in error estimation when using cross-validation for model selection. *BMC bioinformatics*. 2016;7, 91-91.
9. Byers, L. A., Diao, L., Wang, J., et al. An epithelial-mesenchymal transition gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance. *Clin Cancer Res*. 2017; 19, 279-290.
10. Nicolau, M., Levine, A. J. & Carlsson, G. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proc Natl Acad Sci U S A*. 2011;108, 7265-7270.
11. Gholami, A.M., Hahne, H., Wu, Z., et al. Global Proteome Analysis of the NCI-60 Cell Line Panel. *Cell Rep*. 2013; 4, 609-620.
12. Fersini, E., Messina, E. & Archetti, F. A p-Median approach for predicting drug response in tumour cells. *BMC Bioinformatics*. 2014; 15.
13. Jang, I. S., Neto, E. C., Guinney, J., Friend, S. H. & Margolin, A. A. Systematic assessment of analytical methods

- for drug sensitivity prediction from cancer cell line data. Pacific Symposium on Biocomputing. Pac Symp Biocomput. 2014;63-74.
14. Falgreen, S., Dybkaer, K., Young, K. H., et al. Predicting response to multidrug regimens in cancer patients using cell line experiments and regularised regression models. *BMC Cancer*. 2015;15.
 15. Neto, E. C., Jang, I. S., Friend, S. H. & Margolin, A. A. The Stream algorithm: computationally efficient ridge-regression via Bayesian model averaging, and applications to pharmacogenomic prediction of cancer cell line sensitivity. *Pac Symp Biocomput*. 2014; 27-38.
 16. Amin, S. B., Yip, W. K., Minvielle, S., et al. Gene expression profile alone is inadequate in predicting complete response in multiple myeloma. *Leukemia*. 2014; 28, 2229-2234.
 17. Chen, B., Sirota, M., Fan-Minogue, H., Hadley, D. & Butte, A. J. Relating hepatocellular carcinoma tumor samples and cell lines using gene expression data in translational research. *BMC Med Genomics*. 2015;8 Suppl 2, S5-S5
 18. Cortés-Ciriano, I., Van Westen, G. J. P., Bouvier, G., et al. Improved large-scale prediction of growth inhibition patterns using the NCI60 cancer cell line panel. *Bioinformatics*. 2015;32, 85-95.
 19. Tran, T. P., Ong, E., Hodges, A. P., Paternostro, G. & Piermarocchi, C. Prediction of kinase inhibitor response using activity profiling, in vitro screening, and elastic net regression. *BMC Syst Biol*. 2014;8, 74-74.
 20. Stetson, L. C., Pearl, T., Chen, Y. & Barnholtz-Sloan, J. S. Computational identification of multi-omic correlates of anticancer therapeutic response. *BMC Genomics*. 2014;15, S2.
 21. Consortium, C. C. L. E. C. G. O. D. S. I. C. Pharmacogenomic agreement between two cancer cell line data sets. *Nature*. 2015;528, 84-87.
 22. Wu, Z., Wang, Y. & Chen, L. (2013). Network-based drug repositioning. *Mol Biosyst*. 2013;9, 1268.
 23. Xue, H., Li, J., Xie, H. & Wang, Y. Review of Drug Repositioning Approaches and Resources. *Int J Biol Sci*. 2018;14, 1232-1244.
 24. Godinez, W. J., Hossain, I., Lazic, S. E., Davies, J. W. & Zhang, X. A multi-scale convolutional neural network for phenotyping high-content cellular images. *Bioinformatics*. 2017;33, 2010-2019.
 25. Ferrero, E., Dunham, I. & Sanseau, P. In silico prediction of novel therapeutic targets using gene-disease association data. *Journal of translational medicine. J Transl Med*. 2017; 15, 182-182.
 26. Anderson, A. C. Structure-based functional design of drugs: from target to lead compound. *Methods Mol Biol*. 2012; 823, 359-366.
 27. Zhu, T., Cao, S., Su, P.-C., et al. Hit Identification and Optimization in Virtual Screening: Practical Recommendations Based on a Critical Literature Analysis. *J Med Chem*. 2013; 56, 6560-6572.
 28. Mak, K.-K. & Pichika, M. R. Artificial intelligence in drug development: present status and future prospects. *Drug Discov Today*. 2019;24, 773-780.
 29. Vamathevan, J., Clark, D., Czodrowski, P., et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov*. 2019; 18, 463-477.
 30. Mamoshina, P., Volosnikova, M., Ozerov, I. V., et al. Machine Learning on Human Muscle Transcriptomic Data for Biomarker Discovery and Tissue-Specific Drug Target Identification. *Front Genet*. 2018; 9, 242-242.
 31. Li, B., Shin, H., Gulbekyan, G., et al. Development of a Drug-Response Modeling Framework to Identify Cell Line Derived Translational Biomarkers That Can Predict Treatment Outcome to Erlotinib or Sorafenib. *PLoS one*. 2015; 10, e0130700-e0130700.
 32. Kraus, V. B. Biomarkers as drug development tools: discovery, validation, qualification and use. *Nat Rev Rheumatol*. 2018; 14, 354-362.
 33. Gonczarek, A., Tomczak, J. M., Zareba, S., Kaczmar, J., Dąbrowski, P. & Walczak, M. J. Interaction prediction in structure-based virtual screening using deep learning. *Comput Biol Med*. 2018; 100, 253-258.
 34. Ekins, S. The Next Era: Deep Learning in Pharmaceutical Research. *Pharm Res*. 2016; 33, 2594-2603.
 35. Angermueller, C., Pärnamaa, T., Parts, L. & Stegle, O. Deep learning for computational biology. *Mol Syst Biol*. 2016; 12, 878-878.
 36. Lamberti, M. J., Wilkinson, M., Donzanti, B. A., et al. A Study on the Application and Use of Artificial Intelligence to Support Drug Development. *Clin Ther*. 2019; 41, 1414-1426.

ΠΡΩΤΟΤΥΠΟ ΑΡΘΡΟ

Η γενετική σπουδαιότητα της ναυμαχίας της Σαλαμίνας

Κωνσταντίνος Τριανταφυλλίδης*

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Τμήμα Βιολογίας, Θεσσαλονίκη

ΠΕΡΙΛΗΨΗ

Στο πρώτο μισό του 5ου αιώνα π.Χ. οι Πέρσες εισέβαλαν στην Ευρώπη, αλλά οι Έλληνες τους νίκησαν στη μάχη του Μαραθώνα, στη ναυμαχία της Σαλαμίνας και τις μάχες των Πλαταιών και της Μυκάλης. Στόχος της εργασίας είναι να διερευνηθεί εάν υπήρξε γενετική (DNA) ροή από το Ιράν στον Ευρωπαϊκό/Ελλαδικό χώρο. Προς τούτο συγκεντρώθηκαν επιστημονικά στοιχεία για τη γενετική σύσταση των κατοίκων της Ελλάδας. Τα δεδομένα αυτά συγκρίθηκαν με αντίστοιχα δεδομένα από τον Ιράν, την Εγγύς Ανατολή, αλλά και όλη την Ευρώπη. Από αυτά

προκύπτει ότι δεν υπάρχουν ενδείξεις ότι στην προαναφερόμενη χρονική περίοδο συνέβησαν μεγάλα μεταναστευτικά γεγονότα και ροή πληθυσμιακών ομάδων, άρα και γονιδίων, από το Ιράν προς τον Ελλαδικό γεωγραφικό χώρο. Ως εκ τούτου, οι συνέπειες της Ναυμαχίας της Σαλαμίνας και των δύο άλλων μαχών υπήρξαν καθοριστικές για την εξέλιξη των ιστορικών γεγονότων για την Ελλάδα, και για την Ευρώπη γενικότερα, αλλά και για τη γενετική σύσταση των κατοίκων των Βαλκανίων και της Ευρωπαϊκής ηπείρου.

ΛΕΞΕΙΣ ΕΥΡΕΤΗΡΙΟΥ: γενετική ροή, Ναυμαχία της Σαλαμίνας, Περσική εισβολή στην Ευρώπη, Απλοομάδες, μιτοχονδρικό DNA

Εισαγωγή

Παλαιολιθική εποχή. Ο ανατομικά σύγχρονος άνθρωπος (*Homo sapiens*) πρωτοεμφανίστηκε στην Αφρική πριν από 320.000 χρόνια. Πριν από 220.000 χρόνια πληθυσμιακές ομάδες κυνηγών-τροφοσυλλεκτών *Homo sapiens* άρχισαν να πορεύονται προς τη Μέση Ανατολή και από εκεί τόσο προς την Ευρώπη, όσο και την Ανατολική Ασία. Κατά την *Παλαιολιθική εποχή*, λοιπόν, η Μέση Ανατολή ήταν ένα πέρασμα για τον αρχέγονο άνθρωπο από την Αφρική προς την Ευρώπη και την υπόλοιπη Ασία.

Νεολιθική εποχή. Πριν 12.000 χρόνια οι κυνηγοί-τροφοσυλλέκτες της *Εύφορης Ημισελήνου* (από τη Μικρά Ασία και την Παλαιστίνη μέχρι το Δυτικό Ιράν) αναπτύσσουν την καλλιέργεια των δημητριακών (σιτάρι, κριθάρι, φακή, μπιζέλι) και εκτρέφουν παραγωγικά ζώα, π.χ. πρόβατα, κατσίκια, βοοειδή και χοίρους. Η μετάβαση από τη ζωή του κυνηγού-τροφοσυλλέκτη στην ανάπτυξη της γεωργοκτηνοτροφίας θεωρείται το πιο σημαντικό πολιτισμικό και τεχνολογικό γεγονός στην ιστορία του ανατομικά σύγχρονου ανθρώπου. Κατά τη *Νεολιθική εποχή*, οι γεωργοί από αυτήν την περιοχή

* Αντεπιστέλλων Συγγραφέας

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Τμήμα Βιολογίας, Θεσσαλονίκη. Email: triant@bio.auth.gr

εξαπλώθηκαν τόσο Δυτικά, μέσω Μικράς Ασίας στην Ευρώπη, όσο και Ανατολικά στην υπόλοιπη Ασία.

Το Ιράν: Τα υψίπεδα του Ιράν (έως το 1935 η επικρατούσα ονομασία ήταν Περσία) βρίσκονται στο σταυροδρόμι των ανθρωπίνων μεταναστεύσεων από την πρώτη διασπορά των ανατομικά σύγχρονων ανθρώπων από την Αφρική στην υπόλοιπη Ασία την *Παλαιολιθική εποχή*, καθώς και τη Νεολιθική εποχή, λόγω της γεωστρατηγικής τους θέσης. Σε αρκετές προϊστορικές τοποθεσίες στον Ιρανικό γεωγραφικό χώρο βρέθηκαν στοιχεία που δείχνουν ότι εκεί αναπτύχθηκαν οι πρώτοι οργανωμένοι οικισμοί από το 7.000 π.Χ. Ήδη κατά τη τρίτη χιλιετία π.Χ., οι Μήδοι, οι Πέρσες και οι Πάρθοι (φύλα ινδοευρωπαϊκά και συγγενικά μεταξύ τους) ζούσαν στο Δυτικό μέρος του Ιράν. Οι Πέρσες επικράτησαν στην περιοχή και δημιούργησαν την αυτοκρατορία των Αχαιμενιδών (648-330 π.Χ.), που έφτασε στα όρια της ακμής της τον 6^ο και 5^ο π.Χ. αιώνα, και εκτεινόταν από τη Μικρά Ασία έως τον ποταμό Ινδό.

Τα γεγονότα: Την περίοδο 499-449 π.Χ. οι Πέρσες συγκρούστηκαν με τους Έλληνες στην ευρύτερη περιοχή της Μικράς Ασίας και του Αιγαίου Πελάγους και οι προσπάθειες τους να εισβάλουν στην Ευρώπη αναχαιτίστηκαν. Ειδικότερα, ο Δαρείος Α΄ εισέβαλε στην Ελλάδα, αλλά ηττήθηκε στη μάχη του Μαραθώνα, το 490 π.Χ. Η μάχη του Μαραθώνα ήταν καθοριστική στιγμή στην ιστορία της αθηναϊκής δημοκρατίας.

Ο γιος του Ξέρξης Α΄, επανέλαβε την προσπάθεια, αλλά ηττήθηκε στη Ναυμαχία της Σαλαμίνας το 480 π.Χ. και απωθήθηκε οριστικά στη Μάχη των Πλαταιών και της Μυκάλης το επόμενο έτος. Πολλοί ιστορικοί θεωρούν ότι η ναυμαχία της Σαλαμίνας είναι μια από τις πιο σημαντικές μάχες στην ιστορία της ανθρωπότητας. Για να υποστηρίξουν αυτή την άποψη, δηλώνουν ότι αν οι Έλληνες είχαν ηττηθεί στη ναυμαχία, η πιθανή κατάληψη της Ελλάδας από τους Πέρσες θα σταματούσε την ανάπτυξη του Ελληνικού και του μετέπειτα Δυτικού πολιτισμού.

Το διάστημα 356-323 π.Χ. ο Μέγας Αλέξανδρος κατάφερε να επιβληθεί του Πέρση βασιλιά Δαρείου Γ΄ και να κατακτήσει τα εδάφη της αυτοκρατορίας του. Η κυριαρχία των διαδόχων του Μεγάλου Αλεξάνδρου στην περιοχή της Περσίας υπήρξε όμως πρόσκαιρη. Μέσα σε λιγότερο από έναν αιώνα, οι Πάρθοι επικράτησαν σε ολόκληρη την περιοχή. Μουσουλμάνοι Άραβες κατέλυσαν το Περσικό βασίλειο τον 7^ο αιώνα μ.Χ. και από τότε σταδιακά ξεκίνησε ο εξισλαμισμός της περιοχής. Μετέπειτα υπήρχε εισροή Μογγολικών πληθυσμών από την Κεντρική Ασία.

Στη βιβλιογραφία υπάρχουν πλέον στοιχεία ταυτόχρονης μελέτης της γενετικής (DNA) σύστασης των Ελλήνων και των Ιρανών και μάλιστα χρησιμοποιώντας πολλούς γενετικούς δείκτες. Μπορούμε λοιπόν πλέον να δούμε αν και ποια ήταν η γενετική σημασία των Περσικών εισβολών για την Ελλάδα, αλλά και για την Ευρώπη γενικότερα.

Για να μελετηθεί η γενετική σύσταση κάποιου πληθυσμού, μπορούν να χρησιμοποιηθούν πολλοί DNA δείκτες, όπως του μιτοχονδριακού DNA, ή του χρωμοσώματος Υ. Σήμερα κυρίως εξετάζονται χιλιάδες μονονουκλεοτιδικοί πολυμορφισμοί. Τέλος, η ολοκλήρωση της ανάλυσης του DNA με τους κατάλληλους γενετικούς δείκτες οδηγεί στην ταξινόμηση των μελών ενός πληθυσμού σε ομάδες που μοιράζονται την ίδια ακριβώς αλληλουχία DNA, τις λεγόμενες *απλοομάδες*.

Ποσοστά απλοομάδων Ελλήνων και Ιρανών

Οι κύριες απλοομάδες (γενεαλογικές γραμμές) του χρωμοσώματος Υ και του μιτοχονδριακού DNA που παρατηρήθηκαν στον Ελληνικό και τον Ιρανικό πληθυσμό παρουσιάζονται στους **Πίνακες 1** και **2**, αντίστοιχα [1, 2, 3]. Από τους Πίνακες προκύπτει ότι η πλειοψηφία (ποσοστό >95%) τόσο των Ελλήνων, όσο και των Ιρανών ανήκουν σε τουλάχιστον 15 απλοομάδες, είτε μιλάμε για απλοομάδες μητροπλευρικές, είτε πατροπλευρικές.

Από τον **Πίνακα 1** προκύπτει ότι στην πατροπλευρική σύσταση των Ιρανών υψηλότερες συχνότητες καταγράφηκαν για τις απλοομάδες J (31,4%), R (29,1%), G (11,8%) και E (9,2%), ενώ των Ελλήνων R (28,1%), J (26%), E (19,3%) και I (14,5%). Επιπρόσθετα, πέρα από το γεγονός ότι οι δύο πιο συχνές απλοομάδες (R και J) είναι διαφορετικές στους δύο πληθυσμούς, ακόμη πιο σημαντικό είναι ότι οι συχνότητες των δύο υπο-απλοομάδων (R1 και R2, καθώς και J1 και J2) έχουν στατιστικά σημαντικές διαφορές μεταξύ τους. Ειδικότερα, η υπο-απλοομάδα R1 δείχνει πρόσφατη μετακίνηση πληθυσμών από τη Δυτική Ευρασία, ενώ της απλοομάδας Q από την Κεντρική Ασία προς το Ιράν. Με απλά λόγια, διαπιστώνουμε ότι μετακινήσεις πληθυσμιακών ομάδων από την Κεντρική Ασία μέσω του Ιράν προς τον Ελλαδικό χώρο ήταν σχεδόν μηδαμινές. Από την άλλη, γνωρίζουμε ότι το πρότυπο κατανομής του κλάδου J1 τόσο στην Ευρώπη, όσο και στη Ασία σχετίζεται με την εξάπλωση των Αράβων στην Ασία, τη Βόρεια Αφρική και την Ιβηρική χερσόνησο τον 7^ο αιώνα μ.Χ. Το υψηλότερο ποσοστό της πατροπλευρικής υπο-απλοομάδας J1 στους Ιρανούς μάλλον οφείλεται στην εξάπλωση των Αράβων στην περιοχή. Γνωρίζουμε, ότι Μουσουλμάνοι Άραβες κατέλυσαν το Περσικό βασίλειο

Πίνακας 1. Κατανομή των συχνοτήτων των απλοομάδων του χρωμοσώματος Y στους Έλληνες (1707 άτομα) και τους Ιρανούς (938 άτομα)

| Απλοομάδα | I | R1 | R2 | G | J1 | J2 | E | T | Q | N | K | L | C | Άλλες |
|-----------|------|------|------|------|-----|------|------|-----|-----|-----|---|-----|-----|-------|
| Ελλάδα | 14,5 | 11,5 | 16,6 | 6,1 | 3,0 | 23,0 | 19,3 | 2,6 | 0,1 | 0,1 | 1 | 0,3 | 0,3 | 2,2 |
| Ιράν | 0,7 | 26,3 | 2,8 | 11,8 | 8,9 | 22,5 | 9,2 | 3,4 | 5,6 | 0,4 | - | 5,0 | 1,0 | 2,4 |

τον 7^ο αιώνα μ.Χ. και από τότε σταδιακά ξεκίνησε ο εξισλαμισμός της περιοχής. Αντίθετα, η πλειονότητα των χρωμοσωμάτων με απλοομάδα J στην Ελλάδα και στο Ιράν ανήκουν στον υποκλάδο J2, που σχετίζεται με τη Βόρεια Μεσοποταμία, την έναρξη ενασχόλησης με τη γεωργία και την εξάπλωση των γεωργών προς την Ευρώπη και την υπόλοιπη Ασία.

Ανάλυση κύριων συνιστωσών. Από την στατιστική επεξεργασία των παραπάνω γενετικών στοιχείων με τη μεθοδολογία ανάλυσης κυρίων συνιστωσών προέκυψε η **Εικόνα 1**. Σε αυτήν διαπιστώνουμε ότι οι πληθυσμοί ταξινομούνται ανάλογα με τη γεωγραφική και εθνική καταγωγή τους και διακρίνονται πέντε κύριες πληθυσμιακές ομάδες (Ανατολική Αφρική, Άραβες, Ευρωπαίοι, Εγγύς Ανατολή και Νότια Ασία): Από όλες τις ομάδες, ο πληθυσμός του Ιράν ξεχωρίζει σαφώς από τους Ευρωπαϊκούς, συμπεριλαμβανομένου του Ελληνικού.

Οι διάφορες Ιρανικές πληθυσμιακές ομάδες δεν συγκεντρώνονται όλες μαζί, αλλά καταλαμβάνουν ενδιάμεσες θέσεις μεταξύ Αραβικών πληθυσμών, και πληθυσμών προερχομένων από την Εγγύς Ανατολή και τη Νότια Ασία.

Από τον **Πίνακα 2** προκύπτει, επίσης, ότι οι πιο συχνές μιτοχονδριακές απλοομάδες που καταγράφηκαν στους Ιρανούς (U, H, J και M) έχουν στατιστικά σημαντικές διαφορές από τις αντίστοιχες συχνότητες των Ελλήνων [2, 3]. Στην μητροπλευρική γενετική σύσταση των Ιρανών παρατηρήθηκαν γενεαλογικές γραμμές που χαρακτηρίζουν Ευρωπαϊκούς πληθυσμούς (δηλαδή H και U), αλλά και γενεαλογικές γραμμές που χαρακτηρίζουν πληθυσμούς της Ανατολικής και Κεντρικής Ασίας (δηλαδή M και B), κάτι που δεν συμβαίνει στον Ελληνικό πληθυσμό, εκτός της πολύ χαμηλής συχνότητας της απλοομάδας M.

Διαχρονική-χωρική σχέση σημερινών Ιρανών με αρχαία δείγματα DNA

Σε μελέτη [4] διερευνήθηκε η γενετική σύνθεση των σημερινών Ιρανών και η χρονική-χωρική προέλευσή

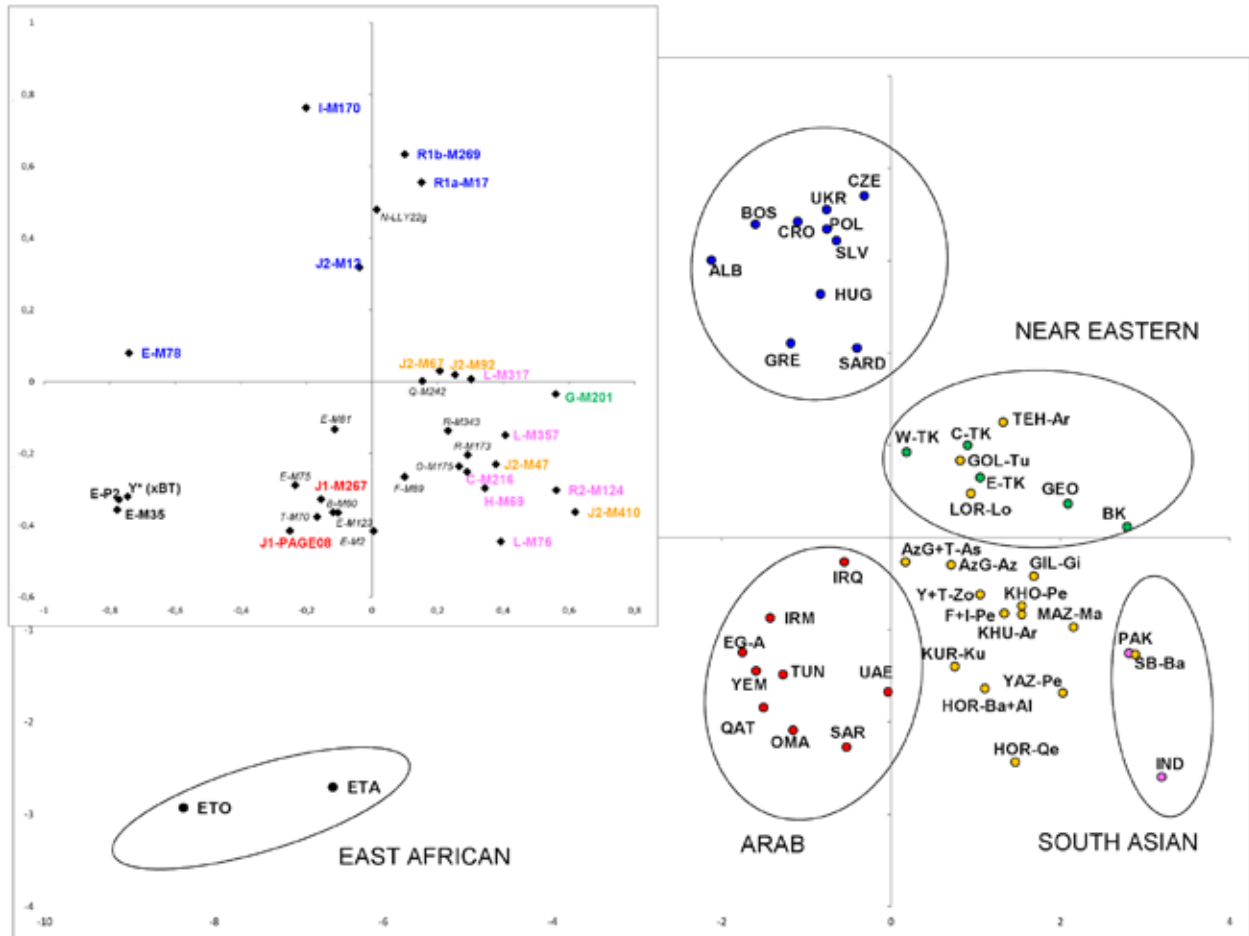
τους. Πληθυσμιακές ομάδες του Κεντρικού Ιράν σε μεγάλο βαθμό είχαν αυτόχθονη ανάπτυξη, τουλάχιστον τα τελευταία 5.000 χρόνια. Είναι αξιοσημείωτο ότι στον τοπογραφικό χάρτη τα δείγματα της *Πρώιμης Νεολιθικής Εποχής* από το Δυτικό Ιράν και την αρχαιολογική θέση Tarreh Hesar (B.A. Ιράν) εντοπίζονται μαζί με δείγματα από τον απομακρυσμένο Περσικό Κόλπο, ενώ δείγματα της *Εποχής του Χαλκού* από την θέση Tarreh Hesar έδειξαν τάση συγγένειας με δείγματα από το Κεντρικό Ιράν.

Η γενετική θέση των Ελλήνων και των Ιρανών στον Παγκόσμιο γενετικό χάρτη

Οι πληθυσμιακές γενετικές μελέτες στην Ευρώπη έχουν καταλήξει στο ακόλουθο εξαιρετικά σημαντικό συμπέρασμα: "Παρά το χαμηλό βαθμό της μέσης γενετικής διαφοροποίησης ανάμεσα στους Ευρωπαίους, διαπιστώθηκε και σημαντικός βαθμός αντιστοιχίας ανάμεσα στη γενετική ποικιλότητα και τη γεωγραφική θέση στον Ευρωπαϊκό γενετικό χάρτη των πληθυσμών" [5, 6]. Ή με απλά λόγια, όλοι οι Ευρωπαίοι μοιάζουν πολύ γενετικά μεταξύ τους, αλλά πάντοτε ο κάθε πληθυσμός γενετικά μοιάζει πολύ περισσότερο με τους γεωγραφικούς του γείτονες. Παρόμοια συσχέτιση διαπιστώθηκε και στους κατοίκους της Ασίας [4].

Ειδικότερα, σε δύο μελέτες [4, 7] εξετάστηκε η γενετική θέση των Ελλήνων και των Ιρανών στον παγκόσμιο γενετικό χάρτη με τις πλέον σύγχρονες γενετικές μεθόδους, δηλ. χρησιμοποιώντας χιλιάδες *μονο νουκλεοτιδικούς πολυμορφισμούς* από Ευρωπαϊκά, Ασιατικά και Αφρικανικά πληθυσμιακά δείγματα (**Εικ. 2**).

Τα γενετικά στοιχεία της πρώτης εργασίας επιβεβαιώνουν τη στενή γενετική συγγένεια μεταξύ Ελλήνων και Ιταλών, ενώ υπάρχει σαφής διάκριση των Ελλήνων από τους Ιρανούς, μια και εντάσσονται σε εντελώς διαφορετική γενετική πληθυσμιακή ομάδα. Επιπρόσθετα, οι Ιρανοί εντάσσονται στην ίδια γενετική πληθυσμιακή ομάδα με τους Τούρκους. Στη δεύτερη εργασία οι ερευ-



Εικ. 1. Η στατιστική ανάλυση κυρίων συνιστωσών έγινε με βάση τις συχνότητες των απλοομάδων του χρωμοσώματος Υ των Ιρανών (κίτρινο χρώμα). Αυτές συγκρίθηκαν με αντίστοιχες από την Ανατολική Αφρική (μαύρο χρώμα), τη Βόρεια Αφρική και τη Μέση Ανατολή (κόκκινο χρώμα), Ευρωπαίους (μπλε χρώμα) και Τούρκους/περιοχή του Καυκάσου (πράσινο χρώμα) και τη Νότια Ασία (ροζ χρώμα). Από Grugni et al. 2011.

νητές επιβεβαιώνουν ότι ο Ιρανικός πληθυσμός συγγενεύει με τους γειτονικούς πληθυσμούς, αλλά ταυτόχρονα παρουσιάζει χαρακτηριστική γενετική ποικιλότητα που είναι απόρροια της μακροχρόνιας γενετικής συνέχειας.

Συνολικά, και αυτά τα αποτελέσματα δεν παρέχουν ενδείξεις μεγάλων μεταναστευτικών γεγονότων και ροής ανθρώπινων πληθυσμιακών ομάδων, άρα και γονιδίων, από το Ιράν στον Ελλαδικό/Ευρωπαϊκό χώρο.

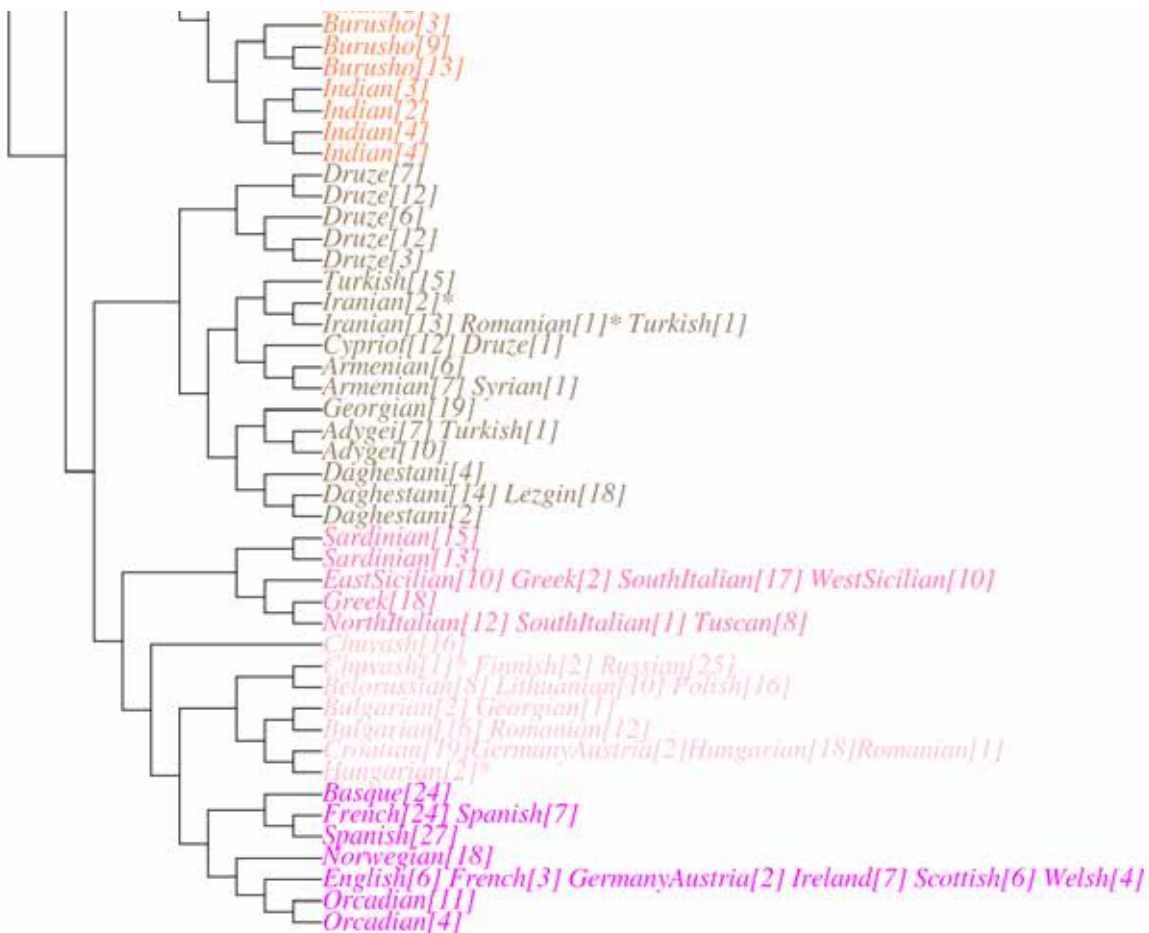
Συμπεράσματα

Συνοπτικά, οι Ιρανοί παρουσιάζουν ξεχωριστή γενετική ποικιλότητα, που προκύπτει από τη μακράιωνη γενετική συνέχειά τους στην περιοχή. Η παρατηρούμενη αντιστοιχία μεταξύ της γενετικής ποικιλότητας και της γεωγραφικής θέσης συνάδουν με τα αρχαιο-

λογικά και ιστορικά στοιχεία για το Ιράν. Η γενετική σύσταση των Ελλήνων και των Ιρανών σήμερα διαφέρει σημαντικά και δεν φαίνεται να υπήρξε γενετική ροή από την Περσία προ την Ελλάδα στο πρώτο μισό του 5^{ου} αιώνα π.Χ. Μπορεί κανείς να πει ότι οι νίκες των Ελλήνων εναντίον των Περσών ανέκοψαν την πορεία των Περσών προς Δυσμάς με αποτέλεσμα να μην υπάρχουν γενετικές ενδείξεις ότι σε εκείνη την χρονική περίοδο υπήρξε ροή γονιδίων, συνεπώς ούτε και μεγάλων μεταναστευτικών γεγονότων και ροής πληθυσμιακών ομάδων, από το Ιράν (Περσία) προς τον Ελλαδικό γεωγραφικό χώρο. Ως εκ τούτου, οι συνέπειες της νικηφόρου ναυμαχίας της Σαλαμίνας και των άλλων τριών νικηφόρων μαχών υπήρξαν καθοριστικές για την εξέλιξη των ιστορικών γεγονότων και πιθανόν επέδρασαν σημαντικά και στα γενετικά

Πίνακας 2. Κατανομή των συχνοτήτων των μιτοχονδριακών απλοομάδων στους Έλληνες (999 άτομα) και τους Ιρανούς (509 άτομα)

| Απλοομάδα | L | HV | H | J | T | U | K | I | W | X | N | M | B | Άλλες |
|-----------|-----|-----|------|------|------|------|-----|-----|-----|-----|-----|------|---|-------|
| Ελλάδα | 0,2 | 4,9 | 38,7 | 10,0 | 10,1 | 14,8 | 4,8 | 2,0 | 1,3 | 4,2 | 5,8 | 2,6 | - | 1,0 |
| Ιράν | 0,3 | 7,4 | 17,5 | 13,8 | 8,0 | 18,5 | 7,3 | 2,4 | 2,4 | 2,8 | 7,0 | 10,0 | 2 | 0,6 |

**Εικ. 2.** Φυλογενετικό δένδρο που δείχνει τη γενετική θέση των Ελλήνων και των Ιρανών στον παγκόσμιο γενετικό χάρτη [7].

πράγματα, όχι μόνο για την Ελλάδα, αλλά και για την Ευρώπη γενικότερα. Σύμφωνα με ιστορικούς, αν οι Πέρσες νικούσαν στη Σαλαμίνα, θα είχε σταματήσει η ανάπτυξη του αρχαίου Ελληνικού και του μετέπειτα Δυτικού πολιτισμού, και κατά συνέπεια ο Δυτικός πολιτισμός δεν θα ήταν αυτό που είναι σήμερα. Αν νικούσαν οι Πέρσες, τότε ενδεχομένως να υπήρχε διάδοση του Περσικού πολιτισμού, αλλά η εξέλιξη των ιστορικών γεγονότων, και η γενετική σύσταση των

κατοίκων της Ελλάδας και της Ευρώπης γενικότερα θα είχε ακολουθήσει άλλες διαδρομές. Γι' αυτό, η ναυμαχία της Σαλαμίνας θεωρείται από τις πιο σημαντικές μάχες στην ανθρώπινη ιστορία.

Δυόμιση χιλιάδες χρόνια πέρασαν από την κεφαλαιώδους σημασία Ναυμαχία της Σαλαμίνας. Η σημασία αυτού του γεγονότος είναι επίκαιρη, μια και το παρελθόν αυτό σηματοδοτεί ακόμη αμείωτα το παρόν και το μέλλον του ελεύθερου ανθρώπου. ●

ABSTRACT

The genetic importance of the naval battle of Salamis

Costas Triantaphyllidis

Aristotle University of Thessaloniki, Department of Biology, Thessaloniki, Greece

In the first half of the 5th century BC, the Persians invaded Europe, but the Greeks defeated them at the Battle of Marathon, the Naval Battle of Salamis and the Battles of Plataea and Mykale. The aim of the article is to investigate whether there was a genetic flow from Iran to the European/Greek geographical area. To this end, scientific data were collected on the genetic composition of the inhabitants of Greece. These data were compared with similar data from Iran, the Middle East, but also the whole Eu-

rope. It follows that there are no indications for this period of large migratory events and population group flows, and therefore genes, from the Iran geographical area to the Greek area. Therefore, the consequences of the Naval Battle of Salamis and the other two battles were decisive for the evolution of historical developments for Greece, and Europe in general, but also for the genetic composition of the inhabitants of the Balkans and the European continent.

KEY WORDS: genetic flow, Battle of Salamis, Persian invasion of Europe, lineage, mitochondrial DNA (mtDNA)

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Grougni V., et al. Ancient migratory events in the Middle East: New clues from the Y-chromosome variation of modern Iranians. *PloS One*, 2012;7 e41252.
2. Zarei F. and Alipanah H. Mitochondrial DNA variation, genetic structure and demographic history of Iranian populations. *Molecular Biology Research Commun.* 2014;3: 45-65.
3. Τριανταφυλλίδης Κ. (2020). Η ΓΕΝΕΤΙΚΗ ΙΣΤΟΡΙΑ ΤΩΝ ΕΛΛΗΝΩΝ. Το DNA των Ελλήνων. Εκδόσεις Κυριακίδη.
4. Mehrjoo Z., et al. Distinct genetic variation and heterogeneity of the Iranian population. *PloS Genetics* 2019;15: e1008385.
5. Lao O., et al. (). Correlation between genetic and geographic structure in Europe. *Curr. Biol.*, 2008; 18: 1241-1248.
6. Novembre J., et al. Genes mirror geography within Europe. *Nature*, 2008; 456: 98-101.
7. Hellenthal G., et al. A Genetic Atlas of Human Admixture History. *Science*, 2014; 343: 747-751.

Η Ένωση Φοιτητών Φαρμακευτικής Ελλάδας (GPSF) συναντά τη Φαρμακογονιδιωματική και Εξατομικευμένη Ιατρική

Τατιάνα Γραμματικοπούλου¹, Δήμητρα Δεδούση^{2*}, Μαρία Θεριανού², Ίρις Τριπολιτσιώτη¹

¹Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών, Τμήμα Φαρμακευτικής, Αθήνα,

²Πανεπιστήμιο Πατρών, Σχολή Επιστημών Υγείας, Τμήμα Φαρμακευτικής, Πάτρα

ΠΕΡΙΛΗΨΗ

Η Ένωση Φοιτητών Φαρμακευτικής Ελλάδας είναι μία εθελοντική ομάδα με δραστηριοποίηση τόσο σε εθνικό επίπεδο, μέσω των μελών της που ανήκουν στα τμήματα Φαρμακευτικής, όσο και στην Ευρώπη, με την συμμετοχή της στην Ευρωπαϊκή Ένωση Φοιτητών Φαρμακευτικής. Κύριοι άξονές της αποτελούν δραστηριότητες εκπαιδευτικού, επιστημονικού και κοινωνικού ενδιαφέροντος. Τα μέλη της ομάδας συμμετέχουν σε μια πληθώρα δραστηριοτήτων διευρύνοντας τις γνώσεις τους

και βελτιώνοντας δεξιότητές τους. Η ανάδειξη της Φαρμακευτικής και ο συνδυασμός της με έτερες επιστήμες, αποδεικνύει το ευρύ φάσμα αντικειμένων με το οποίο μπορεί να ασχοληθεί ένας φαρμακοποιός. Μείζον αντικείμενο ενασχόλησης είναι και ο τομέας της Εξατομικευμένης Ιατρικής και Φαρμακογονιδιωματικής. Τέλος, με στόχο την καλύτερη εξοικείωση των φοιτητών με τα επιμέρους τμήματά του, διοργανώνεται σειρά ομιλιών με θέμα την Βιοπληροφορική.

ΛΕΞΕΙΣ ΚΥΡΙΑΡΧΕΙΟΥ: Ένωση Φοιτητών Φαρμακευτικής Ελλάδας, Φαρμακευτική, δράσεις, Εξατομικευμένη Ιατρική, Βιοπληροφορική

Εισαγωγή

Η Ένωση Φοιτητών Φαρμακευτικής Ελλάδας - Ε.Φ.Φ.Ε. (Greek Pharmaceutical Students' Federation - GPSF) είναι μια μη κερδοσκοπική, μη κυβερνητική, μη συνδικαλιστική οργάνωση στην οποία μπορούν να συμμετέχουν προπτυχιακοί, μεταπτυχιακοί φοιτητές και υποψήφιοι διδάκτορες Τμημάτων Φαρμακευτικής ελληνικών Πανεπιστημιακών Ιδρυμάτων [1]. Η Ένωση εκπροσωπεί

τα μέλη της σε εθνικό και ευρωπαϊκό επίπεδο και σκοπό έχει να προάγει τη συνεργασία για την εξέλιξη της φαρμακευτικής και υγειονομικής περίθαλψης.

Η ομάδα ξεκίνησε τη λειτουργία της στα μέσα της δεκαετίας του 2000 και κίνητρο αποτέλεσε η επιθυμία να δημιουργηθεί μια Ένωση στην οποία θα μπορούσαν να συμμετέχουν όλοι οι φοιτητές Φαρμακευτικής Ελλάδας, να οργανώνουν δράσεις και να εκπροσωπούν

* Αντεπιστέλλων Συγγραφέας

Πανεπιστήμιο Πατρών, Σχολή Επιστημών Υγείας, Τμήμα Φαρμακευτικής, Πανεπιστημιούπολη, Ρίο, 265 04, Πάτρα, Email: gpsf-patras.contact@gmail.com

την χώρα στην Ευρώπη. Για τον λόγο αυτό, αμέσως μετά την ίδρυσή της, η Ένωση γίνεται μέλος της EPSA (Associate Member) της Ευρωπαϊκής Ένωσης Φοιτητών Φαρμακευτικής (European Pharmaceutical Students' Association - EPSA), παρέχοντας τη δυνατότητα συμμετοχής τόσο στις δράσεις της, όσο και στη Γενική της Συνέλευση [2]. Επιπλέον, στην 60η Γενική Συνέλευση της EPSA έγινε ομόφωνα δεκτή η αίτηση της GPSF να γίνει Τακτικό Μέλος (Ordinary Member), δηλαδή να έχει δικαίωμα ψήφου στις Γενικές Συνελεύσεις της EPSA, ενώ κατάφερε να αναδειχθεί Μέλος της Χρονιάς 2019-2020 στην EPSA.

Έδρα της Ένωσης αποτελεί η Αθήνα με εγγεγραμμένα μέλη της στα τμήματα Φαρμακευτικής της Αθήνας, της Θεσσαλονίκης και της Πάτρας. Τον Νοέμβριο του 2020, η GPSF και τα μέλη της που φοιτούν στο τμήμα Φαρμακευτικής στην Πάτρα, ίδρυσαν ένα καινούριο Παράρτημα στην πόλη αυτή. Βασιζόμενοι στις αρχές που προσβύει η Ένωση και στοχεύοντας στην ενημέρωση και δραστηριοποίηση περισσότερων φοιτητών, η GPSF πλέον θα έχει έναν πιο ενεργό ρόλο και στην Πάτρα.

Δράσεις της GPSF

Οι δράσεις της Ένωσης επικεντρώνονται σε επιστημονικά και κοινωνικά θέματα και στην ενημέρωση για τον επαγγελματικό προσανατολισμό. Η ανάδειξη της πολύπλευρης φύσης του Φαρμακοποίου και ο ρόλος του στη σημερινή κοινωνία, είναι οι κύριοι στόχοι της Ένωσης.

Πιο συγκεκριμένα, έχουν διοργανωθεί πλήθος εκστρατειών ενημέρωσης εκπαιδευτικού ενδιαφέροντος που σχετίζονται με θέματα ιατροφαρμακευτικού, αλλά και κοινωνικού χαρακτήρα, άλλοτε με φυσική παρουσία σε δημόσιους χώρους, και άλλοτε διαδικτυακά, όπως στην περίοδο της Πανδημίας CoViD-19 [3]. Η ενημέρωση αυτή μπορεί να λάβει χώρα και με τη δημιουργία επιμορφωτικού βίντεο, στα πλαίσια μιας χαρακτηριστικής δράσης της GPSF που ξεκίνησε το 2019, τα 'Learn From the Experts' [4]. Πρόκειται για συνεντεύξεις που δημοσιεύονται μέσω πλατφόρμας κοινωνικής δικτύωσης της ομάδας στις οποίες ο ομιλητής που εξειδικεύεται σε ένα αντικείμενο απαντάει σε σχετικές ερωτήσεις.

Ανταποκρινόμενοι στις ανάγκες των νέων και της κοινωνίας, γίνεται σημαντική προσπάθεια για την ενημέρωση σε θέματα κρίσιμα για τη δημόσια υγεία. Οι πρωτόγνωρες συνθήκες στις οποίες όλοι έχουμε εκτεθεί εξαιτίας της πανδημίας, όχι μόνο δεν περιορίσαν τις δράσεις της ομάδας, αλλά ενίσχυσαν το αίσθημα των μελών για παροχή αξιόπιστων και απαραίτητων στο κοινό πληροφοριών. Για παράδειγμα, η παρατήρηση

της αμφισβήτησης της κοινής γνώμης απέναντι στην αποτελεσματικότητα των εμβολίων, ώθησε την ομάδα να δημιουργήσει ένα Learn From the Experts σχετικό με τα εμβόλια.

Με άξονα την εκπαίδευση και την ενημέρωση για τις επιστημονικές εξελίξεις, η GPSF διοργανώνει διαδικτυακά σεμινάρια, ημερίδες και συνέδρια στα οποία συμμετέχουν ως ομιλητές ειδικοί επιστήμονες, τα οποία στόχο έχουν να εμπλουτίσουν τις γνώσεις των φοιτητών Φαρμακευτικής [5]. Τέτοια παραδείγματα αποτελούν το Mentor Up, το PharmACon (**Εικόνα 1**) και το Career Month. Το PharmACon αποτέλεσε μία από τις σημαντικότερες επιστημονικές εκδηλώσεις που διοργάνωσε η GPSF, με διακεκριμένους εκπροσώπους της ελληνικής επιστημονικής κοινότητας, προσφέροντας μία ευρεία θεματολογία. Το Career Month, πρόσφατο δρώμενο για τον επαγγελματικό προσανατολισμό, είναι εμπνευσμένο από την αντίστοιχη δράση της EPSA, Career Week. Καταξιωμένοι Έλληνες και διεθνείς ομιλητές ενημέρωσαν τους συμμετέχοντες για την επαγγελματική εξέλιξη σε διάφορα πεδία της Φαρμακευτικής, ενώ υπήρχαν και ομιλίες αναφορικά με την προσωπική εξέλιξη και βελτίωση. Άλλες θεματολογίες που έχουν αναφερθεί αποτελούν η μικροβιακή αντοχή, η νευρογένεση στη θεραπευτική της νευροεκφύλισης και τα λιπιδωμικά ναοφάρμακα. Παράλληλα, λοιπόν, με τις τρέχουσες εξελίξεις και διαβλέποντας το μέλλον στον χώρο της Φαρμακευτικής, η GPSF επιδιώκει την προώθηση και την ενημέρωση σε φλέγοντα ζητήματα, όπως είναι και η Εξατομικευμένη Ιατρική.

Η GPSF μπορεί επίσης να συμμετέχει και σε άλλα συνέδρια, όπως το Πανόραμα Σταδιοδρομίας και Επιχειρηματικότητας, που αφορά τον επαγγελματικό προσανατολισμό των νέων, ενώ το 2019 παρέστη ως χορηγός επικοινωνίας στο 1ο Συνέδριο Εξατομικευμένης Ιατρικής. Μέλη της Ένωσης μπορούν να συμμετέχουν και με παρουσιάσεις συνεδρίων, όπως συνέβη στο 24ο και 25ο ΕΣΦΙΕ, με θεματολογίες που αφορούσαν τα φυσικά προϊόντα και τα βιο-ομοειδή αντίστοιχα. Ο συνδυασμός και η συνεργασία των επιστημών της Ιατρικής και της Φαρμακευτικής συνιστούν κύριο μέλημα της GPSF, συνεπώς ένας πυλώνας τον οποίο θέλει να προωθήσει ενεργά είναι και η Φαρμακογονιδιωματική.

Για την εξοικείωση των φοιτητών με τη βιβλιογραφική αναζήτηση και την ανάγνωση επιστημονικών άρθρων, έχουν θεσπιστεί δύο σημαντικές δράσεις. Αρχικά, τα 'Fun Facts', τα οποία συνιστούν ενημερωτικά άρθρα μικρής έκτασης, που έχουν συντάξει μέλη της ομάδας, παρουσιάζουν ενδιαφέροντα θέματα που άπτονται δια-



Εικόνα 1. Εθελοντές την GPSF και μέλη της Οργανωτικής Επιτροπής στο PharmACon, Πολεμικό Μουσείο Αθηνών, Οκτώβριος 2018.

φόρων κλάδων, όπως είναι για παράδειγμα η Βιοπληροφορική [6], η ανάλυση του γονιδιώματος, το τρισδιάστατα εκτυπωμένο σκεύασμα αλλά και νέες μελέτες στην φαρμακευτική επιστήμη που αποτελούν μερικά από τα περιεχόμενά τους. Η δεύτερη δράση δημιουργήθηκε ώστε να υπάρχει συνεχής ενημέρωση σχετικά με το ευρύ φάσμα αντικειμένων γύρω από την Φαρμακευτική. Έτσι, κάθε μήνα αναρτώνται Επιστημονικά Άρθρα διάφορης θεματολογίας στην ιστοσελίδα της GPSF [7]. Η δράση εγκαινιάστηκε το 2020 σε συνεργασία με μέλη ΔΕΠ από τα Τμήματα Φαρμακευτικής Αθηνών και Πατρών έχοντας ως στόχο τη συμμετοχή όλων των τμημάτων. Μέσω της συγκεκριμένης δράσης γνωστοποιούνται ερευνητικά προγράμματα στα οποία δραστηριοποιούνται τα ελληνικά Πανεπιστήμια, ένα από τα οποία είναι και το Φορητό Εργαστήριο Μοριακής Βιολογίας (2MoBiL) που συντονίζεται από το Εργαστήριο Φαρμακογονιδιωματικής και Εξατομικευμένης Θεραπείας του Τμήματος Φαρμακευτικής του Πανεπιστημίου Πατρών [8]. Προσφέρεται, με τον τρόπο αυτό, η δυνατότητα στους φοιτητές να έρθουν σε επαφή με τις επιστημονικές ανακαλύψεις και τις πιο πρόσφατες εξελίξεις.

Ακόμη, τα μέλη της GPSF μετέχουν σε μια πληθώρα εθελοντικών δράσεων [9]. Έχοντας ως στόχο την εθελοντική προσφορά, έχουν καθιερωθεί δράσεις ανθρωπιστικού περιεχομένου, όπως η ετήσια συλλογή

τροφίμων και φαρμάκων αλλά και η διοργάνωση ή η συμμετοχή σε αθλητικά δρώμενα. Δράσεις που προάγουν την αλληλοβοήθεια περιλαμβάνονται και σε μελλοντικά πλάνα της ομάδας. Μια ομάδα με υπόσταση, αντανάκλαση και επιρροή στην κοινωνία οφείλει να συμβάλλει ενεργά στο κοινωνικό σύνολο.

Δράσεις της GPSF στην Φαρμακογονιδιωματική και Εξατομικευμένη Θεραπεία

Ο κλάδος της βιοπληροφορικής αποτελεί την επιστημονική περιοχή όπου η Βιολογία συναντά όχι μόνο την Πληροφορική, αλλά την Μαθηματική Επεξεργασία και τη Στατιστική Ανάλυση, παρέχοντας στους επιστήμονες νέες προσεγγίσεις στις βασικές αρχές της Βιολογίας και καινοτόμες μεθόδους επίλυσης σύγχρονων βιολογικών προβλημάτων. Με άξονα την συλλογή διαφόρων στοιχείων από μια πληθώρα διαφορετικών γνωστικών αντικειμένων, η Βιοπληροφορική και Γονιδιωματική Πληροφορική χρησιμοποιεί υπολογιστικές μεθόδους και μαθηματικές τεχνικές με στόχο την ανάλυση και περιγραφή Βιολογικών συστημάτων. Πιο αναλυτικά, ο τομέας της Βιοπληροφορικής και Γονιδιωματικής Πληροφορικής παρουσιάζει ένα στοχευμένο αλλά παράλληλα ευρύ πεδίο εφαρμογών όχι μόνο στην μοριακή αλλά και στην περιβαλλοντική, πληθυσμιακή βιολογία καθώς και στους τομείς της φαρμα-

κογονιδιωματικής και εξατομικευμένης θεραπείας. Με την ταυτόχρονη άνθιση του επιστημονικού τομέα της Εξατομικευμένης Ιατρικής, η Βιοπληροφορική και Γονιδιωματική Πληροφορική αποκτούν νέες διαστάσεις. Αντιμετωπίζουν επιστημονικά ερωτήματα, όπως τη διαχείριση των μεγάλων όγκων δεδομένων που προκύπτουν (Big Data), ενώ παράλληλα διαδραματίζουν καθοριστικό ρόλο σε εφαρμογές του τομέα, όπως η πρόβλεψη προτύπων γονιδιακής έκφρασης και δομών αγνώστων πρωτεϊνών, ο σχεδιασμός και η ανακάλυψη νέων φαρμάκων, καθώς και η εξατομίκευση της προτεινόμενης θεραπείας, είτε με την δημιουργία αλγορίθμων συσχέτισης γονιδιακών παραλλαγών και συνιστώμενης θεραπευτικής δόσης φαρμάκου, είτε μέσω άλλων μεθόδων. Τα προαναφερθέντα καθιστούν την επαφή με τον τομέα της Βιοπληροφορικής επιτακτική ανάγκη τόσο για τον μελλοντικό φαρμακοποιό όσο και για όλους τους θετικούς επιστήμονες και επιστήμονες των κλάδων της υγείας.

Η GPSF, αναλογιζόμενη την ανάγκη επιμόρφωσης των φοιτητών στο νέα αυτά αντικείμενα, βρίσκεται στην ευχάριστη θέση να οργανώσει μία σειρά διαδικτυακών ομιλιών με την Δρ Κορομηνά Μαρία, ερευνήτρια στη Βιοπληροφορική και τη Στατιστική Γενετική στο Εργαστήριο Φαρμακογονιδιωματικής και Εξατομικευμένης Θεραπείας, Τμήμα Φαρμακευτικής, Πανεπιστήμιο Πατρών. Η κα Κορομηνά, είναι διακεκριμένη ερευνήτρια στον τομέα της, και με τρόπο εύληπτο και κατανοητό θα μεταδώσει στους συμμετέχοντες στις βασικές αρχές της Βιοπληροφορικής.

Συμπεράσματα

Η Φαρμακευτική, όπως και κάθε επιστήμη, εξελίσσεται και η ανάγκη για μια περισσότερο ανθρωποκεντρική προσέγγιση γίνεται όλο και πιο προφανής. Όπως προ-

αναφέρθηκε, η συνεργασία όλων των επαγγελματιών Υγείας αποτελεί ύψιστη προτεραιότητα ώστε να μπορέσει να δοθεί στους ασθενείς η βέλτιστη αγωγή και φροντίδα, μειώνοντας τα έξοδα θεραπείας αλλά και τις περιπτώσεις λανθασμένης χορήγησης φαρμακευτικής αγωγής. Οι εξελίξεις αυτές δεν αφορούν μόνο νέα επιστημονικά επιτεύγματα, αλλά σχετίζονται και με την ίδια τη φύση του επαγγέλματος. Εκτός από τα επιστημονικά εφόδια που πρέπει να διαθέτει ο Φαρμακοποιός ώστε να μπορεί να ανταποκριθεί βέλτιστα στις ανάγκες των ασθενών, θα πρέπει να αναπτύξει και τις λεγόμενες οριζόντιες δεξιότητες (soft skills), τις προσωπικές και κοινωνικές του δεξιότητες.

Όλες αυτές οι δράσεις που προαναφέρθηκαν σχεδιάζονται και υλοποιούνται, στο μεγαλύτερο βαθμό, από τα μέλη της ομάδας, με τον Υπεύθυνο Εκπαιδευτικών Προγραμμάτων να φροντίζει για το συντονισμό τους.

Η ομάδα διαπνέεται από διάθεση προς εξωστρέφεια κι ένα σύστημα αξιών, το οποίο καθορίζει τις πράξεις της. Έχει συνάψει στο παρελθόν αξιόλογες συνεργασίες και προβλέπεται να καλωσορίσει νέους συνεργάτες σύντομα. Το ευρύ φάσμα δραστηριοτήτων που περιλαμβάνεται στην ατζέντα της, σε συνδυασμό με τους νέους ορίζοντες δραστηριοποίησης που ανοίγονται στο άμεσο μέλλον, δίνει το χώρο στα μέλη της για έκφραση και απασχόληση σε ποικίλους τομείς ευρείας αλλά και επίκαιρης θεματολογίας. Την ομάδα απαρτίζουν άνθρωποι με θέληση για ανιδιοτελή προσφορά, που εργάζονται με αφοσίωση, συνέπεια και αγάπη για αυτή, με οδηγό τις αρχές και τους σκοπούς της. Συνδυάζοντας αυτά τα χαρακτηριστικά με την επιστήμη της Φαρμακευτικής, επιτυγχάνονται σπουδαία αποτελέσματα, όπως αναφέρει χαρακτηριστικά και η φράση που συνοδεύει την ομάδα: 'When science meets creativity, great things happen!' ●

ABSTRACT

The Greek Pharmaceutical Student's Federation (GPSF) meets Pharmacogenomics and Personalized Medicine

Tatiana Grammatikopoulou¹, Dimitra Dedousi², Maria Therianou², Iris Tripolitsioti¹

¹National and Kapodistrian University of Athens, School of Pharmacy, Athens, Greece,

²University of Patras School of Health Sciences, Department of Pharmacy, Laboratory of Pharmacogenomics and Individualized Therapy, Patras, Greece

The Greek Pharmaceutical Student's Federation is a voluntary group active both nationally, through its members belonging to the Departments of Pharmacy, and in Europe, with its participation in the European Pharmaceutical Students' Association. Its main activities areas are of educational, scientific and social interest. At the same time team members engage in a variety of activities expanding their knowledge and improving their scientific and soft skills. The increasing

interest in the pharmaceutical discipline and its interplay with other scientific disciplines provides a wide range of topics a pharmacist can deal with. The field of Personalized Medicine and Pharmacogenomics is also a major area of interest and a core component in recent GPSF's activities. Lastly, in order to familiarize students with the emerging field of personalized medicine, a series of lectures on Bioinformatics is being organized.

KEY WORDS: Greek Pharmaceutical Students Federation, Pharmacy, Personalized Medicine, Bioinformatics

BIBΛΙΟΓΡΑΦΙΑ

1. Ένωση Φοιτητών Φαρμακευτικής Ελλάδας. Εισαγωγή. Διαθέσιμο <https://gpsf.gr/>. Επίσκεψη Δεκέμβριος 27, 2020.
2. European Pharmaceutical Students' Association. EPSA Memberships. Available via <https://www.epsa-online.org/how-to-apply-as-a-member/>. Accessed December 26, 2020.
3. Ένωση Φοιτητών Φαρμακευτικής Ελλάδας. Health Campaigns. Διαθέσιμο <https://gpsf.gr/category/health-campaigns/>. Επίσκεψη Δεκέμβριος 26, 2020.
4. Ένωση Φοιτητών Φαρμακευτικής Ελλάδας. Learn from the expert. Διαθέσιμο <https://gpsf.gr/?s=learn+from+the+expert>. Επίσκεψη Δεκέμβριος 26, 2020.
5. Ένωση Φοιτητών Φαρμακευτικής Ελλάδας. Επιστημονικές/Εκπαιδευτικές Εκδηλώσεις. Διαθέσιμο <https://gpsf.gr/category/επιστημονικές-εκδηλώσεις-συνέδρια/>. Επίσκεψη Δεκέμβριος 26, 2020.
6. Ένωση Φοιτητών Φαρμακευτικής Ελλάδας. Fun Facts. Διαθέσιμο <https://gpsf.gr/category/fun-facts/>. Επίσκεψη Δεκέμβριος 26, 2020.
7. Ένωση Φοιτητών Φαρμακευτικής Ελλάδας. Επιστημονικά Άρθρα. Διαθέσιμο <https://gpsf.gr/επιστημονικά-άρθρα/>. Επίσκεψη Δεκέμβριος 26, 2020.
8. Σιαμόγλου Σ, Πατρινός ΓΠ. Το φορητό Εργαστήριο Μοριακής Βιολογίας 2MoBiL στην εκπαίδευση μαθητών στο πεδίο της Μοριακής Γενετικής και Εξατομικευμένης Ιατρικής. *Εξατομικευμένη Ιατρική* 2019; 1(1): 17-25.
9. Ένωση Φοιτητών Φαρμακευτικής Ελλάδας. Κοινωνικές Εκδηλώσεις. Διαθέσιμο <https://gpsf.gr/κοινωνικές-εκδηλώσεις/>. Επίσκεψη Δεκέμβριος 26, 2020.

ΟΔΗΓΙΕΣ ΠΡΟΣ ΣΥΓΓΡΑΦΕΙΣ

1. Θεματολογία και αναγνωστικό κοινό

Το περιοδικό «Εξατομικευμένη Ιατρική» είναι ελληνόγλωσση κρινόμενη επιστημονική περιοδική έκδοση που στοχεύει στην διάδοση της ιδέας και των αρχών, των μορφών και των μέσων της Εξατομικευμένης Ιατρικής, Φαρμακογονιδιωματικής και Εξατομικευμένης Θεραπείας, και στην επαύξηση και βελτίωση του διεπιστημονικού διαλόγου. Δέχεται και δημοσιεύει κλινικά σημαντικές ερευνητικές εργασίες και άρθρα ανασκόπησης στους τομείς την Γονιδιωματικής και της Εξατομικευμένης Ιατρικής, περιλαμβάνοντας ενδεικτικά αλλά όχι αποκλειστικά την Φαρμακογονιδιωματική, τη γονιδιωματική πληροφορική, τη γονιδιωματική σπανίων νόσων και Καρκίνου, τη ψυχιατρική γονιδιωματική, την γονιδιωματική Δημόσιας Υγείας, την δεοντολογία στη Γονιδιωματική και την οικονομική αξιολόγηση τεχνολογιών και θεραπειών με γονιδιωματικές συνιστώσες.

2. Γλώσσα

Δημοτική Ελληνική, στην οποία πρέπει να είναι συγγεγραμμένες όλες οι υποβαλλόμενες εργασίες, συνοδευόμενες από τίτλο, συγγραφική ομάδα και φορείς προέλευσης, περιλήψη και λέξεις ευρετηρίου και στα Αγγλικά.

3. Τρόπος υποβολής

Η υποβολή γίνεται αποκλειστικά ηλεκτρονικά μέσω του οικείου ιστοτόπου www.permed.gr.

Η συντακτική επιτροπή και ο Διευθυντής Σύνταξης πραγματοποιούν την αρχική εξέταση της εργασίας και αν αυτή είναι κατάλληλη για το περιοδικό και η υποβολή έχει γίνει πλήρως και ορθά, το δοκίμιο αποστέλλεται στους αντίστοιχους κριτές. Κατά την υποβολή οι συγγραφείς μπορούν να ζητήσουν εξαίρεση κριτών. Η διαδικασία κρίσης ολοκληρώνεται σε 2-3 εβδομάδες με την υποβολή της εισήγησης των κριτών, με βάση την οποία ο Διευθυντής Σύνταξης αποφασίζει περί της δημοσίευσης της εργασίας. Τα σχόλια, η εισήγηση των κριτών και η τελική απόφαση κοινοποιούνται από τη Διεύθυνση Σύνταξης στους Συγγραφείς.

4. Πρωτοτυπία της εργασίας και πνευματικά δικαιώματα

Οι υποβληθείσες εργασίες πρέπει να είναι πρωτότυπες στο όλον τους, χωρίς τμήματα προηγούμενων δημοσιεύσεων και κατά την υποβολή τους δεν πρέπει να είναι υπό κρίση σε άλλο φορέα δημοσίευσης/περιοδικό. Η υποβολή πρέπει να έχει εγκριθεί από όλους τους συγγραφείς. Ο αντεπιστέλλον συγγραφέας είναι συνήθως το αρχαιότερο μέλος της συγγραφικής ομάδας που αναλαμβάνει τη γενική ευθύνη για όλες τις απόψεις της εργασίας (δεοντολογία, πρωτοτυπία, συναίνεση, διαχείριση δεδομένων και όλες τις διαστάσεις της Ορθής Ιατρικής Πρακτικής). Το περιοδικό αποποιείται κάθε ευθύνης σε περιπτώσεις αιτημάτων αποζημίωσης/επανορθώσεων. Όλοι οι συγγραφείς πρέπει να υπογράψουν τη δή-

λωση παραχώρησης πνευματικών δικαιωμάτων και επίσης πρέπει να έχουν συνεισφέρει στη εργασία σύμφωνα με τα καθοριζόμενα από την ICMJE (<http://www.icmje.org>).

5. Δήλωση αντικρουόμενων συμφερόντων

Κάθε συγγραφέας οφείλει να δηλώσει όλα εκείνα τα οικονομικά στοιχεία και ενδιαφέροντα που τον αφορούν και που σχετίζονται με την μελέτη καθώς μπορούν να οδηγήσουν σε σύγκρουση συμφερόντων. Η προέλευση της χρηματοδότησης/χορηγίας της μελέτης πρέπει να γνωστοποιείται.

Αν δεν υφίσταται θέμα αντικρουόμενων συμφερόντων, αυτό πρέπει να δηλώνεται σαφώς στην υποβαλλόμενη εργασία πριν από τις Αναφορές ως ακολούθως: "Οι συγγραφείς δηλώνουν απουσία αντικρουόμενων συμφερόντων".

6. Δεοντολογία έρευνας και συμμόρφωση

Η Εξατομικευμένη Ιατρική ακολουθεί τις οδηγίες της Διεθνούς Επιτροπής Συντακτών Ιατρικών Περιοδικών (www.icmje.org). Για όλα τα πρωτότυπα ερευνητικά άρθρα απαιτούνται δηλώσεις (i) για έγκριση της έρευνας από την εκάστοτε επιβλέπουσα επιτροπή δεοντολογίας, (ii) για τη διεξαγωγή της έρευνας σύμφωνα με τα δεοντολογικά πρότυπα τα προβλεπόμενα από τη Διακήρυξη του Ελσίνκι και (iii) για την απόκτηση έγγραφης ενημέρωσης συγκατάθεσης από τους μετέχοντες ασθενείς. Για μελέτες με πειραματόζωα απαιτείται και η ακόλουθη δήλωση: «Τηρήθηκαν όλες οι διεθνείς, εθνικές ή/και εταιρικές ισχύουσες προβλέψεις για τη φροντίδα και μεταχείριση ζώων».

7. Άδειες και Λογοκλοπή

Για τη χρήση απεικονιστικού υλικού ήδη δημοσιευμένου απαιτείται η προσκόμιση από τους συγγραφείς γραπτής άδειας από τον/τους κατόχους των πνευματικών δικαιωμάτων και η παρουσίασή της κατά το στάδιο της υποβολής. Η λογοκλοπή δεν γίνεται ανεκτή σε καμία περίπτωση. Εξειδικευμένο λογισμικό χρησιμοποιείται για την ανίχνευσή της και δοκίμια που εμφανίζουν τέτοια φαινόμενα επιστρέφονται χωρίς κρίση στον αντεπιστέλλοντα συγγραφέα.

8. Τύποι εργασιών

Το περιοδικό «Εξατομικευμένη Ιατρική» δημοσιεύει τις κάτωθι κατηγορίες εργασιών:

Πρωτότυπα ερευνητικά άρθρα: Οι πρωτότυπες δημοσιεύσεις πρέπει να συνεισφέρουν στη διεύρυνση της Γνώσης στο πεδίο της φαρμακογονιδιωματικής και της γονιδιωματικής και εξατομικευμένης θεραπείας. Τα συμπεράσματα πρέπει να είναι λογικά και να υποστηρίζονται από στατιστική ανάλυση. Όταν εξετάζεται η ακρίβεια διαγνωστικού δοκιμίου, ενθαρρύνεται η συμμόρφωση με τα Πρότυπα Αναφοράς Διαγνωστικής Ακρίβειας/Standards for Reporting of Diagnostic Accuracy (STARD) και με το αντίστοιχο

ροϊκό διάγραμμα. (<http://www.stard-statement.org>). Οι προδιαγραφές του άρθρου είναι: απλή περιληψη ως 250 λέξεων, 3-5 λέξεις ευρετηρίου, κείμενο ως 10,000 λέξεις, ως έξι εικόνες και ως έξι πίνακες και βιβλιογραφία ως 70 αναφορές.

Άρθρα ανασκόπησης: Το περιοδικό αποδέχεται συστηματικές ανασκοπήσεις, βιβλιογραφικές ανασκοπήσεις επί ενός θέματος και μετααναλύσεις.

Οι προδιαγραφές της Ανασκόπησης είναι: απλή περιληψη ως 250 λέξεων, 3-5 λέξεις ευρετηρίου, κείμενο ως 12,000 λέξεις, ως οκτώ εικόνες και ως έξι πίνακες και βιβλιογραφία ως 120 αναφορές.

Επιστολές στην σύνταξη: Η επικοινωνία με την διεύθυνση σύνταξης είναι ευπρόσδεκτη και δημοσιεύεται αν προφέρει εποικοδομητική και συνεκτική κριτική σε δημοσιευμένα άρθρα του περιοδικού με βάση της διακριτική ευχέρεια της Σύνταξης. Πρέπει να υποβάλλονται το αργότερο τρεις μήνες μετά από την ψηφιακή δημοσίευση του άρθρου ενδιαφέροντος. Μετά την αποδοχή, η υποβολή τίθεται υπόψη της συγγραφικής ομάδας του άρθρου ενδιαφέροντος για απάντηση. Οι προδιαγραφές είναι κείμενο ως 800 λέξεις, ως δύο εικόνες και βιβλιογραφία ως 10 αναφορές. Επιτρέπονται ως 5 συγγραφείς και δεν απαιτούνται περιληψη και λέξεις ευρετηρίου.

Σχόλια/Άρθρα γνώμης/Συνηντεύξεις/Βιβλιοπαρουσιάσεις/Ανασκοπήσεις συνεδρίων/Παρουσιάσεις ιδρυμάτων: Τα άρθρα αυτά έχουν σαν σκοπό τη γνωστοποίηση της άποψης διεθνούς φήμης ακαδημαϊκών και ερευνητών σε θέματα επιστημονικής επικαιρότητας, δημοσίου διαλόγου ή μείζονος σημασίας τρεχουσών ανακαλύψεων στους τομείς της φαρμακογονιδιωματικής αι την γονιδιωματικής και εξατομικευμένης Ιατρικής. Συγγράφονται μόνο κατόπιν πρόσκλησης και οι προδιαγραφές είναι κείμενο ως 1,500 λέξεις, ως δύο εικόνες και βιβλιογραφία ως 10 αναφορές. Επιτρέπονται ως 3 συγγραφείς και δεν απαιτούνται περιληψη και λέξεις ευρετηρίου.

9. Δομή δοκιμίου

Ένα δοκίμιο προς υποβολή πρέπει να ικανοποιεί τις κάτωθι προδιαγραφές

I. Συνοδευτική επιστολή: Κάθε δοκίμιο πρέπει να συνοδεύεται από επιστολή υπογεγραμμένη από τον αντεπιστέλλοντα συγγραφέα εκ μέρους όλων των συγγραφέων δηλώνοντας ότι το υποβληθέν άρθρο δεν έχει υποβληθεί για δημοσίευση και σε άλλο περιοδικό. Σε περίπτωση αναθεωρημένων (revised) άρθρων, μια αναλυτική λίστα με τις διορθώσεις/παρεμβάσεις σύμφωνα με τα σχόλια των κριτών θα πρέπει να υποβάλλεται μαζί με τη συνοδευτική επιστολή.

II. Σελίδα τίτλου: Περιλαμβάνει τον τίτλο του δοκιμίου, τα ονόματα των συγγραφέων και τα ιδρύματα από τα οποία προέρχονται οι συγγραφείς, τις διευθύνσεις ηλεκτρονικού ταχυδρομείου όλων των συγγραφέων, και το ίδρυμα, διεύθυνση, τηλέφωνο και διεύθυνση ηλεκτρονικού ταχυδρομείου του συγγραφέα αλληλογραφίας.

III. Αταυτοποιητό δοκίμιο

1. Τίτλος: Αταυτοποιητή σελίδα τίτλου που αναφέρει μόνο τον τίτλο και όχι συγγραφείς και στοιχεία υπαγωγής.

2. Περιληψη: Η περιληψη περιλαμβάνει τα σημαντικότερα ευρήματα και συμπεράσματα της εργασίας.

3. Λέξεις ευρετηρίου: Μετά την περιληψη απαιτούνται 3 ως 5 λέξεις ευρετηρίου, οι οποίες πρέπει να επιλέγονται από τη βάση δεδομένων Επικεφαλίδων Ιατρικής Θεματολογίας/ Medical Subject Headings (MeSH) της Εθνικής Ιατρικής Βιβλιοθήκης των ΗΠΑ.

4. Δομή σώματος κειμένου: Το κείμενο των πρωτότυπων άρθρων πρέπει να έχει Εισαγωγή, Υλικά και Μεθόδους, Αποτελέσματα και Συζήτηση. Των άρθρων ανασκόπησης πρέπει να έχει Εισαγωγή, Συζήτηση και Μελλοντικές Προοπτικές, ενώ η ανάπτυξη των υπολοίπων παραγράφων επαφίεται καθ' ολοκληρίαν στους συγγραφείς.

Γραμματοσειρά: Η προτεινόμενη επιλογή είναι η Times New Roman 12 στιγμών (12 pt) σε διπλό διάστημα και με εμφανείς τις αριθμήσεις σειράς και σελίδας.

Συνημμένες: Πρέπει να χρησιμοποιούνται όσο το δυνατόν λιγότερο, αλλά σε περίπτωση χρήσης ορίζονται την πρώτη φορά που αναφέρονται στο κείμενο με την ολογραφική μορφή να ακολουθείται από το ακρωνύμιο ή την σύντμηση εντός παρένθεσης.

Ευχαριστίες, χορηγίες και χρηματοδότηση: Οι Ευχαριστίες βρίσκονται στο τέλος του σώματος του κειμένου, πριν από τον βιβλιογραφικό κατάλογο/αναφορές. Η οποιαδήποτε χρηματοδότηση ή χορηγία από φαρμακευτικές εταιρείες, κατασκευαστές ιατρικού εξοπλισμού και άλλες εταιρείες ή επιχειρήσεις των οποίων προϊόντα ή υπηρεσίες έχουν χρησιμοποιηθεί στη μελέτη πρέπει να δηλώνονται στην Δήλωση Αντικρουόμενων συμφερόντων αλλά και στις Ευχαριστίες.

Μονάδες μέτρησης: Όλες οι μετρήσεις πρέπει να αποδίδονται στο Διεθνές Σύστημα (SI). Η υποδιαστολή αποδίδεται με κόμμα (π.χ. 3,5 cm) με κενά εμπρός και πίσω από το σύμβολο συν/πλην (π.χ. 13,6 ± 1,2). Δεν πρέπει να υπάρχουν κενά εκατέρωθεν ενδεικτών κύμανσης (π.χ. 15-20) ή συμβόλων ισότητας/ανισότητας (π.χ. $r=0,37$, $p<0,005$).

10. Εικόνες και Πίνακες

Όλες οι εικόνες και οι πίνακες πρέπει να αναφέρονται στο κείμενο κατά σειράν εμφανίσεως με αραβικά νούμερα εντός παρενθέσεως, π.χ. (Εικ. 1) και (Πιν. 1). Τμήματα εικόνων αναφέρονται με πεζά γράμματα π.χ. (Εικ. 1a).

Οι εικόνες πρέπει να είναι υψηλής ανάλυσης. Ανυσματικά γραφήματα, σαρωμένα και πρωτότυπα γραμμικά σχέδια πρέπει να υποβάλλονται σε μορφή bitmap και σε ελάχιστη ανάλυση 1,200 στιγμών (dpi). Φωτογραφίες, σχέδια και ζωγραφιές πρέπει να υποβάλλονται σε μορφή TIFF ή JPEG, σε διαστάσεις 174 mm πλάτος και 234 mm ύψος και σε ελάχιστη ανάλυση 300 στιγμών (dpi).

Η ανωνυμία των ασθενών πρέπει να διασφαλίζεται. Όλα τα δεδομένα που μπορούν να οδηγήσουν σε ταυτοποίηση (ονόματα, αρχικά, αριθμοί ταυτότητας/ταυτοποίησης/καταγραφής) πρέπει

να μην εμφανίζονται στο κείμενο στις εικόνες και σε πίνακες. Αν είναι απαραίτητο να εμφανίζεται το πρόσωπο ασθενούς, πρέπει να μαυρίζονται τα μάτια. Αν είναι δυνατή η ταυτοποίηση του ασθενούς από φωτογραφία, λεζάντα ή αντίστοιχο μέρος του κειμένου, πρέπει να συνοποβάλλεται γραπτώς η συγκατάθεσή του.

Οι λεζάντες των εικόνων πρέπει να περιλαμβάνονται σε ξεχωριστή σελίδα του κειμένου ενώ των Πινάκων πρέπει να αναγράφονται πάνω από τον κάθε πίνακα. Ερμηνευτικά σύμβολα (αστερίσκοι, βέλη κλπ) πρέπει να χρησιμοποιούνται όταν τα απεικονιστικά δεδομένα είναι ασαφή, να μην χρησιμοποιούνται σε έγχρωμη οροφή αλλά σε άσπρο, μαύρο και τόνους του γκρι και σε μέγεθος που να αναλογεί με αυτό της εικόνας.

Οι πίνακες πρέπει να υποβάλλονται μετά από τον βιβλιογραφικό κατάλογο, σε ξεχωριστή σελίδα ο καθένας και με αραβική αρίθμηση και τον τίτλο τους σε μορφή επικεφαλίδας, ενώ τυχούσες συντμήσεις πρέπει να επεξηγούνται σε υποσημείωση. Κάθε στήλη του πίνακα πρέπει να έχει κεφαλίδα που να περιγράφει το πληροφοριακό της περιεχόμενο.

11. Βιβλιογραφία

Η ακρίβεια της βιβλιογραφικής τεκμηρίωσης βαρύνει αποκλειστικά τους συγγραφείς. Ο κατάλογος βιβλιογραφίας πρέπει να περιλαμβάνει μόνο εργασίες που αναφέρονται εντός του κειμένου και είτε έχουν δημοσιευτεί είτε έχουν γίνει αποδεκτές προς δημοσίευση. Προσωπική επικοινωνία και αδημοσίευτες εργασίες απλά αναφέρονται εντός του κειμένου και μόνο. Δεν πρέπει να χρησιμοποιούνται σημειώσεις ως υποκατάστατο βιβλιογραφικού καταλόγου.

Η καταλογογράφηση γίνεται με αλφαβητική σειρά από το επίθετο του πρώτου συγγραφέα κάθε αναφερόμενης εργασίας. Προκειμένου περί κοινού πρώτου συγγραφέα η καταλογογράφηση γίνεται βάσει του αρχικού του επωνύμου του δεύτερου, τρίτου κ.ο.κ. συγγραφέα, ενώ για πανομοιότυπη συγγραφική ομάδα χρησιμοποιείται απόλυτη χρονολογική σειρά με επιθέματα a, b, c.

Οι εργασίες που κατά την υποβολή υπάρχουν μόνο σε μορφή ψηφιακής προδημοσίευσης πρέπει να αναφέρονται με το αντίστοιχο αριθμό DOI (digital object identifier).

Ως έξι συγγραφείς αναγράφονται πλήρως με επώνυμο και αρχικά ονόματος/-ων. Αν είναι περισσότεροι αναγράφονται οι 3 πρώτοι και ακολουθεί το 'et al.'

Για κεφάλαια βιβλίων αναγράφονται ο τίτλος και οι συγγραφείς του κεφαλαίου, ο τίτλος του βιβλίου και ο/οι επιμελητής/-ές έκδοσης, η πόλη, χώρα, εκδοτικός οίκος, χρονολογία έκδοσης και οι σελίδες.

Για αποκλειστικώς δικτυακά διαθέσιμα αρχεία πρέπει να αναγράφονται οι συγγραφείς αν υπάρχουν, ο τίτλος της ιστοσελίδας, η δικτυακή διεύθυνση URL και η ημερομηνία τελευταίας επίσκεψης των συγγραφέων στην ιστοσελίδα.

Παραδείγματα βιβλιογραφικών αναφορών:

Άρθρα περιοδικών:

Özdemir V, Patrinos GP. David Bowie and the Art of Slow Innovation: A Fast-Second Winner Strategy for Biotechnology and Precision Medicine Global Development. OMICS 2017; 21(11): 633-637.

ή
Sosnay PR, Siklosi KR, Van Goor F, et al. Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. Nat Genet 2013; 45(10): 1160-1167.

ή
El Rouby N, McDonough CW, Gong Y, et al. Genome-wide association analysis of common genetic variants of resistant hypertension. Pharmacogenomics J. 2018; doi: 10.1038/s41397-018-0049-x. Epub 2018 Sep 20.

Βιβλία:

Correa-Lopez C, Patrinos GP (eds). Genomic Medicine in developing and emerging economies. Elsevier/Academic Press, Burlington, CA, USA, 2018.

Κεφάλαια βιβλίων:

Squassina A, Manchia M, Mitropoulou C, et al. An introduction to Pharmacogenomics and Personalized Medicine, In Lanzer P (ed) PanVascular Medicine (2nd edition), Springer, Heidelberg, Germany, 2015, pp. 1053-1065.

Πρακτικά συνεδρίων:

Innocenti F. Safety of Drug Treatment in Oncology: A Genomic Perspective. 2018 Golden Helix Summer School, Public Health Genomics 2018; 21(Suppl 1): 2.

Ψηφιακά έγγραφα:

National Institute for Health and Care Excellence). SIR-Spheres for treating inoperable hepatocellular carcinoma. Available via nice.org.uk/guidance/mib63. Published May 10, 2013. Accessed January 25, 2014.

Σε περίπτωση αμφιβολίας, αναγράψτε τον πλήρη τίτλο του περιοδικού.

12. Διαδικασία Κρίσης και αναθεώρησης υποβληθεισών εργασιών

Η αποδοχή προς δημοσίευση αποφασίζεται από τον Διευθυντή Σύνταξης βάσει των εισηγήσεων των κριτών. Οι συγγραφείς οφείλουν να ολοκληρώσουν τις απαιτούμενες διορθώσεις στα τελικά δοκίμιο εντός 5 ημερών από την αποστολή αυτών σε μορφή αρχείου pdf, οι οποίες περιλαμβάνουν χωρίς να περιορίζονται σε έλεγχο της αρτιότητας του κειμένου, αποδοχή ορθογραφικών, γραμματικών και συντακτικών διορθώσεων και έλεγχο της εισαγωγής και ορθής αρίθμησης και αναφοράς πινάκων και εικόνων. Μετά την ψηφιακή δημοσίευση της εργασίας δεν είναι εφικτές διορθώσεις, εκτός από τη δημοσίευση διορθώσεων (Erratum) σε επόμενα τεύχη. ●