



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ**  
**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**  
**ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ**

Συστήματα τεχνητής νοημοσύνης στην διάγνωση και θεραπεία  
του καρκίνου

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

της

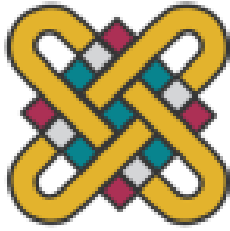
**ΕΥΣΤΑΘΙΑΣ ΜΠΕΛΤΑΟΥ**

(ΑΕΜ: 1473 )

**Επιβλέπων : ΣΙΝΑΤΚΑΣ ΙΩΑΝΝΗΣ**  
**ΚΑΘΗΓΗΤΗΣ**

Καστοριά Δεκέμβριος - 2023





**ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ**  
**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**  
**ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ**

**Συστήματα τεχνητής νοημοσύνης στην  
διάγνωση και θεραπεία του καρκίνου**

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

της

**ΕΥΣΤΑΘΙΑΣ ΜΠΕΛΤΑΟΥ**

(ΑΕΜ: 1473 )

**Επιβλέπων : ΣΙΝΑΤΚΑΣ ΙΩΑΝΝΗΣ**  
**ΚΑΘΗΓΗΤΗΣ**

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 08/12/2023

.....  
ΣΙΝΑΤΚΑΣ ΙΩΑΝΝΗΣ  
ΚΑΘΗΓΗΤΗΣ

.....  
ΦΩΤΙΑΔΗΣ ΔΗΜΗΤΡΙΟΣ  
ΕΠΙΚΟΥΡΟΣ ΚΑΘΗΓΗΤΗΣ

.....  
ΑΓΓΕΛΗΣ ΣΤΥΛΙΑΝΟΣ  
Ε.Τ.Ε.Π.

Καστοριά Δεκέμβριος - 2023

Copyright © 2023 – ΕΥΣΤΑΘΙΑ ΜΠΕΛΤΑΟΥ

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν αποκλειστικά τον συγγραφέα και δεν αντιπροσωπεύουν τις επίσημες θέσεις του Πανεπιστημίου Δυτικής Μακεδονίας.

Ως συγγραφέας της παρούσας εργασίας δηλώνω πως η παρούσα εργασία δεν αποτελεί προϊόν λογοκλοπής και δεν περιέχει υλικό από μη αναφερόμενες πηγές.

## Ευχαριστίες

Θέλω να ευχαριστήσω τον επιβλέποντά μου Καθηγητή Ιωάννη Σινάτκα για την εμπιστοσύνη και τη τιμή που έδειξε προς το πρόσωπό μου, μέσω της ανάθεσης αυτής της εργασίας. Με την υπομονή του, την επιμονή του, τη μεθοδικότητα και το προσωπικό του ενδιαφέρον, κατάφερα να συντάξω αυτή την εργασία. Ήταν σύμβουλος και δάσκαλός μου.

## Περίληψη

---

Η τεχνητή νοημοσύνη (AI) και ειδικότερα ο τομέας της που αφορά στην βαθιά μάθηση μεταξύ άλλων βρίσκει ευρεία εφαρμογή στην ιατρική και ειδικότερα στην παθολογοανατομία. Σκοπός της εργασίας είναι να αναδείξει την εφαρμογή της στον τομέα αυτό και ως μεθοδολογία βασίζεται στην βιβλιογραφική κριτική ανασκόπηση.

Τα αποτελέσματα της έρευνας δείχνουν ότι στράφηκε η διάγνωση στην τεχνολογία της AI γιατί αφενός μεν γιατί η αύξηση των περιστατικών καρκίνου και η θνησιμότητα ήταν μεγάλη και αφετέρου η τεχνολογία της AI ήταν ελπιδοφόρα.

Η έρευνα εστιάζεται και αφορά στη συμβολή της τεχνολογίας και ως εκ τούτου αντιμετώπισε προβλήματα και περιορισμούς που αφορούν στο ιατρικό μέρος της. Για παράδειγμα αντιμετωπίστηκαν ασάφειες σχετικά με περιγραφές μεθόδων, διακρίσεις μεταξύ στρατηγικών και τεκμηριωμένων παρεμβάσεων, τακτικές αξιολόγησης αποτελέσματος καθώς και απροσδιόριστο επίπεδο για το χαρακτηρισμό της επιτυχίας.

Η εργασία συστηματοποιεί και οργανώνει την υφιστάμενη γνώση και εμπειρία με στόχο η τεχνητή νοημοσύνη να βοηθήσει στον τομέα της διάγνωσης και κατηγοριοποίησης του καρκίνου με ταχύτερα τρόπο και με διαρκώς αυξανόμενο ποσοστό επιτυχίας αποτελεσμάτων.

*Λέξεις Κλειδιά: Τεχνητή νοημοσύνη (AI), διαγνωστικά, μηχανική μάθηση, καρκίνος, ογκολογία, παθολογία, ακτινολογία*

## Abstract

---

Artificial intelligence (AI) and in particular its field related to deep learning among others finds wide application in medicine and especially in pathology. The purpose of the work is to highlight its application in this field and as a methodology it is based on the bibliographic critical review.

The research results show that diagnosis turned to AI technology because, on the one hand, the increase in cancer incidence and mortality was high, and on the other hand, AI technology was promising.

The research focuses and concerns the contribution of technology and therefore faced problems and limitations related to its medical part. For example, ambiguities regarding method descriptions, distinctions between strategic and evidence-based interventions, outcome evaluation tactics, and an unspecified standard for characterizing success were addressed.

The work systematizes and organizes existing knowledge and experience with the goal of artificial intelligence helping the field of cancer diagnosis and categorization in a faster way and with an ever-increasing success rate of results.

**Keywords:** Artificial intelligence (AI), diagnostics, machine learning, cancer, oncology, pathology, radiology

## Πίνακας Περιεχομένων

---

Εισαγωγή.....	10
1. Υφιστάμενη κατάσταση .....	13
1.1 Διαδικασία ΑΙ στην κλινική έρευνα για τον καρκίνο.....	15
1.2 Επισκόπηση της τεχνητής νοημοσύνης στην ογκολογία.....	17
1.2.1 Τύποι δεδομένων: Ηλεκτρονικά Μητρώα Υγείας.....	20
1.3 Κλινικές Εφαρμογές .....	26
1.3.1 Στρωματοποιημένος προσυμπτωματικός έλεγχος κινδύνου ασυμπτωματικών ασθενών .....	26
1.3.2 Συμπτωματική Διαλογή Ασθενούς.....	27
2. Τρέχουσες προσπάθειες εξειδίκευσης και γνώσης τομέα .....	29
2.1 Προσπάθειες της εμπειρογνωμοσύνης .....	29
2.1.1 Ενσωματωμένη γνώση του τομέα .....	30
2.1.2 Δομική μάθηση βασισμένη σε βιο-δίκτυα .....	35
2.2 ΑΙ στην κλινική έρευνα για τον καρκίνο: μέθοδος και εφαρμογή .....	35
2.2.1 Αναγνώριση απεικόνισης καρκίνου .....	36
2.2.2 Γονιδιωματική ανάλυση.....	38
2.2.3 Προκλήσεις και μελλοντικές κατευθύνσεις .....	40
Συμπεράσματα .....	47
Βιβλιογραφία .....	49



## Λίστα Εικόνων

---

Εικόνα 1: Η τεχνητή νοημοσύνη και οι υποδιαίρεσεις της.....	18
Εικόνα 2: Παράδειγμα αρχιτεκτονικής ANN ενός κρυφού επιπέδου.....	21
Εικόνα 3: Κλινικές εφαρμογές της AI στην έγκαιρη διάγνωση του καρκίνου.....	26
Εικόνα 4: Αναγνώριση υπολογιστικής απεικόνισης για κλινική έρευνα καρκίνου.....	37
Εικόνα 5: Υπολογιστική ανάλυση που χρησιμοποιείται συνήθως στην έρευνα για τον καρκίνο στη γονιδιωματική.....	39

## Εισαγωγή

---

Η Τεχνητή Νοημοσύνη (AI) έχει εφαρμοστεί εκτενώς σε εργασίες σε ιατρικές ειδικότητες [1]. Η έλευση των τεχνολογιών τεχνητής νοημοσύνης στη βασική βιολογία, τη φαρμακολογία και την ιατρική οδήγησε σε πολλαπλές ανακαλύψεις επιδόσεων και πέτυχε επιδόσεις συγκρίσιμες με τους ειδικούς του ανθρώπου σε ορισμένους τομείς [2]. Σε μια έρευνα σχετικά με την επίδραση της τεχνητής νοημοσύνης, η τεχνητή νοημοσύνη αναμένεται να έχει σημαντικό αντίκτυπο σε πολλές δραστηριότητες σε τομείς όπως η υγεία και η επιστήμη, και υπάρχει 50% πιθανότητα η τεχνητή νοημοσύνη να ξεπεράσει τον άνθρωπο σε όλες τις εργασίες σε 45 χρόνια [3].

Οι πρώτες προσεγγίσεις της τεχνητής νοημοσύνης κυριαρχούνταν από συστήματα εμπειρογνομών που βασίζονται σε παραδοσιακά σύμβολα και βασισμένα σε πληροφορίες. Στη συνέχεια, η εμφάνιση της μηχανικής μάθησης (ML) έφερε επαναστατική πρόοδο στην τεχνητή νοημοσύνη. Η ML, ως παραδοσιακή τεχνολογία AI, παρέχει μια πληθώρα αλγορίθμων που μπορούν να βελτιώσουν την ακρίβεια προσδιορισμού ή πρόβλεψης με άφθονα, υψηλής ποιότητας δεδομένα [4].

Το Deep Learning (DL) είναι ένας τύπος αλγορίθμων που αναπτύχθηκε χρησιμοποιώντας μοντέλα νευρωνικών δικτύων για την επίλυση προβλημάτων που είναι δύσκολο να λυθούν με την παραδοσιακή ML. Από την πρώτη εφαρμογή στην αναγνώριση εικόνας το 2012 [5], το DL έχει γίνει η de facto προσέγγιση για την ανάλυση της όρασης υπολογιστή και έχει βελτιωθεί σημαντικά στην ακρίβεια της αναγνώρισης εικόνας από τότε [1]. Οι προσεγγίσεις DL ταξινομούνται με βάση την αρχιτεκτονική, τον τύπο των επιπέδων, τους αλγόριθμους ενημέρωσης του βάρους σύνδεσης, τον μηχανισμό ανάδρασης κ.λπ. Τα πιο συνηθισμένα μοντέλα DL είναι τα βαθιά νευρωνικά δίκτυα (DNN) [6], τα συνελκτικά νευρωνικά δίκτυα (CNN) [7] και επαναλαμβανόμενα νευρωνικά δίκτυα (RNNs) [8]. Αναφορικά με το τελευταίο πρόκειται για ένα επαναλαμβανόμενο νευρωνικό δίκτυο (RNN) είναι ένας από τους δύο ευρείς τύπους τεχνητού νευρωνικού δικτύου, που χαρακτηρίζεται από την κατεύθυνση της ροής πληροφοριών μεταξύ των στρωμάτων του. Σε αντίθεση με το μονοκατευθυντικό νευρωνικό δίκτυο τροφοδοσίας, είναι ένα αμφίδρομο τεχνητό νευρωνικό δίκτυο, πράγμα που σημαίνει ότι επιτρέπει στην έξοδο από ορισμένους κόμβους να επηρεάσει την επακόλουθη είσοδο στους ίδιους κόμβους [8].

Τις τελευταίες δύο δεκαετίες, η τεχνητή νοημοσύνη έχει ανακτήσει τη δημοτικότητά της και έχει δείξει υποσχέσεις σε όλες τις διαστάσεις της έρευνας για τον καρκίνο μαζί με την έκρηξη των ψηφιακών πληροφοριών. Με την ταχεία ανάπτυξη της ποικιλίας και του όγκου των συνόλων δεδομένων, ένας μεγάλος όγκος πληροφοριών όγκου σε μοριακό επίπεδο από ασθενείς με καρκίνο μπορεί να αποκτηθεί εύκολα.

Η αποστολή πολλών δημόσιων βάσεων δεδομένων είναι να επιτρέψουν την ανταλλαγή δεδομένων σε μελέτες καρκίνου για την υποστήριξη της ιατρικής ακριβείας. Για παράδειγμα, ο Άτλας του Γονιδιώματος του Καρκίνου (TCGA) [9], μια ευρέως χρησιμοποιούμενη δημόσια βάση δεδομένων για την έρευνα για τον καρκίνο, χαρακτηρίζει πάνω από 20.000 δείγματα καρκίνου που εκτείνονται σε 33 τύπους καρκίνου και δημιουργεί διάφορους τύπους δεδομένων και πόρων, συμπεριλαμβανομένων γονιδιωματικών δεδομένων και ψηφιακής αποθήκης διαφανειών.

Επιπλέον, η ανάπτυξη υπολογιστικού υλικού υψηλής απόδοσης, όπως οι μονάδες γραφικής επεξεργασίας (GPUs), έχει παράσχει την παράλληλη επεξεργαστική ισχύ για αλγόριθμους τεχνητής νοημοσύνης σε εντατικούς αριθμητικά υπολογισμούς, π.χ. πολλαπλά επίπεδα αφαίρεσης και εκατομμύρια υπολογιστικούς κόμβους. Ως αποτέλεσμα, η τεχνητή νοημοσύνη υπερέρχει στο χειρισμό μεγάλων όγκων και πολύπλοκων δεδομένων και στον εντοπισμό χαρακτηριστικών από τα δεδομένα, τα οποία ο ανθρώπινος εγκέφαλος δεν μπορεί να αναγνωρίσει.

Αν και η τεχνητή νοημοσύνη έχει ενσωματωθεί ταχέως στην ογκολογική έρευνα, η ανάπτυξη λύσεων τεχνητής νοημοσύνης είναι ακόμη στα σπάργαλα. Μόνο μερικές εφαρμογές που βασίζονται σε τεχνητή νοημοσύνη έχουν εγκριθεί για χρήση στην πράξη, π.χ. νοσοκομεία, φαρμακευτικές εταιρείες κ.λπ. Είναι ακόμα υπό συζήτηση εάν η τεχνητή νοημοσύνη είναι ικανή να αντικαταστήσει τους ειδικούς γιατρούς ως επαγγελματίες. Μεγάλο μέρος της δημοφιλούς συζήτησης για την τεχνητή νοημοσύνη επικεντρώνεται στην πρόοδο μέχρι στιγμής στην εφαρμογή της τεχνητής νοημοσύνης σε τομείς κλινικής έρευνας για τον καρκίνο.

Κατά συνέπεια, η έρευνα εφαρμογών τεχνητής νοημοσύνης έχει επιταχύνει επιδόσεις συγκρίσιμες με τους ειδικούς σε ανθρώπους στον βιοϊατρικό τομέα. Επιπλέον, η τεχνητή νοημοσύνη θα εξοπλίσει τους ανθρώπινους ειδικούς με περισσότερες πληροφορίες στη λήψη αποφάσεων και θα γίνει βασικό συστατικό της ιατρικής ομάδας.

Αυτή η ανασκόπηση παρέχει μια επισκόπηση σχετικά με τις βασικές έννοιες της τεχνητής νοημοσύνης στην κλινική έρευνα για τον καρκίνο, συμπεριλαμβανομένης της κατάστασης κλινικής έρευνας με χρήση τεχνητής νοημοσύνης τις τελευταίες δύο δεκαετίες, των διαθέσιμων δεδομένων, των τεχνικών και των τρεχουσών εφαρμογών. Περιγράφονται οι προκλήσεις που αντιμετωπίζει η μετάφραση της τεχνητής νοημοσύνης από τις θεωρητικές μελέτες στην κλινική χρήση του πραγματικού κόσμου. Ενώ αναμένεται να παρέχει μελλοντικές προοπτικές που θα βοηθήσουν στην προώθηση ουσιαστικών ερευνών που θα πραγματοποιήσουν τελικά την πραγματική συμμετοχή της τεχνητής νοημοσύνης στη θεραπεία του καρκίνου.

## 1. Υφιστάμενη κατάσταση

---

Η τεχνητή νοημοσύνη χρησιμοποιείται στην έρευνα για τον καρκίνο εδώ και σχεδόν 20 χρόνια. Σημαντικές πρόοδοι στην έρευνα για τον καρκίνο έχουν αρχίσει να δείχνουν πολλά υποσχόμενες και έχουν επιτευχθεί επιδόσεις σε επίπεδο ειδικών [1]. Υπάρχουν χιλιάδες εργασίες για την κλινική ογκολογία του καρκίνου από μοντέλα τεχνητής νοημοσύνης και ορισμένες βασικές μελέτες.

Επιπλέον, πολλές εταιρείες και βιομηχανικές ερευνητικές ομάδες έχουν ενταχθεί στη χρήση της τεχνητής νοημοσύνης για τον εντοπισμό, τη διάγνωση και τη θεραπεία του καρκίνου. Η IBM είναι η πρώτη εταιρεία που έκανε μεγάλη ώθηση για να φέρει την τεχνητή νοημοσύνη στην κλινική. Το 2014, η IBM ανέπτυξε τον Watson για να παρέχει ιατρική τεχνητή νοημοσύνη για την έρευνα για τον καρκίνο [10].

Ομοίως, στα ερευνητικά εργαστήρια της Microsoft, μια ομάδα επιστημόνων υπολογιστών και ερευνητών προσπαθούν να χρησιμοποιήσουν την ML και την Επεξεργασία Φυσικής Γλώσσας (NLP) για να προγραμματίσουν τη βιολογία για τη θεραπεία του καρκίνου [11]. Ως αποτέλεσμα, η τεχνητή νοημοσύνη είναι έτοιμη να έχει επιπτώσεις που αλλάζουν την πρακτική στη βελτίωση της ακρίβειας και της ταχύτητας της διάγνωσης, βοηθώντας τις προτάσεις και συστάσεις επιλογών θεραπείας και οδηγώντας σε καλύτερα αποτελέσματα πρόγνωσης

Η χρήση της τεχνητής νοημοσύνης (AI) για να ξεπεραστούν οι πολυπλοκότητες της ιατρικής έχει προαναγγελθεί εδώ και καιρό ως ένα κοντινό μέλλον και ανατρεπτικό ενδεχόμενο. Έχει μια μακρά ιστορία που χρονολογείται από τη δεκαετία του 1970, όταν τα συστήματα υποστήριξης κλινικών αποφάσεων (CDSS) απαιτούσαν από τους ανθρώπους να παρέχουν κανόνες για τεχνικές δέντρου αποφάσεων και να επιλέγουν χειροκίνητα χαρακτηριστικά για συμπερίληψη σε αυτά τα έμπειρα συστήματα [11].

Το ευρύ πεδίο της επιστήμης των υπολογιστών στο οποίο οι μηχανές ή οι αλγόριθμοι προγραμματίζονται για να προσομοιώνουν την ανθρώπινη νοημοσύνη καλύπτεται από τον όρο AI. Η μηχανική μάθηση (ML) είναι ένας κλάδος της τεχνητής νοημοσύνης στον οποίο οι υπολογιστές εκτελούν καθορισμένες εργασίες και εφαρμόζουν στατιστικές μεθόδους για την ανίχνευση κρυφών μοτίβων στα δεδομένα και τη βελτίωση της απόδοσης του μοντέλου [11].

Το υποπεδίο ML του DL, σε αντίθεση με το κλασικό ML, δεν απαιτεί ευρετικές που ορίζονται από τον άνθρωπο για να βρεθεί μια λύση για μια εργασία. Αντίθετα, η DL λειτουργεί με τη δύναμη των πολυεπίπεδων νευρωνικών δικτύων, επιτρέποντας έτσι την αυτοανακάλυψη χαρακτηριστικών άγνωστων ή απρόβλεπτων από τον άνθρωπο και εξαλείφοντας τη χειρωνακτική ανθρώπινη προσπάθεια για εξαγωγή χαρακτηριστικών. Τα συνελκτικά νευρωνικά δίκτυα (CNNs), ένας τύπος DL, μαζί με την εξαιρετικά αυξανόμενη υπολογιστική ισχύ έχουν οδηγήσει σε επιτάχυνση της ανάπτυξης εφαρμογών που βασίζονται στην τεχνητή νοημοσύνη, ιδιαίτερα στην ιατρική απεικόνιση [11].

Η επεξεργασία φυσικής γλώσσας (NLP) είναι μια παρακείμενη ειδικότητα στην τεχνητή νοημοσύνη που προσπαθεί να διασυνδεθεί η ανθρώπινη γλώσσα με τη μηχανική ερμηνεία. Χρησιμοποιείται για τη μετατροπή μη δομημένων δεδομένων - από κλινικές σημειώσεις EHR και διαγνωστικές ή διαδικαστικές αναφορές - σε διακριτά στοιχεία δεδομένων [11].

Μια πρόσφατη πρόοδος στον τομέα έχει οδηγήσει σε σημαντικά αυξημένη αποτελεσματικότητα της τεχνολογίας, η οποία μπορεί να χρησιμοποιηθεί για την αυτοματοποίηση της συλλογής και της τεκμηρίωσης της ημερομηνίας διάγνωσης, της επιβίωσης χωρίς εξέλιξη και άλλων χαρακτηριστικών όγκου που σχετίζονται με τον καρκίνο και των αποτελεσμάτων των ασθενών. Ένας τέτοιος αυτοματισμός θα μπορούσε να υποστηρίξει πολύπλοκη ανάπτυξη βάσεων δεδομένων και μητρώων όγκων, η οποία αυξάνει αναδρομικά τη δύναμη των παραγόμενων μοντέλων. Μόνο του ή σε συνδυασμό με τεχνικές ML/DL, το NLP έχει χρησιμοποιηθεί για την αντιστοίχιση κλινικών δοκιμών και για τον εντοπισμό πιθανών ανεπιθύμητων ενεργειών του φαρμάκου [11].

Έχει σημειωθεί σημαντική πρόοδος στον εντοπισμό κυκλοφορούντος DNA χωρίς καρκινικά κύτταρα για την πρόγνωση του καρκίνου, η οποία στη συνέχεια οδήγησε στην αξιολόγησή του για τον προσυμπτωματικό έλεγχο και την ανίχνευση του καρκίνου, καθώς και για την παρακολούθηση της υποτροπής του καρκίνου. Ένα πανκαρκινικό τεστ προσυμπτωματικού ελέγχου είναι ελκυστικό, καθώς ο συνδυασμένος επιπολασμός λιγότερο κοινών καρκίνων μπορεί να επιτρέψει μια οικονομικά αποδοτική μέθοδο για τη διευκόλυνση της έγκαιρης ανίχνευσής τους [11].

Οι αλγόριθμοι ML/DL μπορούν να ξεπεράσουν τους περιορισμούς των τυπικών υπολογιστικών μεθόδων μαθαίνοντας μοτίβα από ολόκληρο το μεταγραφώμιο. Για παράδειγμα, μια μέθοδος ML που χρησιμοποιεί δεδομένα αλληλούχισης ολόκληρου του

μεταγραφώματος RNA και ενσωματώνει πολλαπλά προφίλ όγκων έχει βρεθεί ότι αναγνωρίζει με ακρίβεια μια καρκινική κατάσταση και τη διακρίνει από τα φυσιολογικά κύτταρα. είχε καλή απόδοση για σπάνιους τύπους καρκίνου και έδειξε χρησιμότητα στην πρόβλεψη της θέσης προέλευσης του όγκου. Ομοίως, νευρωνικά δίκτυα έχουν εφαρμοστεί σε μεταγραφικά δεδομένα για την ταξινόμηση μοριακών υποτύπων διαφόρων όγκων [11].

## **1.1 Διαδικασία AI στην κλινική έρευνα για τον καρκίνο**

Αν και ορισμένοι αναμένουν ότι η τεχνητή νοημοσύνη μπορεί να αντικαταστήσει τους ανθρώπινους ειδικούς στη διαγνωστική απεικόνιση, τη λήψη αποφάσεων θεραπείας κ.λπ., η τεχνητή νοημοσύνη έπαιξε ρόλο μόνο στην επικουρική ιατρική. Σήμερα, η τεχνητή νοημοσύνη πρέπει να αντιμετωπίσει πολλά ζητήματα στην ανάπτυξη και την επικύρωση λύσεων στον τομέα της έρευνας για τον καρκίνο. Η δυνατότητα της τεχνητής νοημοσύνης μπορεί πρωτίστως να επιβεβαιωθεί σε προσεκτικά σχεδιασμένα πειράματα. Ως εκ τούτου, ένα ιδανικό σχήμα θα ωφελήσει τη βελτίωση των πρακτικών των λύσεων τεχνητής νοημοσύνης [12].

Πριν την αρχή της λύσης τεχνητής νοημοσύνης, είναι σημαντικό να ορίσετε και να χαρακτηρίσετε τα ζητήματα που πρέπει να αντιμετωπιστούν και να προβλέψετε εάν μπορεί να λυθεί (ή αξίζει να λυθεί) από την τεχνητή νοημοσύνη [12]. Πολλά κλινικά δεδομένα παρουσιάζουν μη ισορροπημένη φυσική κατανομή δειγμάτων μεταξύ διαφορετικών κατηγοριών. Το μη ισορροπημένο χαρακτηριστικό μπορεί να δημιουργήσει μια πρόκληση για αλγόριθμους ταξινόμησης που είναι γενικά σχεδιασμένοι για ισορροπημένες κλάσεις. Οι απλούστερες μέθοδοι επαναδειγματοληψίας είναι η τυχαία υπερδειγματοληψία και η τυχαία υποδειγματοληψία.

Η πρώτη αυξάνει την κατηγορία μειοψηφίας αντιγράφοντας τα δείγματα στην κατηγορία μειοψηφίας, ενώ η δεύτερη διαγράφει τυχαία ορισμένα δείγματα στην κατηγορία πλειοψηφίας [13]. Για να έχετε μια αμερόληπτη αξιολόγηση του μοντέλου AI, όλα τα διαθέσιμα δεδομένα χωρίζονται σε τρία μέρη, δηλαδή σε σύνολο εκπαίδευσης, σετ επικύρωσης και σετ δοκιμών. Η κοινή αναλογία είναι 60–70, 15–20 και 15–20%. Το σύνολο δεδομένων εκπαίδευσης χρησιμοποιείται για την εκπαίδευση του μοντέλου. Το σύνολο δεδομένων επαλήθευσης χρησιμοποιείται για την προσαρμογή των παραμέτρων

και την επιλογή χαρακτηριστικών. και το σύνολο δεδομένων δοκιμής χρησιμοποιείται για την αξιολόγηση της απόδοσης του εκπαιδευμένου μοντέλου ΑΙ.

Δεν είναι όλες αυτές οι λειτουργίες χρήσιμες για υπολογιστές μοντέλων τεχνητής νοημοσύνης, ενώ ορισμένα θορυβώδη ή άσχετα χαρακτηριστικά ενδέχεται να επηρεάσουν αρνητικά την απόδοση του ταξινομητή. Η επιλογή χαρακτηριστικών μπορεί να εκχωρήσει τιμές σχετικά με τα χαρακτηριστικά που είναι πιο σημαντικά για τον προσδιορισμό της ευνοϊκής βιοδραστικότητας. Πολλές μέθοδοι είναι διαθέσιμες για την επιλογή χαρακτηριστικών—t-test, false discovery rate (FDR), recursive feature elimination (RFE), Z-score, Wilcoxon, κ.λπ.

Η διασταυρούμενη επικύρωση είναι μια τεχνική επικύρωσης που χρησιμοποιείται για την αξιολόγηση ενός μοντέλου ΑΙ σε περιορισμένα δείγματα. Ένα δεδομένο σύνολο δεδομένων χωρίζεται τυχαία σε ομάδες  $K$  και μια ομάδα ορίζεται για δοκιμή και οι υπόλοιπες ομάδες  $K-1$  χρησιμοποιούνται για εκπαίδευση. Έτσι, η διαδικασία ονομάζεται συχνά διασταυρούμενη επικύρωση  $K$ -fold. Μια συνάρτηση απώλειας ή συνάρτηση κόστους μπορεί απλώς να μετρήσει την απόλυτη διαφορά μεταξύ του αποτελέσματος πρόβλεψης και της πραγματικής τιμής και να αντιπροσωπεύει κάποιο «κόστος» που σχετίζεται με το γεγονός.

Η επιλογή του καταλληλότερου μοντέλου θα μπορούσε να μεγιστοποιήσει τις πιθανότητες επιτυχίας στη λύση τεχνητής νοημοσύνης. Υπάρχουν τρεις συνήθεις τύποι μοντέλων που χρησιμοποιούνται στην τεχνητή νοημοσύνη: εποπτευόμενη μάθηση, μάθηση χωρίς επίβλεψη και ημι-εποπτευόμενη μάθηση. Η εποπτευόμενη μάθηση μπορεί να προβλέψει έναν ή περισσότερους στόχους που σχετίζονται με μια δεδομένη ετικέτα, τυπικές εφαρμογές συμπεριλαμβανομένης της παλινδρόμησης και της ταξινόμησης.

Η παλινδρόμηση είναι το έργο της πρόβλεψης ενός συνεχούς πραγματικού αριθμού ενώ η ταξινόμηση είναι το έργο της πρόβλεψης μιας διακριτής ετικέτας κλάσης. Οι δημοφιλείς αλγόριθμοι παλινδρόμησης περιλαμβάνουν το δέντρο απόφασης, τη γραμμική παλινδρόμηση, τους  $K$ -πλησιέστερους γείτονες (KNN), την παλινδρόμηση διανυσμάτων υποστήριξης (SVR). Εν τω μεταξύ, οι δημοφιλείς αλγόριθμοι ταξινόμησης περιλαμβάνουν τυχαίο δάσος, λογιστική παλινδρόμηση και μηχανές υποστήριξης διανυσμάτων (SVM).

Υπάρχει επίσης κάποια αλγοριθμική επικάλυψη, όπως τα νευρωνικά δίκτυα, τα οποία μπορούν να χρησιμοποιηθούν τόσο για ταξινόμηση όσο και για παλινδρόμηση. Αντίθετα,



η μάθηση χωρίς επίβλεψη επιτρέπει την ανακάλυψη λανθάνοντων κανόνων ή τάσεων στα δεδομένα, ή αλγορίθμων ομαδοποίησης, για τη διερεύνηση συλλογών δεδομένων και συσχετισμών μεταξύ δειγμάτων [14, 15].

Οι τυπικές μέθοδοι μάθησης χωρίς επίβλεψη περιλαμβάνουν την ιεραρχική ομαδοποίηση, τη διάδοση συνάφειας, το κρυφό μοντέλο Markov, την ομαδοποίηση k-means και τη χωρική ομαδοποίηση εφαρμογής με θόρυβο βάσει πυκνότητας (DBSCAN). Ωστόσο, η ημι-εποπτευόμενη μάθηση συνδυάζει την εποπτευόμενη μάθηση και την μάθηση χωρίς επίβλεψη, η οποία χρησιμοποιεί μεγάλο αριθμό δεδομένων χωρίς ετικέτα, καθώς και δεδομένα με ετικέτα ταυτόχρονα, για την αναγνώριση προτύπων.

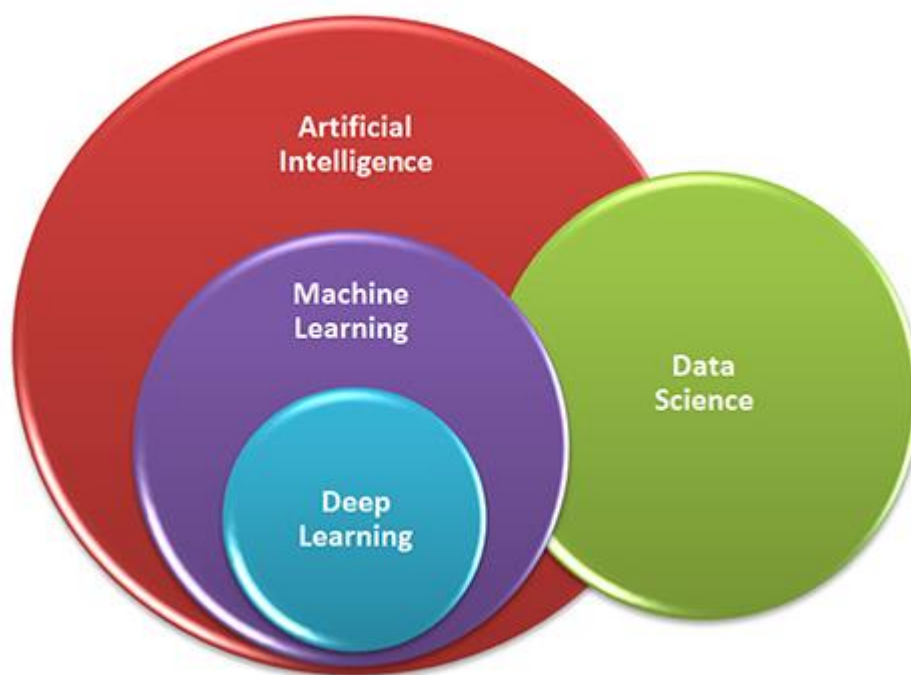
Η απλή αυτο-εκπαίδευση, η συνεκπαίδευση, ο αλγόριθμος διάδοσης ετικετών, η δημιουργία ημι-εποπτευόμενων μοντέλων και οι ημι-εποπτευόμενες μηχανές διανυσμάτων υποστήριξης (S3VM) είναι όλες ημι-εποπτευόμενες μέθοδοι μάθησης. Μια ανάπτυξη μοντέλου AI είναι να μάθεις τις παραμέτρους του και στη συνέχεια να κάνεις ακριβείς προβλέψεις ή προσδιορισμούς σε αόρατα δεδομένα.

Η απόδοση του μοντέλου αξιολογείται με διάφορες μετρήσεις [16]. Για ένα πρόβλημα δύο τάξεων, οι κοινές μέθοδοι απόδοσης περιλαμβάνουν την ευαισθησία, την ειδικότητα, την ακρίβεια, τον συντελεστή συσχέτισης Matthew (MCC), την τιμή F και την περιοχή κάτω από την καμπύλη (AUC). Μερικές φορές, για να εκτιμηθεί η γενίκευση της απόδοσης του μοντέλου AI, το μοντέλο εφαρμόζεται επίσης στο ανεξάρτητο σύνολο επικύρωσης για την αξιολόγηση της γενίκευσης του ταξινομητή σε νέα σύνολα δεδομένων [17].

## **1.2 Επισκόπηση της τεχνητής νοημοσύνης στην ογκολογία**

Το AI είναι ένας γενικός όρος που περιγράφει τη μίμηση της ανθρώπινης νοημοσύνης από υπολογιστές (Εικόνα 1). Η μηχανική μάθηση (ML), μια υποδιαίρεση της τεχνητής νοημοσύνης, αναφέρεται σε αλγόριθμους εκπαίδευσης υπολογιστών για να κάνουν προβλέψεις με βάση την εμπειρία και μπορεί να χωριστεί ευρέως σε εποπτευόμενο (όπου επιτρέπεται στον υπολογιστή να δει τα δεδομένα αποτελέσματος) ή χωρίς επίβλεψη (δεν παρέχονται δεδομένα αποτελέσματος) μάθηση.

Και οι δύο προσεγγίσεις αναζητούν μοτίβα δεδομένων για να επιτρέψουν προβλέψεις για τα αποτελέσματα, όπως η παρουσία ή η απουσία καρκίνου, τα ποσοστά επιβίωσης ή οι ομάδες κινδύνου. Κατά την ανάλυση μη δομημένων κλινικών δεδομένων, μια τεχνική που χρησιμοποιείται συχνά, τόσο στην ογκολογία όσο και ευρύτερα, είναι η επεξεργασία φυσικής γλώσσας (NLP) [12]. Το NLP μετατρέπει το μη δομημένο ελεύθερο κείμενο σε μορφή αναλύσιμη από υπολογιστή, επιτρέποντας την αυτοματοποίηση εργασιών με ένταση πόρων [13].



Εικόνα 1: Η τεχνητή νοημοσύνη και οι υποδιαιρέσεις της.

Πηγή: <https://www.vinsys.com/blog/a-complete-understanding-of-data-science-ai-machine-learning-deep-learning>

Είναι κοινή πρακτική στο ML να χωρίζονται δεδομένα σε διαμερίσματα, έτσι ώστε τα μοντέλα να αναπτύσσονται και να βελτιστοποιούνται σε υποσύνολα εκπαίδευσης και επικύρωσης, αλλά να αξιολογούνται σε ένα αόρατο σύνολο δοκιμών για να αποφευχθεί η υπερβολική αισιοδοξία. Τέτοιες μέθοδοι περιλαμβάνουν παραδοσιακά στατιστικά μοντέλα όπως η λογιστική παλινδρόμηση (LR) καθώς και νέα δέντρα αποφάσεων και αλγόριθμους DL [14].

Η βαθιά μάθηση (DL) είναι μια υποομάδα της ML, όπου κατασκευάζονται σύνθετες αρχιτεκτονικές ανάλογες με τους διασυνδεδεμένους νευρώνες του ανθρώπινου εγκεφάλου. Τα δημοφιλή πλαίσια για βαθιά μάθηση που βασίζονται σε Python περιλαμβάνουν το Tensorflow (Google) και το PyTorch (Facebook), τα οποία παρέχουν δυνατότητες για ανάπτυξη μοντέλων, εκπαίδευση και αξιολόγηση. Η Google παρέχει επίσης ένα δωρεάν διαδικτυακό περιβάλλον φορητού υπολογιστή, το Google Colaboratory, που επιτρέπει τη χρήση Python που βασίζεται σε σύννεφο και πρόσβαση σε μονάδες επεξεργασίας γραφικών (GPU) χωρίς εγκατάσταση τοπικού λογισμικού [15].

Αν και η λεπτομερής περιγραφή των δομών νευρωνικών δικτύων ξεφεύγει από το πεδίο εφαρμογής αυτού του άρθρου, τα τεχνητά νευρωνικά δίκτυα (ANN) μπορούν να χρησιμοποιηθούν για να επεξηγήσουν τις πρωταρχικές αρχές. Ως πρόσφατο παράδειγμα, οι Muhammad et al. χρησιμοποίησε ένα ANN για να προβλέψει τον κίνδυνο καρκίνου του παγκρέατος χρησιμοποιώντας κλινικές παραμέτρους όπως η ηλικία, η κατάσταση καπνίσματος, η χρήση αλκοόλ και η εθνικότητα [18]. Στην πιο βασική τους μορφή, τα ANN αποτελούνται από:

(1) ένα στρώμα εισόδου,

(2) ένα «κρυφό στρώμα», που αποτελείται από πολλούς κόμβους που πολλαπλασιάζουν την είσοδο με βάρη και προσθέτουν μια τιμή πόλωσης, και

(3) το επίπεδο εξόδου, περνώντας το σταθμισμένο άθροισμα των κρυφών κόμβων του στρώματος σε μια συνάρτηση ενεργοποίησης για να γίνουν προβλέψεις.

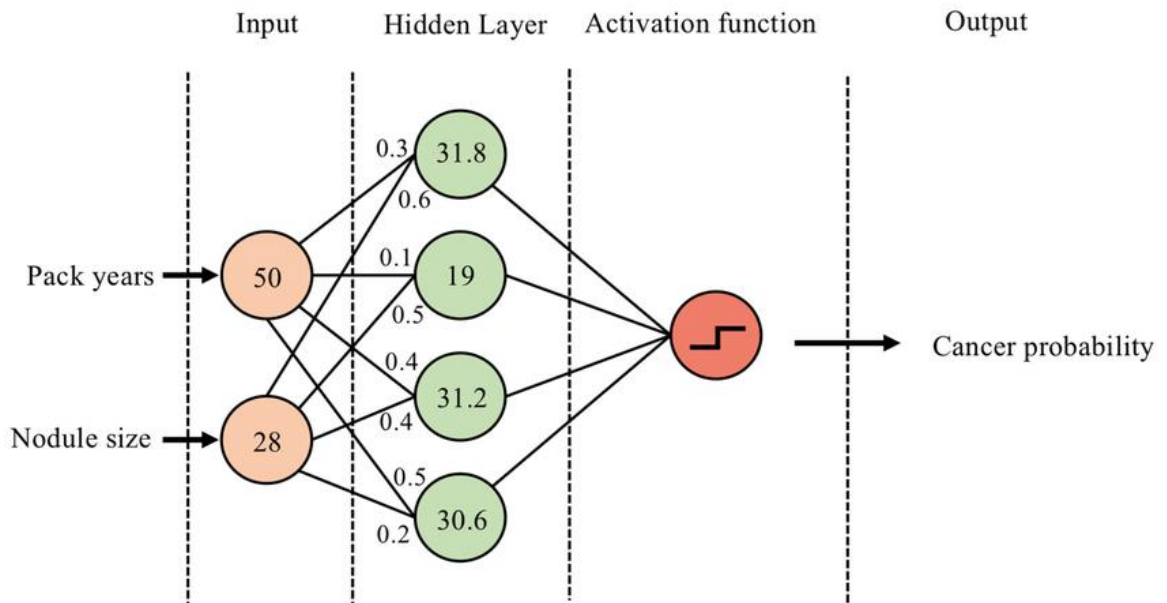
Η βαθιά εκμάθηση αναφέρεται απλώς σε δίκτυα με περισσότερα από ένα κρυφά επίπεδα. Πολλά μοντέλα πρώιμης διάγνωσης έχουν εκμεταλλευτεί αρχιτεκτονικές συνελκτικών νευρωνικών δικτύων (CNN), οι οποίες οδήγησαν σε επανάσταση στην έρευνα για την όραση υπολογιστή, επιτρέποντας τη χρήση έγχρωμων εικόνων ως δεδομένων εισόδου [19].

Ενώ τα κατάντη πλήρως συνδεδεμένα επίπεδα μοιάζουν με αυτά ενός ANN, τα δεδομένα εισόδου επεξεργάζονται από μια σειρά πυρήνων που ολισθαίνουν πάνω από τα κανάλια χρώματος της εικόνας και εξάγουν χαρακτηριστικά, όπως άκρες και χρωματικές διαβαθμίσεις. Αυτές οι εισοδοί στη συνέχεια συγκεντρώνονται και ισοπεδώνονται πριν περάσουν στο πλήρως συνδεδεμένο στρώμα. Πολλές προκαθορισμένες αρχιτεκτονικές

CNN με διάφορους βαθμούς πολυπλοκότητας είναι διαθέσιμες για χρήση, συμπεριλαμβανομένων των AlexNet [20], EfficientNet [21], InceptionNet [22], ResNet [23] και DenseNet [24]. Όπως συζητάμε περαιτέρω σε αυτό το άρθρο, τα CNN έχουν ένα ευρύ φάσμα εφαρμογών στην ακτινολογία και την ψηφιακή παθολογία.

### **1.2.1 Τύποι δεδομένων: Ηλεκτρονικά Μητρώα Υγείας**

Ένας αριθμός αναδυόμενων τρόπων δεδομένων υγειονομικής περίθαλψης είναι κατάλληλες για ανάλυση με τεχνητή νοημοσύνη. Τα τελευταία χρόνια, έχει συμβεί μια παγκόσμια επέκταση στις υποδομές ηλεκτρονικών αρχείων υγειονομικής περίθαλψης (EHR), επιτρέποντας την αποθήκευση και την αποτελεσματική πρόσβαση τεράστιων ποσοτήτων κλινικών δεδομένων [25]. Πολλές συναρπαστικές ψηφιακές συνεργασίες προκύπτουν για να διευκολυνθεί η έρευνα έγκαιρης διάγνωσης με χρήση ΗΜΥ, συμπεριλαμβανομένου του κόμβου DATA-CAN σε όλο το Ηνωμένο Βασίλειο [26]. Άλλες ψηφιακές βάσεις δεδομένων καταγράφουν μέτρα αποτελέσματος και δεδομένα διαδρομής. Για παράδειγμα, η Digital Cancer Waiting Times Database στοχεύει στη βελτίωση των οδών παραπομπής του καρκίνου μέσω μετρήσεων απόδοσης που ανεβαίνουν από χρήστες [27].



Εικόνα 2: Παράδειγμα αρχιτεκτονικής ANN ενός κρυφού επιπέδου. (1) Η κατάσταση καπνίσματος σε πακέτα ετών και το μέγεθος του όζου του πνεύμονα (mm) εισάγονται ως οι δύο κόμβοι εισόδου. (2) Στο κρυφό στρώμα, κάθε κόμβος πολλαπλασιάζει τις τιμές από τους εισερχόμενους νευρώνες με ένα βάρος (που εμφανίζεται ως δεκαδικά ψηφία στους εισερχόμενους νευρώνες). (3) Τα αποτελέσματα μεταβιβάζονται σε μια συνάρτηση ενεργοποίησης, μετατρέποντας την έξοδο σε πιθανότητα καρκίνου μεταξύ 0 και 1. Πολλαπλοί κύκλοι εκμάθησης χρησιμοποιούνται για την ενημέρωση των βαρών κρυφών επιπέδων για τη βελτίωση της απόδοσης. Πηγή : [https://www.researchgate.net/figure/Example-of-a-single-hidden-layer-ANN-architecture-1-The-smoking-status-in-pack-years\\_fig2\\_359258230](https://www.researchgate.net/figure/Example-of-a-single-hidden-layer-ANN-architecture-1-The-smoking-status-in-pack-years_fig2_359258230)

Είναι σημαντικό να γίνει διάκριση μεταξύ των τοπικών νοσοκομειακών δεδομένων EHR και των εθνικών μητρώων δεδομένων δημόσιας υγείας, συμπεριλαμβανομένων εκείνων που χρησιμοποιούνται από πολυκεντρικές μελέτες προσυμπτωματικού ελέγχου. Με τα μητρώα, εφαρμόζονται ενοποιημένες δομές βάσεων δεδομένων για συνέπεια μεταξύ των ιδρυμάτων. Ένας βασικός στόχος του προγράμματος NHSx «ψηφιακός μετασχηματισμός του ελέγχου» είναι να διασφαλίσει τη διαλειτουργικότητα των συστημάτων, έτσι ώστε τα δεδομένα να μπορούν να ρέουν απρόσκοπτα σε ολόκληρη τη διαδρομή ελέγχου, συμπεριλαμβανομένων των εθνικών βάσεων δεδομένων μητρώου [28].

Παράδειγμα ενοποίησης βάσεων δεδομένων είναι το νέο σύστημα διαχείρισης προσυμπτωματικού ελέγχου του καρκίνου του τραχήλου της μήτρας του Ηνωμένου Βασιλείου, το οποίο θα απλοποιήσει 84 διαφορετικές βάσεις δεδομένων σε μια ενιαία εθνική βάση δεδομένων και στοχεύει στον εξορθολογισμό της εισαγωγής δεδομένων και

την παροχή απλής πρόσβασης βάσει cloud για τους χρήστες [29]. Οι ψηφιακές βάσεις δεδομένων, είτε τοπικές είτε εθνικές, είναι ώριμες για ανάλυση με τεχνητή νοημοσύνη, η οποία είναι εγγενώς ικανή να επεξεργάζεται μεγάλες ποσότητες πληροφοριών («Big Data») [30].

Τα δεδομένα EHR περιλαμβάνουν συνήθως δομημένα, εύκολα μετρήσιμα δεδομένα, όπως ημερομηνίες εισαγωγής ή αποτελέσματα αίματος, και μη δομημένο ελεύθερο κείμενο, όπως κλινικές σημειώσεις ή διαγνωστικές αναφορές. Το τελευταίο μπορεί να αναλυθεί χρησιμοποιώντας προσεγγίσεις NLP. Μια επισκόπηση του NLP στην ογκολογία παρέχεται από τους Yim et al. [12] και παραδείγματα χρήσεων έγκαιρης διάγνωσης περιλαμβάνουν τον εντοπισμό μη φυσιολογικών αποτελεσμάτων προσυμπτωματικού ελέγχου καρκίνου [31], τον έλεγχο προτύπων κολonosκόπησης ή κυστεοσκόπησης [32,33] και τον εντοπισμό ή τη διαστρωμάτωση κινδύνου προκακοήθων βλαβών [34-38].

Το NLP έχει επίσης χρησιμοποιηθεί για την αυτοματοποίηση της ταυτοποίησης ασθενών για κλινικές δοκιμές, μειώνοντας τον φόρτο των ελέγχων καταλληλότητας [39]. Ο Morin και οι συνεργάτες του δημοσίευσαν ένα συναρπαστικό παράδειγμα για το πώς η τεχνολογία AI και NLP μπορεί να ενσωματωθεί στα συστήματα EHR: το μοντέλο τους μπορεί να αναλύσει εκατομμύρια σημεία δεδομένων και να εκτελέσει πρόγνωση καρκίνου σε πραγματικό χρόνο με βάση τη συνεχή εκμάθηση κλινικών δεδομένων που συλλέγονται τακτικά [40].

### 1.2.1.1 Διαθέσιμα δεδομένα της διαδικασίας AI στον βιοϊατρικό τομέα

---

Για τη βελτιστοποίηση των επιθυμητών αποτελεσμάτων, είναι σημαντικό να κατανοήσουμε ποια είδη συνόλων δεδομένων απαιτούνται για μια πιθανή χρησιμότητα και πώς να αποκτήσετε αυτά τα σύνολα δεδομένων. Οι ευαίσθητοι και χρήσιμοι δείκτες ή χαρακτηριστικά για την έρευνα για τον καρκίνο που εκτελείται με τεχνητή νοημοσύνη συνήθως περιλαμβάνουν:

- (i) ακτινογραφία,
- (ii) γονιδίωμα του καρκίνου,
- (iii) ιατρικά αρχεία,
- (iv) πληροφορίες για φάρμακα και
- (v) βιοϊατρικές βιβλιογραφία.

Οι τεχνικές ιατρικής απεικόνισης μπορούν να παρέχουν μια άποψη των εσωτερικών δραστηριοτήτων μέσα στο ανθρώπινο σώμα χωρίς τομές από γιατρούς, ιατρούς και ερευνητές [41-46]. Οι μέθοδοι σάρωσης εικόνας διαδραματίζουν σημαντικό ρόλο στην κλινική έρευνα των καρκίνων. Η μαγνητική τομογραφία (MRI), η αξονική τομογραφία (CT), η υπεριώδης ακτινοβολία, οι ιστοπαθολογικές διαφάνειες, η ακτινογραφία και η μαστογραφία είναι οι πιο κοινές μέθοδοι σάρωσης εικόνας στην ακτινολογία [46-50]. Ως αποτέλεσμα, είναι πιθανό ότι η τεχνητή νοημοσύνη θα μπορούσε να χρησιμοποιηθεί για τη διάκριση περιπτώσεων ασθένειας από υγιείς μάρτυρες εξάγοντας πληροφορίες, όπως σταδιοποίηση του όγκου, μέγεθος όγκου, εντοπισμό καρκινώματος και ούτω καθεξής, από εικόνες [51-55].

Επιπλέον, η γονιδιωματική ανάλυση έχει επικεντρωθεί στη βελτίωση της ακρίβειας της δοκιμής συμπεριλαμβάνοντας περισσότερα χαρακτηριστικά σχετικά με την ασθένεια [22] και στην κατανόηση των σχέσεων μεταξύ γενετικών μορφών και καταστάσεων ασθένειας. Οι υπογραφές γονιδιακής έκφρασης μπορούν να χρησιμοποιηθούν για την εύρεση πιθανών βιοδεικτών ή θεραπευτικών στόχων, οι οποίοι μπορούν να ληφθούν από την ανάλυση μικροσυστοιχιών DNA, δεδομένων έκφρασης RNA και μεταγραφής και έχουν αποδειχθεί χρήσιμοι στην ταξινόμηση καρκίνων σε επίπεδο γονιδιακής έκφρασης [33].

Τα τσιπ ολιγονουκλεοτιδίου και οι συστοιχίες cDNA είναι οι δύο κοινώς χρησιμοποιούμενες μικροσυστοιχίες [56-57]. Ορισμένα δημοφιλή σύνολα δεδομένων μικροσυστοιχιών καρκίνου υψηλών διαστάσεων, όπως το σύνολο δεδομένων μικρού στρογγυλού μπλε κυττάρου (SRBCT) [58] και διάχυτο μεγάλο λέμφωμα Β-κυττάρων (DLBCL) [59], περιέχουν μεγάλο αριθμό πειραματικών δειγμάτων με αντίστοιχες εκφράσεις γονιδίων. Το Genomics μπορεί να χρησιμοποιήσει τις γενετικές αλλαγές στον όγκο του ασθενούς για να καθορίσει ένα κατάλληλο σχέδιο θεραπείας για ιατρική ακριβείας.

Επιπλέον, ενισχυμένοι από την πρόοδο των βιοτεχνολογιών υψηλής απόδοσης, οι πολυάριθμοι νέοι καρκινικοί βιοδείκτες είχαν βρεθεί σε πρωτεώματα του σωματικού υγρού, όπως τα κυκλικά RNA (circRNAs) που χρησιμεύουν ως βιοδείκτες για τον καρκίνο του προστάτη και μπορούν να ανιχνευθούν στα ούρα [60]. Η εύρεση της πιο διακριτικής γονιδιωματικής έκφρασης σε διαφορετικά στάδια καρκίνου είναι μια σημαντική πρόκληση για την έρευνα για τον καρκίνο τα τελευταία χρόνια.

Εν τω μεταξύ, τα ιατρικά αρχεία, π.χ. οικογενειακό ιστορικό, ηλικία, διαίτα, βάρος (παχυσαρκία), συνήθειες υψηλού κινδύνου (κάπνισμα, μεγάλη κατανάλωση αλκοόλ) και έκθεση σε περιβαλλοντικές καρκινογόνες ουσίες (ακτινοβολία UV, ραδόνιο ή αμίαντος), μπορεί να σχετίζονται με την εμφάνιση και την εξέλιξη του καρκίνου. Ωστόσο, αυτά τα υπάρχοντα δεδομένα για τα μοντέλα εκπαίδευσης είναι περιορισμένα και ενιαία.

Πολλές συνεχιζόμενες κλινικές πληροφορίες στις αναφορές χειρουργικής παθολογίας μπορούν να χρησιμοποιηθούν για τον προσδιορισμό της καταλληλότητας των ασθενών που στρατολογήθηκαν για τη μελέτη. Για τους ειδικούς της επιστήμης δεδομένων είναι δύσκολο να αποκτήσουν κλινικά δεδομένα λόγω έλλειψης ευκαιρίας για κλινική πρακτική ή λόγω της ανάγκης έγκρισης από το ινστιτούτο.

Επιπλέον, πολλά κλινικά δεδομένα από πρώτο χέρι ήταν δύσκολο να ενσωματωθούν σε μοντέλα λόγω της δυσκολίας στη χειροκίνητη λήψη. Τα δεδομένα του Ηλεκτρονικού Μητρώου Υγείας (EHR) θα βελτιώσουν δυνητικά τα αποτελέσματα της βιοϊατρικής έρευνας. Τα συστήματα EHR ενσωματώνουν εργαστηριακά αποτελέσματα, αναφορές διαδικασίας και ακτινολογίας και κλινικές αφηγήσεις, όπως σημειώσεις της πρωτοβάθμιας περίθαλψης και της γαστρεντερολογικής κλινικής.



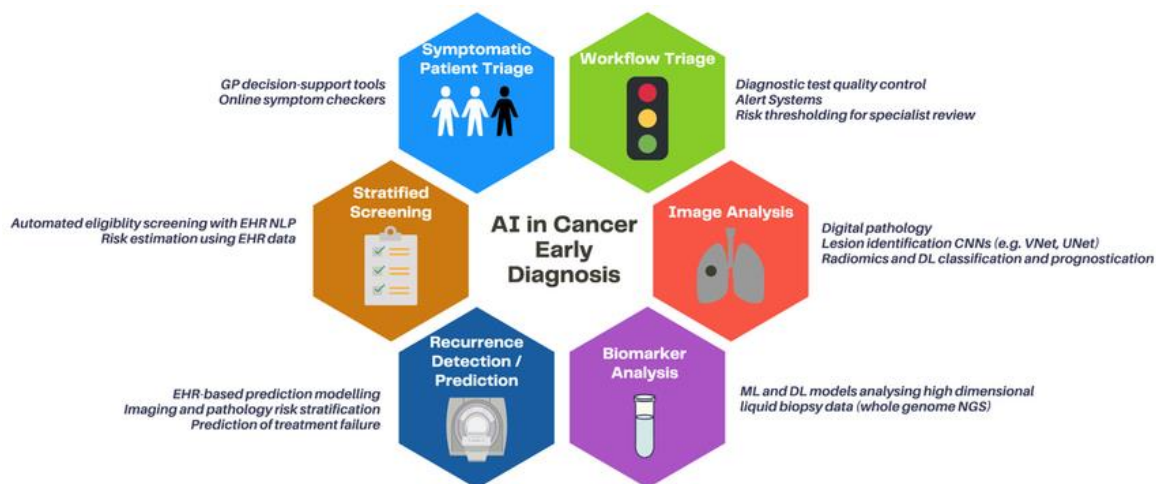
Πρόσφατα, πολλά υποσχόμενα αποτελέσματα έχουν αποδειχθεί με τη χρήση της τεχνητής νοημοσύνης στην ανάπτυξη φαρμάκων, στο προφίλ φαρμάκων-στόχων και στην επανατοποθέτηση/ανατοποθέτηση φαρμάκων. Η φαρμακοδυναμική, η φαρμακοκινητική και η τοξικολογική μπορούν να βελτιώσουν την εξειδίκευση και την εκλεκτικότητα του στόχου στο σχεδιασμό του φαρμάκου με μικρά μόρια. Διαφορετικοί τύποι δεδομένων έχουν χρησιμοποιηθεί στη βιβλιογραφία ανακάλυψης φαρμάκων που σχετίζεται με τον καρκίνο με βάση την τεχνητή νοημοσύνη. Οι κλασικοί τύποι δεδομένων περιλαμβάνουν χημικές δομές φαρμάκων, φυσικοχημικές ιδιότητες και μοριακούς στόχους [61].

Ειδικά, η μικροσυστοιχία RNA, η συστοιχία πολυμορφισμού απλού νουκλεοτιδίου (SNP), η αλληλουχία RNA (RNA-Seq), η σειρά πρωτεϊνών αντίστροφης φάσης, η αλληλουχία εξωμίου και η κατάσταση μεθυλίωσης DNA είναι διαθέσιμα για την εύρεση βιοδεικτών και τη δημιουργία προγνωστικών μοντέλων ευαισθησίας στα φάρμακα [2]. Οι υπάρχοντες πόροι για τη διευκόλυνση της ανακάλυψης φαρμάκων για τον καρκίνο περιλαμβάνουν τα DepMap [62], Genomics of Drug Sensitivity in Cancer (GDSC) [63], canSAR [64], Open Targets [65], TG-GATE [66], drugBank [67] και οι υπολοίποι. Χρησιμοποιώντας αυτές τις βάσεις δεδομένων και τους πόρους, μπορεί να συσχετιστεί η ευαισθησία στα φάρμακα και να παρασχεθούν πιθανοί βιοδείκτες της απόκρισης στα φάρμακα.

Επιπλέον, η βιοϊατρική βιβλιογραφία είναι μεγάλη και αυξάνεται ραγδαία. Έχουν δημοσιευτεί αρκετές επιτυχημένες εφαρμογές της τεχνητής νοημοσύνης σε διάφορα στάδια της έρευνας για τον καρκίνο [70-75]. Για να συγκεντρώσει και να παρουσιάσει ομοιόμορφα τους διαθέσιμους πόρους από την ανάκτηση βιβλιογραφίας υψηλής απόδοσης της κλινικής έρευνας για τον καρκίνο, πολλές κεντρικές, ελεύθερα προσβάσιμες και ανοιχτές βάσεις γνώσης της κοινότητας, π.χ. TCGA, έχουν δημιουργηθεί για να παρέχουν κλινικά και μοριακά δεδομένα για κλινικούς ιατρούς και ερευνητές. Αυτές οι βάσεις γνώσεων ενσωματώνουν ετερογενή δεδομένα, συμπεριλαμβανομένων πληροφοριών γονιδίων, πρωτεϊνών και έκφρασης σε ιστούς ελέγχου και όγκου, καθώς και πληροφορίες ακτινογραφικής απεικόνισης ασθενών με καρκίνο [76-79].

### 1.3 Κλινικές Εφαρμογές

Παρακάτω, παρατίθενται οι τομείς, όπου η τεχνητή νοημοσύνη είναι πιθανό να έχει κλινικό αντίκτυπο στο εγγύς μέλλον, χρησιμοποιώντας υποδειγματικές ομάδες καρκίνου.



Εικόνα 3: Κλινικές εφαρμογές της AI στην έγκαιρη διάγνωση του καρκίνου. Κλινικές Συνομογραφίες: GP: γενικός ιατρός, NLP: επεξεργασία φυσικής γλώσσας, EHR: ηλεκτρονικό αρχείο υγειονομικής περίθαλψης, ML: μηχανική μάθηση, DL: βαθιά μάθηση, NGS: αλληλουχία επόμενης γενιάς.

Πηγή : [https://www.researchgate.net/figure/Clinical-applications-of-AI-in-early-cancer-diagnosis-Abbreviations-GP-general\\_fig3\\_359258230](https://www.researchgate.net/figure/Clinical-applications-of-AI-in-early-cancer-diagnosis-Abbreviations-GP-general_fig3_359258230)

#### 1.3.1 Στρωματοποιημένος προσυμπτωματικός έλεγχος κινδύνου ασυμπτωματικών ασθενών

Αρκετές μελέτες μεγάλης κλίμακας έχουν δείξει ότι ο προσυμπτωματικός έλεγχος για τον καρκίνο του πνεύμονα σε ασθενείς σε κίνδυνο προσδίδει οφέλη επιβίωσης [80,81]. Στη συνέχεια, στις ΗΠΑ, τα Κέντρα για Υπηρεσίες Medicare & Medicaid (CMS) έκριναν ότι ασθενείς ηλικίας 55–77 ετών με ιστορικό καπνίσματος  $\geq 30$  πακέτου έτους είναι επιλέξιμοι για έλεγχο αξονικής τομογραφίας, με νέες κατευθυντήριες γραμμές να προτείνουν ότι αυτό θα πρέπει να χαλαρώσει περαιτέρω [82 ,83].

Ωστόσο, στην πράξη, μόνο ένα μικρό ποσοστό των επιλέξιμων ασθενών ελέγχεται στην πραγματικότητα, εν μέρει λόγω της κακής τεκμηρίωσης της κατάστασης του καπνίσματος και των πιέσεων χρόνου του γιατρού [84,85]. Για τη βελτίωση της επιλογής διαλογής, οι Lu et al. ανέπτυξε ένα μοντέλο CNN που ενσωματώνει ακτινογραφίες θώρακα και

ελάχιστα δεδομένα EHR (ηλικία, φύλο, τρέχουσα κατάσταση καπνίσματος) για την πρόβλεψη του κινδύνου εμφάνισης καρκίνου 12 ετών, το οποίο συγκρίθηκε με τα κριτήρια CMS [86].

Το στοιχείο απεικόνισης εκπαιδεύτηκε χρησιμοποιώντας ένα δίκτυο Inception V4 σε 85.748 ακτινογραφίες από τη δοκιμή PLCO και επικυρώθηκε σε 5615 και 5493 ακτινογραφίες από τις μελέτες PLCO και NLST, αντίστοιχα. Η ομάδα διαπίστωσε ότι το μοντέλο βελτιώθηκε με βάση τα κριτήρια καταλληλότητας CMS, αναφέροντας AUC 0,755 σε σύγκριση με 0,634 και πέτυχε ισοτιμία με πιο σύνθετες βαθμολογίες κινδύνου που απαιτούσαν 11 σημεία δεδομένων (PLCOM2012) [86].

Πιο πρόσφατα, οι Gould et al. δημοσίευσε ένα μοντέλο ML που βασίζεται σε μη απεικονιστικά δεδομένα EHR [87]. Χρησιμοποιώντας ένα σύνολο δεδομένων 6505 ασθενών με καρκίνο του πνεύμονα και 189.597 μαρτύρων, το μοντέλο ήταν πιο ακριβές από τα κριτήρια PLCO στην πρόβλεψη του καρκίνου του πνεύμονα μέσα στους επόμενους 9-12 μήνες (AUC 0,86). Επιπλέον, βελτιώνονται τα τυπικά κριτήρια καταλληλότητας για τον προσυμπτωματικό έλεγχο καρκίνου του πνεύμονα, παρέχοντας αποδείξεις ότι η αξιολόγηση των κλινικών δεδομένων ρουτίνας με ενισχυμένη τεχνητή νοημοσύνη μπορεί να βοηθήσει στον εντοπισμό ασθενών για στοχευμένα προγράμματα προσυμπτωματικού ελέγχου. Η χρήση της τεχνητής νοημοσύνης για τη βελτίωση της επιλογής ασθενών για προσυμπτωματικό έλεγχο μπορεί να είναι μια χρήσιμη διαδρομή για την έγκαιρη διάγνωση στο μέλλον.

### **1.3.2 Συμπτωματική Διαλογή Ασθενούς**

Οι γενικοί ιατροί (GPs) είναι συχνά το πρώτο επίκεντρο για ασθενείς με συμπτώματα καρκίνου και έχουν κρίσιμο ρόλο να διαδραματίσουν ως φύλακες στη δευτεροβάθμια περίθαλψη [88]. Κατά την τελευταία δεκαετία, έχουν εμφανιστεί διάφορα εργαλεία υποστήριξης αποφάσεων για να βοηθήσουν τους γενικούς ιατρούς να προσδιορίσουν ποια συμπτώματα καρκίνου απαιτούν παραπομπή για περαιτέρω έρευνα [89]. Για παράδειγμα, το εργαλείο υποστήριξης αποφάσεων με τη σήμανση CE, «C The Signs», εφαρμόζεται επί του παρόντος σε μια σειρά από πρακτικές για να βοηθήσει τους γενικούς ιατρούς στη διαστρωμάτωση του κινδύνου καρκίνου [90,91].

Το εργαλείο παρέχει έναν πίνακα ελέγχου για χρήση σε πραγματικό χρόνο και προτείνει έρευνες ή παραπομπές με βάση τα προφίλ συμπτωμάτων του καρκίνου. Οι πρώιμες αναφορές αξιολόγησης υποδηλώνουν αυξημένο ποσοστό ανίχνευσης καρκίνου 6,4% [91]. Θα πρέπει να σημειωθεί ότι δεν υπάρχουν επί του παρόντος στη βιβλιογραφία δημοσιεύσεις με κριτές σχετικά με αυτό το εργαλείο, και παρόλο που το μάρκετινγκ υποδεικνύει τη χρήση της τεχνητής νοημοσύνης για τη χαρτογράφηση των αποφάσεων με τα πιο πρόσφατα στοιχεία, δεν είναι δυνατή η πλήρης κριτική της υποδομής του χωρίς δημοσιευμένη μεθοδολογία.

Αναδύονται επίσης τεχνολογίες για τη διάγνωση και τη διαλογή ασθενών απευθείας σύμφωνα με τα συμπτώματα που περιγράφονται μόνοι τους, χρησιμοποιώντας chat-bots ή διαδικτυακούς ελέγχους συμπτωμάτων. Ο εμπορικός πάροχος υγειονομικής περίθαλψης digital, Babylon Health, παρέχει στους ασθενείς πρόσβαση σε ιδιωτικές συμβουλές μέσω τηλεφώνου ή εφαρμογών υπολογιστή [92]. Η Babylon χρησιμοποιεί ένα Bayesian δίκτυο που βασίζεται σε προφίλ πιθανότητας ασθενειών που ενημερώνονται από επιδημιολογικά δεδομένα και γνώμες ειδικών για να διαγνώσει ασθένειες και να προτείνει ενέργειες, όπως η παρακολούθηση ατυχήματος και έκτακτης ανάγκης ή η κράτηση ενός μη επείγοντα ραντεβού για τον γιατρό, σύμφωνα με τα συμπτώματα που έχουν εισαχθεί από τον ασθενή [93].

Η διαλογή και το διαγνωστικό του σύστημα αναφέρεται ότι έχουν συγκρίσιμη ακρίβεια και ασφάλεια με τους ανθρώπινους γιατρούς, με την προειδοποίηση ότι η χρήση προσομοιωμένων συμβουλών περιορίζει την εξωτερική εγκυρότητα αυτής της αξιολόγησης [93]. Είναι και πάλι σημαντικό να σημειωθεί ότι αυτό το εργαλείο έχει εφαρμοστεί κλινικά παρά την έλλειψη δημοσιεύσεων με κριτές από ομοτίμους που περιγράφουν λεπτομερείς διαδικασίες δοκιμών και επικύρωσης, κάτι που έχει προκαλέσει κριτική από την MHRA και την ογκολογική κοινότητα [94].

Η τρέχουσα βιβλιογραφία υποδηλώνει ότι η τεχνητή νοημοσύνη μπορεί να διαδραματίσει ρόλο στη διαλογή συμπτωματικών ασθενών στην κοινότητα που διατρέχουν κίνδυνο καρκίνου στο μέλλον [95]. Ωστόσο, απαιτούνται περαιτέρω στοιχεία, συμπεριλαμβανομένων ισχυρών προοπτικών μελετών επικύρωσης, για να επιβεβαιωθεί η αποτελεσματικότητα και η ασφάλειά τους για κλινική ανάπτυξη.

## 2. Τρέχουσες προσπάθειες εξειδίκευσης και γνώσης τομέα

---

### 2.1 Προσπάθειες της εμπειρογνωμοσύνης

Υπάρχει μια γενική έλλειψη διαθέσιμων ετικετών ειδικών για μια παραγωγή [2]. Ενώ τα ακατέργαστα δεδομένα μπορούν να τροφοδοτηθούν απευθείας στα μοντέλα ΑΙ, τα σύνολα δεδομένων εξακολουθούν να απαιτούν χειροκίνητο σχολιασμό ή τουλάχιστον επιμέλεια [1]. Θα πρέπει να συμμετέχουν πολλοί ειδικοί στα θέματα στον σχολιασμό δεδομένων για να παρέχουν ακριβή αξιολόγηση των ετικετών δεδομένων. Για παράδειγμα, ο σχολιασμός ιατρικών εικόνων απαιτεί δεσμεύσεις από κλινικούς ειδικούς για την εκ των προτέρων εξαγωγή τομέων περιοχών ενδιαφέροντος (ROI). Μερικές φορές, τα αποτελέσματα του μοντέλου εκπαίδευσης πρέπει επίσης να αξιολογηθούν ποιοτικά από τους ειδικούς για την προσαρμογή των υπερπαραμέτρων του μοντέλου.

Ειδικότερα, σπάνιες περιπτώσεις, οι οποίες είναι πολύ σημαντικές στην έρευνα για τον καρκίνο, χρειάζονται ειδικούς ανθρώπους για να αναγνωρίσουν το όφελος από την εκπαίδευση και την εμπειρία τους. Για παράδειγμα, η ανίχνευση της περιοχής του καρκίνου του πνεύμονα από εικόνες παθολογίας που έχουν χρωματιστεί με αιματοξυλίνη και ηωσίνη (H&E) από ΑΙ είναι πολύ δύσκολη λόγω της πολυπλοκότητας των δομών του ιστού του καρκίνου του πνεύμονα. Χρειάζεται ανθρώπινους παθολόγους για να κυκλώσουν τα ακριβή όρια του όγκου και να υποδείξουν όλο τον όγκο που εξαπλώνεται μέσω των εναέριων χώρων (STAS) [48].

Παραδείγματα επιτυχημένων περιπτώσεων εμπειρογνωμοσύνης και μοντελοποίησης της συνεργασίας μεταξύ της τεχνητής νοημοσύνης και των ειδικών περιλαμβάνουν την εργασία των Fan et al. [49] και Cha et al. [37], όπου οι καρκινικές βλάβες κατατμήθηκαν χειροκίνητα με συναίνεση από έμπειρους ειδικούς ακτινολόγους και οι επιδόσεις των μοντέλων συγκρίθηκαν με ειδικούς ακτινολόγους. Αυτό θα είναι πιθανότατα ένας κοινός ρόλος για εμπειρογνωμοσύνη στο εγγύς μέλλον.

### 2.1.1 Ενσωματωμένη γνώση του τομέα

Οι αιτιολογικοί παράγοντες (π.χ. σχέσεις μεταξύ κλινικών συμβάντων) έχουν συχνά κοινούς καθοριστικούς παράγοντες σε πολλαπλά επίπεδα. Η ολοκληρωμένη γνώση μπορεί να βελτιώσει την ποιότητα της έρευνας της τεχνητής νοημοσύνης σε μεμονωμένα ιατρικά συμβάντα με καρκίνους, ανοίγοντας παράλληλα το δρόμο για την ερμηνευσιμότητα των μοντέλων τεχνητής νοημοσύνης. Αυτή η ολοκλήρωση έχει το διπλό πλεονέκτημα της δημιουργίας μεγαλύτερης γνώσης σχετικά με τους περίπλοκους μηχανισμούς των καρκίνων, καθώς και τη βελτίωση της κατανόησής μας για τις επιρροές των διαφορετικών αιτιολογικών παραγόντων. Καμία υπάρχουσα μέθοδος δεν εκμεταλλεύεται δυναμικά την περίπλοκη ιατρική γνώση [96].

Έχουν δημοσιευτεί τώρα αρκετές επιτυχημένες περιπτώσιολογικές μελέτες στις οποίες διαφορετικοί τομείς γνώσης συνεργάζονται για την επίλυση προβλημάτων από την ιατρική πρακτική. Οι πρόσφατες εξελίξεις με τη χρήση ψηφιακών εικόνων παθολογίας έχουν εφαρμοστεί παραγωγικά σε απεικονιστικές εργασίες (π.χ. ταξινόμηση, πρόγνωση και πρόβλεψη) σε όλη τη δερματολογία, την οφθαλμολογία, την ακτινολογία και την ιστοπαθολογία.

Οι Nguyen et al. [50] ανέπτυξε μια μελέτη σκοπιμότητας για την πρόβλεψη της βέλτιστης κατανομής της δόσης ακτινοθεραπείας σε ασθενείς με καρκίνο του προστάτη από την ανατομία του ασθενούς χρησιμοποιώντας βαθιά μάθηση. Οι Mobadersany et al. [51] συνδύασε την ιστολογία με τη γονιδιωματική για να βελτιώσει την προγνωστική ακρίβεια των μοριακών υποτύπων γλοιώματος αντί να βασίζεται μόνο στην ιστολογία. Esteva et al. [52] έδειξε ταξινόμηση των δερματικών βλαβών χρησιμοποιώντας CNN σε ένα σύνολο δεδομένων 129 450 κλινικών εικόνων.

Η εφαρμογή της τεχνητής νοημοσύνης για την ανάλυση πολυομικών δεδομένων μεγάλου όγκου (εξώμιο, μεταγραφώματος και επιγονιδιώματος) σε συνδυασμό με κλινικά σχολιασμένα σύνολα δεδομένων έχει επίσης οδηγήσει στον εντοπισμό γονιδίων ευαισθησίας σε φάρμακα, στην ανίχνευση παραλλαγών, σε νέες γνώσεις βιολογίας του καρκίνου και στην πρόβλεψη τοποθεσιών συγκόλλησης RNA.

Μια σημαντική πρόκληση στην ανάπτυξη μοντέλων τεχνητής νοημοσύνης είναι η έλλειψη δομημένων δεδομένων υγείας που σχετίζονται με τον καρκίνο, καθώς και η έλλειψη τυποποίησης στον τρόπο συλλογής και αποθήκευσης μη δομημένων δεδομένων

στο πλαίσιο του EHR ή της ενοποιημένης πλατφόρμας δεδομένων ενός ενιαίου συστήματος υγειονομικής περίθαλψης. Η έλλειψη τυποποίησης μεταξύ των συστημάτων υγειονομικής περίθαλψης και των παγκόσμιων κοινοτήτων είναι ακόμη πιο σημαντική καθώς περιορίζει τη διαλειτουργικότητα και την ευρεία ανταλλαγή δεδομένων και πληροφοριών για την υγεία. Για να αντιμετωπιστεί αυτό, η πρωτοβουλία ελάχιστων κοινών ογκολογικών στοιχείων δεδομένων καθιέρωσε καθολικούς όρους και ορισμούς για συχνά χρησιμοποιούμενα χαρακτηριστικά ασθενών και όγκων, ταξινομήσεις της κατάστασης της νόσου και θεραπευτικές παρεμβάσεις [97].

Η εφαρμογή του απαιτεί σημαντικούς πόρους τεχνολογίας πληροφοριών και συστημάτων και η σκοπιμότητα της υιοθέτησής του στη συνήθη κλινική πρακτική παραμένει υπό διερεύνηση. Τα αποτελέσματα που αναφέρονται από τους ασθενείς και τα επικυρωμένα ερωτηματολόγια είναι άλλα μέσα με τα οποία μπορούν να συλλεχθούν τυποποιημένα δεδομένα υγείας που παράγονται από τον ασθενή. Αυτά τα δεδομένα συχνά καταγράφονται συστηματικά ως μέρος κλινικών δοκιμών ή συνήθους πρακτικής, αλλά μπορούν επίσης να εξορύσσονται απευθείας από το EHR χρησιμοποιώντας τεχνικές επεξεργασίας φυσικής γλώσσας ή ασύγχρονες ηλεκτρονικές αλληλεπιδράσεις [98].

Είναι σημαντικό ότι τα αποτελέσματα που αναφέρθηκαν από τον ασθενή μπορεί να είναι ισχυροί προγνωστικοί δείκτες επιβίωσης. Αυτά τα εργαλεία μπορεί να έχουν πολλά άλλα οφέλη για την παροχή φροντίδας για τον καρκίνο, την έρευνα και τις κλινικές λειτουργίες. Ωστόσο, μπορούν επίσης να αυξήσουν το γραφειοκρατικό βάρος για τις ομάδες κλινικής φροντίδας, καθώς και για τους ασθενείς και τους φροντιστές τους. Μια διεπιστημονική προσέγγιση στην κλινική εφαρμογή τους θα είναι απαραίτητη για την ελαχιστοποίηση των αλληλεπικαλυπτόμενων προσπαθειών και τη βελτιστοποίηση της πληρότητας της συλλογής δεδομένων. Στην ιδανική περίπτωση, αυτά τα τυποποιημένα εργαλεία συλλογής και διαχείρισης δεδομένων θα πρέπει να εφαρμόζονται πριν από την ανάπτυξη μοντέλων τεχνητής νοημοσύνης [99].

Τα μοντέλα τεχνητής νοημοσύνης θεωρούνται συχνά αντικειμενικά και, ως εκ τούτου, η ενσωματωμένη προκατάληψη μπορεί να είναι ύπουλη ή να παραβλέπεται εάν τα σύνολα δεδομένων κατάρτισης δεν εξετάζονται προσεκτικά ή δεν αποκαλύπτονται. Ομοίως, η κούρτη ελέγχου, απέναντι από την κούρτη δοκιμής, αναπτύσσεται συνήθως από αντιστοιχισμένα άτομα που δεν έχουν την κατάσταση της νόσου που μελετάται. Ωστόσο, λόγω της συμπερίληψής τους, είναι πιο πιθανό να έχουν συχνή ιατρική επαφή και μπορεί

να μην αντιπροσωπεύουν πραγματικά υγιείς ελέγχους. Επιπλέον, η μεροληψία στα σύνολα δεδομένων κατάρτισης μπορεί να περιορίσει τη δυνατότητα μεταφοράς του μοντέλου τεχνητής νοημοσύνης ή να οδηγήσει στην αδυναμία αναπαραγωγής των αποτελεσμάτων του μοντέλου στα συστήματα υγειονομικής περίθαλψης εκτός εκείνων στα οποία αναπτύχθηκαν και εφαρμόστηκαν [99].

Επιπλέον, τα προβλεπόμενα αποτελέσματα ή τα κλινικά τελικά σημεία που χρησιμοποιούνται για την εκπαίδευση μοντέλων (όπως η χρήση φροντίδας, το κόστος περίθαλψης, οι συνταγές κ.λπ.) πρέπει να επιλέγονται προσεκτικά για να διασφαλιστεί ότι δεν σχετίζονται με υποκείμενες κοινωνικοοικονομικές προκαταλήψεις, ειδικά εκείνες που προκύπτουν από εγγενείς ανισότητες στην παροχή υγειονομικής περίθαλψης που οδηγούν συστηματικά σε ανισότητες στα αποτελέσματα. Ένα χαρακτηριστικό παράδειγμα αυτού αποδείχθηκε σε μια μελέτη που αξιολόγησε την ικανότητα ενός εμπορικού αλγορίθμου, εκπαιδευμένου να χρησιμοποιεί το κόστος υγειονομικής περίθαλψης ως υποκατάστατο της ανεξέλεγκτης ασθένειας, να προβλέπει την ανάγκη των ασθενών για επιπλέον φροντίδα [99].

Ο αλγόριθμος αναγνώρισε τους μαύρους ασθενείς ως υγιέστερους (λόγω του χαμηλότερου σχετικού κόστους υγειονομικής περίθαλψης) από τους λευκούς ασθενείς ίσου κινδύνου και αυτό με τη σειρά του μείωσε τον αριθμό των μαύρων ασθενών που εντοπίστηκαν για επιπλέον φροντίδα κατά περισσότερο από το ήμισυ. Καθώς λιγότερα χρήματα δαπανώνται σε μαύρους ασθενείς που έχουν ίση ανάγκη με τους λευκούς ασθενείς, κυρίως λόγω της άνισης πρόσβασης στην περίθαλψη, ο αλγόριθμος κατέληξε εσφαλμένα στο συμπέρασμα ότι οι μαύροι ασθενείς είναι πιο υγιείς από τους λευκούς ασθενείς [98].

Για να αποφευχθεί ή να ελαχιστοποιηθεί η προκατάληψη από την εισαγωγή σε αλγόριθμους AI και CDSS, είναι επιτακτική ανάγκη τα σύνολα δεδομένων εκπαίδευσης και τα κλινικά τελικά σημεία να περιλαμβάνουν τις υποεκπροσωπούμενες κοόρτες και τις ρυθμίσεις υγειονομικής περίθαλψης που προορίζονται να εξυπηρετήσουν. Εάν αυτό δεν συμβεί, θα επηρεαστεί η ακρίβεια και η επεκτασιμότητα του μοντέλου και θα διαδοθούν οι υπάρχουσες ανισότητες στον τομέα της υγειονομικής περίθαλψης και οι συστημικές προκαταλήψεις [98].

Οι μέθοδοι για τον εντοπισμό της μεροληψίας στα σύνολα δεδομένων κατάρτισης βρίσκονται υπό ανάπτυξη και οι διαδικασίες για την υποστήριξη της χρήσης τους στην



ανάπτυξη μοντέλων θα πρέπει να αποτελέσουν απαίτηση. Επιπλέον, η τυποποιημένη αναφορά του πηγαίου κώδικα του αλγορίθμου και των συνθηκών εκπαίδευσης θα ενισχύσει τη διαφάνεια και θα προωθήσει την αναπαραγωγικότητα του μοντέλου σε άλλα παρόμοια περιβάλλοντα συστημάτων υγειονομικής περίθαλψης ή πληθυσμούς ασθενών [98].

Όπως σημειώθηκε προηγουμένως, η αναπαραγωγικότητα της παραγωγής του μοντέλου τεχνητής νοημοσύνης αποτελεί πρόκληση όταν μεταφέρεται σε συστήματα υγειονομικής περίθαλψης και παγκόσμιες κοινότητες. Αλλά ακόμη και στο περιβάλλον στο οποίο αναπτύχθηκε, οι αλγόριθμοι ΑΙ και τα προηγμένα CDSS υπόκεινται σε μετατόπιση δεδομένων με την πάροδο του χρόνου, γεγονός που μπορεί να επηρεάσει την απόδοσή τους [98].

Η μετατόπιση δεδομένων μπορεί να προκληθεί από αλλαγές στη μορφοποίηση ή την ανίχνευση δεδομένων, την ποιότητα των δεδομένων, τη φυσική μετατόπιση που δεν υπήρχε κατά την εκπαίδευση του μοντέλου και την αλλαγή στη σχέση μεταξύ των χαρακτηριστικών (μετατόπιση συμμεταβλητής). Τα πρότυπα θα είναι απαραίτητα για τη συνεχή παρακολούθηση των μοντέλων τεχνητής νοημοσύνης και τη διασφάλιση της διατήρησής τους καθώς εξελίσσονται οι διανομές δεδομένων, οι κοόρτες ασθενών και η πρακτική της ιατρικής [98].

Αν και έχει επισημανθεί η σημασία της αναφοράς και της αναπαραγωγικότητας των τεχνικών τεχνητής νοημοσύνης, αυτό πρέπει να μετριαστεί από τις σημαντικές ανησυχίες για την ασφάλεια των δεδομένων που προκύπτουν από την εργασία με ευαίσθητες ιατρικές πληροφορίες. Ανάλογα με την ποικιλία μάθησης και την υλοποίηση, ορισμένοι αλγόριθμοι επιτρέπουν την αντίστροφη μηχανική του τελικού μοντέλου για τον επαναπροσδιορισμό προηγουμένως αποχαρακτηρισμένων πληροφοριών ασθενών. Ωστόσο, η ψηφιακή εγκληματολογία επιτρέπει πλέον τη συλλογή προσωπικών υπογραφών υγείας από μια ποικιλία μεθόδων δεδομένων, συμπεριλαμβανομένης της αλληλούχισης DNA, της απεικόνισης αξονικής ή μαγνητικής τομογραφίας και των καταλόγων διάγνωσης, ειδικά εάν εκπροσωπούνται σπάνιες ασθένειες [99].

Καθώς τα μοντέλα τεχνητής νοημοσύνης εφαρμόζονται όλο και περισσότερο στην ιατρική πρακτική, θα απαιτηθεί διαχείριση των αλληλεπιδράσεων μηχανής-ανθρώπου και ανάθεση ευθύνης για κλινικές αποφάσεις και σφάλματα. Εκείνοι που ρυθμίζουν την εφαρμογή της τεχνητής νοημοσύνης στην κλινική πρακτική πρέπει να συνεργαστούν με

επιστήμονες και μηχανικούς τεχνητής νοημοσύνης, καθώς και με ιατρούς, για να καλύψουν τα κενά στην ισχύουσα νομοθεσία και να αναπτύξουν κανονισμούς που εξισορροπούν κατάλληλα την ελευθερία καινοτομίας της τεχνολογίας τεχνητής νοημοσύνης με την προστασία των ευάλωτων πληθυσμών [98].

Η επιταχυνόμενη ανάπτυξη της υπολογιστικής ισχύος, τα ολοένα και πιο διαθέσιμα μηχανικά αναγνώσιμα EHRs, η πολυμοική και τα δεδομένα ιατρικής απεικόνισης - καθώς και οι εξελίξεις στην DL, ιδιαίτερα τα συνελκτικά νευρωνικά δίκτυα - έχουν φέρει επανάσταση στην ανάπτυξη και χρήση αλγορίθμων AI και CDSS στην ανάλυση απεικόνισης που σχετίζεται με τον καρκίνο, τη γονιδιωματική και την κλινική πρακτική σε όλο το συνεχές του καρκίνου [98].

Η συνεχιζόμενη έρευνα για την υποστήριξη της εφαρμογής της τεχνητής νοημοσύνης στη γονιδιωματική του καρκίνου αναμένεται να επιτρέψει την έγκαιρη ανίχνευση πολλαπλών καρκίνων και τον προσδιορισμό της θέσης προέλευσης του όγκου. Αυτό μπορεί να μεταμορφώσει τον προσυμπτωματικό έλεγχο του καρκίνου, ιδιαίτερα για λιγότερο διαδεδομένους και σπάνιους καρκίνους, και μπορεί να ενισχύσει τις στρατηγικές επιτήρησης για τους επιζώντες από καρκίνο [96].

Οι συνεχείς εξελίξεις που βασίζονται στην απεικόνιση μπορούν επίσης να οδηγήσουν στην ανάπτυξη μοντέλων που αξιολογούν τον κίνδυνο για διάφορους τύπους καρκίνου, ενισχύουν τη διαγνωστική ακρίβεια των καρκίνων ή προβλέπουν τα σχετικά αποτελέσματα νοσηρότητας και θνησιμότητας. Αυτό μπορεί να επιτρέψει εξατομικευμένες στρατηγικές προσυμπτωματικού ελέγχου και πρόληψης, θεραπευτικές προσεγγίσεις και παρακολούθηση επιζώντων και μπορεί επιπλέον να υποστηρίξει την εικονική βιοψία για την ταξινόμηση των παθολογικών και γονιδιωματικών χαρακτηριστικών που σχετίζονται με τη διάγνωση καρκίνου [96].

### **2.1.2 Δομική μάθηση βασισμένη σε βιο-δίκτυα**

Βιολογικά δίκτυα, όπως δίκτυα μονοπατιών ασθένειας, δίκτυα αλληλεπιδράσεων πρωτεΐνης-πρωτεΐνης (PPIs) και δίκτυα ομοιότητας ασθενειών, παρέχουν δομημένα αποθετήρια γνώσης για την ανακάλυψη των αλληλεπιδράσεων και των ιδιοτήτων των βιολογικών συστημάτων [53]. Οι προσεγγίσεις δικτύου έχουν χρησιμοποιηθεί σε πολλές εργασίες με αξιοσημείωτες ανακαλύψεις στη βιολογία, συμπεριλαμβανομένης της διάγνωσης του καρκίνου, της πρόβλεψης γονιδιωματικής λειτουργίας και της ανακάλυψης φαρμάκων. Επιπλέον, αυτές οι προσεγγίσεις έχουν δείξει ευρεία χρησιμότητα στην αποκάλυψη νέας βιολογίας από το μονοκύτταρο σε επίπεδο πληθυσμού [53].

Η τεχνητή νοημοσύνη σε γραφήματα είναι μια σημαντική και πανταχού παρούσα προσέγγιση για τη χρήση δεδομένων δομημένων γραφημάτων ως πληροφοριών χαρακτηριστικών σε προβλήματα ταξινόμησης και παλινδρόμησης. Το συνελκτικό νευρωνικό δίκτυο γραφήματος (GCN), μία από αυτές τις προσεγγίσεις, συνδυάζει τη δομή του γραφήματος με ένα νευρωνικό δίκτυο για την επίλυση βιολογικών και χημικών προβλημάτων. Οι Rhee et al. [54] πρότεινε ένα υβριδικό μοντέλο που ενσωμάτωσε δύο βασικά στοιχεία GCN και δίκτυο σχέσεων (RN) για την ταξινόμηση του υποτύπου καρκίνου του μαστού. Li et al. [55] ανέπτυξε ένα δίκτυο GCN για την πρόβλεψη του ποσοστού επιβίωσης αποδίδοντας τις βέλτιστες γραφικές αναπαραστάσεις των ολόκληρων διαφανειών του καρκινώματος του πνεύμονα και του εγκεφάλου (WSIs).

## **2.2 AI στην κλινική έρευνα για τον καρκίνο: μέθοδος και εφαρμογή**

Σε συνδυασμό με τον αυξανόμενο πλούτο σε σύγχρονα βιοϊατρικά δεδομένα, την τεχνητή νοημοσύνη και πιο συγκεκριμένα, η DL έχει συγκεντρώσει ορισμένες επιτυχίες στην κλινική έρευνα για τον καρκίνο (Εικόνα 1). Μέθοδοι που βασίζονται στην τεχνητή νοημοσύνη χρησιμοποιούνται όλο και περισσότερο σε διάφορους τομείς της κλινικής έρευνας για τον καρκίνο για τη βελτίωση της ακρίβειας και της αποτελεσματικότητας. Αυτές περιλαμβάνουν τη χρήση της τεχνητής νοημοσύνης στην αναγνώριση απεικόνισης

καρκίνου, τη γονιδιωματική ανάλυση, την εξόρυξη ιατρικών αρχείων, την ανακάλυψη φαρμάκων και τη χρησιμότητα βιοϊατρικής βιβλιογραφίας. Παρακάτω εξετάζουμε τους διαφορετικούς υποτομείς της κλινικής έρευνας για τον καρκίνο, οι οποίοι επωφελήθηκαν από την ενσωμάτωση της ΑΙ.

### 2.2.1 Αναγνώριση απεικόνισης καρκίνου

Με την έλευση των αυξημένων υπολογιστικών δυνατοτήτων και αλγορίθμων, η τεχνητή νοημοσύνη έχει εφαρμοστεί με επιτυχία στην ακτινολογία για να βοηθήσει τον ακτινολόγο στον ορισμό της νόσου [30]. Οι ακατέργαστες εικόνες, πριν τροφοδοτηθούν στο μοντέλο, μπορεί να χρειαστεί να υποβληθούν σε βασική προεπεξεργασία. Για παράδειγμα, για να αποφευχθεί η ανίχνευση άσχετων τμημάτων της εικόνας, τα ROI εξάγονται τμηματοποιώντας τις βλάβες σε μια εικόνα και, στη συνέχεια, μόνο οι πληροφορίες εικόνας εντός των ROI προβλέπονται από το μοντέλο. Οι περιοχές μπορούν να σχολιαστούν από ειδικούς [30] ή να εκχωρηθούν από ετικέτες διάγνωσης [48]. Ωστόσο, σε αντίθεση με άλλους κοινούς τύπους ιατρικών εικόνων, τα WSI είναι πολύ μεγάλα για να υποστούν επεξεργασία από το μοντέλο DL στο σύνολό τους [68].

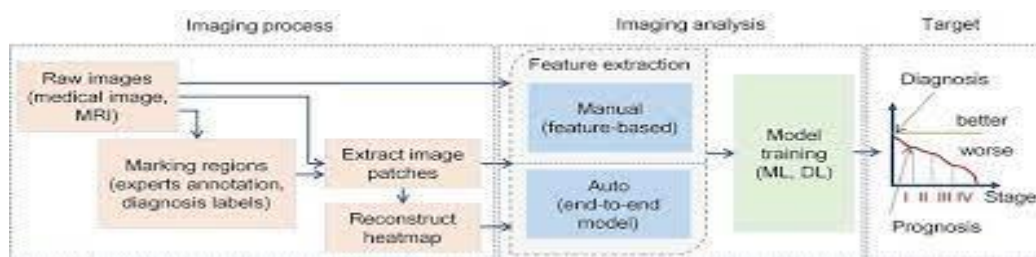
Για να ξεπεραστεί αυτό το ελάττωμα, τα WSI περικόπτονται σε πολυάριθμες μικρές ενημερώσεις εικόνας και στη συνέχεια συγχωνεύουν αυτές τις προβλέψεις επιπέδου patch για να λάβουν πρόβλεψη επιπέδου εικόνας [69]. Μερικές φορές, ένας θερμικός χάρτης πιθανότητας όγκου μπορεί να χρησιμοποιηθεί για την εκτέλεση γεωμετρικής και μορφολογικής επιλογής χαρακτηριστικών, ως είσοδο ενός μοντέλου, και για τον εντοπισμό και τον χαρακτηρισμό μοτίβων ασθενειών σε ψηφιοποιημένες διαφάνειες ιστών [70].

Οι παραδοσιακές προσεγγίσεις αναγνώρισης εικόνας χρησιμοποιούν χειροποίητα, σχεδιασμένα από τον χρήστη χαρακτηριστικά εικόνας, όπως η υφή, το σχήμα, το χρώμα, η πυκνότητα των pixels και η αντίθεση/φωτεινότητα, για να καταγράψουν τη μορφολογία του όγκου ή των κυττάρων [37]. Αυτοί οι αλγόριθμοι που βασίζονται σε χαρακτηριστικά υποφέρουν από ορισμένους περιορισμούς:

- (i) αυτές οι μέθοδοι εξαρτώνται από το βήμα εξαγωγής χαρακτηριστικών [71] και

(ii) αυτά τα χαρακτηριστικά δεν είναι πάντα συνεπή υπό διαφορετικές συνθήκες σάρωσης [1].

Η αυτόματη εξαγωγή χαρακτηριστικών, απαλλαγμένη από το αρχικό βήμα εξαγωγής χαρακτηριστικών, επιτρέπει την απευθείας τροφοδοσία ακατέργαστων εικόνων στο μοντέλο (μοντέλο από άκρο σε άκρο) και την ταυτόχρονη εκτέλεση ταξινόμησης εικόνων (Εικόνα 4) [72].



Εικόνα 4: Αναγνώριση υπολογιστικής απεικόνισης για κλινική έρευνα καρκίνου. Αναγνώριση υπολογιστικής απεικόνισης για κλινική έρευνα καρκίνου. Ένα προγνωστικό μοντέλο για τον καρκίνο (πράσινο κουτί) μπορεί να δημιουργηθεί χρησιμοποιώντας προσεγγίσεις AI σε δεδομένα εικόνας. Το μοντέλο χρησιμοποιεί δεδομένα από κλινικές ακατέργαστες ή προεπεξεργασμένες εικόνες ασθενών. Μόλις επικυρωθεί, το μοντέλο θα μπορούσε να χρησιμοποιηθεί για την πρόβλεψη της διάγνωσης και/ή την εκτίμηση της εξέλιξης του όγκου για να βοηθήσει σημαντικά τους γιατρούς στη λήψη αποφάσεων σχετικά με τη φροντίδα και τη θεραπεία ασθενών με καρκίνο.

Πολλές καθιερωμένες μέθοδοι απεικόνισης έχουν δώσει καλά αποτελέσματα στον προσυμπτωματικό έλεγχο και τη θεραπεία διαφορετικών τύπων καρκίνου με AI. Για παράδειγμα, οι Trebeschi et al. [26] κατασκεύασε έναν ταξινομητή CNN χρησιμοποιώντας πολυπαραμετρική μαγνητική τομογραφία (mpMRI) για να ταξινομήσει κάθε voxel σε όγκο ή μη όγκο.

Στην ανάλυσή τους συμπεριλήφθηκαν μαγνητικές τομογραφίες 140 ασθενών με τοπικά προχωρημένο καρκίνο του ορθού και δύο ειδικοί ακτινολόγοι τμηματοποίησαν κάθε όγκο. Η AUC των χαρτών πιθανοτήτων που προέκυψαν ήταν πολύ υψηλή,  $AUC = 0,99$ . Fan et al. [73] χρησιμοποίησε μια εντελώς ανεπιτήρητη μέθοδο Κυρτής Ανάλυσης Μιγμάτων (CAM) για την πρόβλεψη υποτύπων καρκίνου του μαστού με την αποσύνθεση της MRI με ενισχυμένη αντίθεση (DCE-MRI) από ετερογενείς ιστούς.

Η πρόβλεψη της τάσης για υποτροπή του καρκίνου έχει αναγνωριστεί από τους Wang et al. [36]. Εκπαίδευσαν ένα δίκτυο DL σε 8917 εικόνες CT από την κοόρτη εκμάθησης χαρακτηριστικών για την εξαγωγή των προγνωστικών βιοδεικτών του Υψηλού Βαθμού Ορώδους Καρκίνου των Ωοθηκών (HGSO). Στη συνέχεια, αναπτύχθηκε ένα μοντέλο DL-Cox Proportional Hazard (Cox-PH) για την πρόβλεψη του ατομικού κινδύνου υποτροπής και της πιθανότητας υποτροπής για 3 χρόνια. Μια άλλη πολύτιμη εφαρμογή της ΑΙ είναι η πρόβλεψη των αποτελεσμάτων του καρκίνου, π.χ. επιβιωσιμότητα, προσδόκιμο ζωής, εξέλιξη και ευαισθησία στα φάρμακα του όγκου. Σε πολλές περιπτώσεις, η διαθεσιμότητα της μαστογραφίας έχει επιβεβαιωθεί ως η κύρια μέθοδος απεικόνισης που χρησιμοποιείται για τον έλεγχο του καρκίνου του μαστού με πολλές υπολογιστικές μεθόδους [74].

Li et al. [75] ανέπτυξε μια βελτιωμένη προσέγγιση DL για την ανίχνευση του θηλώδους καρκίνου του θυρεοειδούς σε εικόνες υπερήχων. Σε μια μελέτη των Vang et al. [70], ιστοπαθολογικές διαφάνειες. Fan et al. [76] ανέπτυξε ένα σύστημα υποβοηθούμενης διάγνωσης (CAD) CNN (3D-Mask RCNN) σε περιοχή με 3D μάσκα για ανίχνευση μάζας και τμηματοποίηση με βάση το στήθος στην ψηφιακή τομοσύνθεση μαστού (DBT). Πιο πρόσφατα, τα WSIs έχουν χρησιμοποιηθεί για παθολογική ανάλυση, όπως συμπέρασμα μοριακού υποτύπου, βαθμού όγκου ή κατάσταση υποδοχέα οιστρογόνου [77].

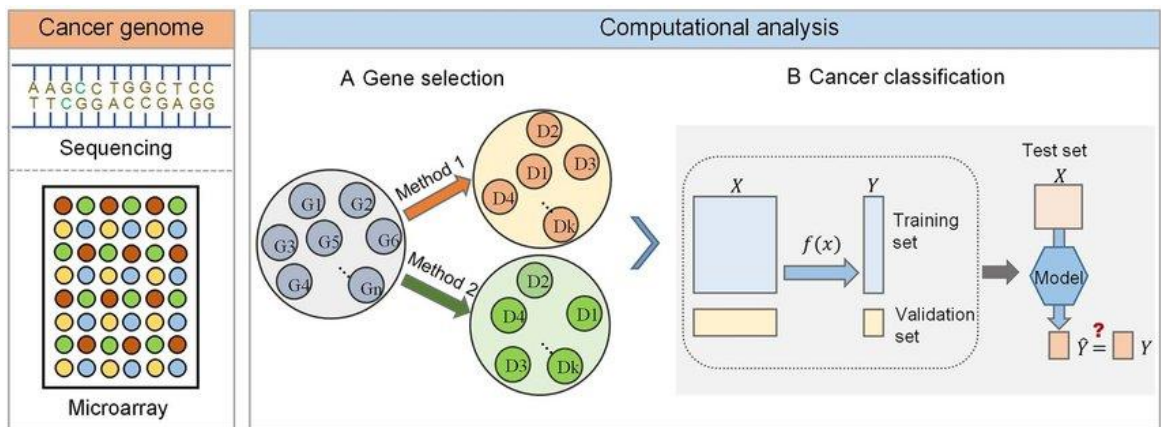
### 2.2.2 Γονιδιωματική ανάλυση

Βασικά, η γονιδιωματική είναι συστηματικές προσεγγίσεις για τον χαρακτηρισμό της λειτουργίας κάθε γονιδιωματικού στοιχείου ενός οργανισμού [78]. Μελέτες συσχέτισης σε επίπεδο γονιδιώματος (GWAS) έχουν εντοπίσει επιτυχώς αλληλεπιδρώντες γενετικές παραλλαγές που συμβάλλουν στον κίνδυνο καρκίνου [29]. Επιπλέον, το μοριακό προφίλ είναι απαραίτητο για την αναγνώριση προγνωστικών βιοδεικτών που σχετίζονται με τους φαινοτύπους του καρκίνου, την πρόγνωση και τα κλινικά αποτελέσματα [79].

Η ανάλυση ενός κυττάρου επιτρέπει ποσοτικές μετρήσεις των τύπων κυττάρων και της μοριακής δραστηριότητας μέσα σε έναν όγκο [80]. Δύο υπολογιστικές αναλύσεις που χρησιμοποιούνται συνήθως στην έρευνα για τον καρκίνο με βάση τη γονιδιωματική περιλαμβάνουν:

- (i) επιλογή γονιδίου (Εικόνα 4A) και
- (ii) ταξινόμηση του καρκίνου (Εικόνα 4B).

Η επιλογή γονιδίου προσπαθεί να επιλέξει γονίδιο υψηλής ρύθμισης ή διαφορετικής έκφρασης και να αφαιρέσει φτωχά από χιλιάδες γονίδια σε πειράματα μικροσυστοιχών αναλύοντας και μετρώντας τα αποτελέσματά τους κατά την κατασκευή ενός ταξινομητή [81].



Εικόνα 5: Υπολογιστική ανάλυση που χρησιμοποιείται συνήθως στην έρευνα για τον καρκίνο στη γονιδιωματική. Το πλαίσιο τεχνητής νοημοσύνης μπορεί να επιλέξει υψηλά ρυθμιζόμενα ή διαφορεικά εκφραζόμενα γονίδια (A) σε διάφορες βασικές εργασίες ταξινόμησης καρκίνου (B) (όπως πρόμνη διάγνωση ασθενειών, ανακάλυψη φαρμάκων ή ταξινόμηση όγκων) χρησιμοποιώντας δεδομένα προσδιορισμού αλληλουχίας ή προφίλ μικροσυστοιχιών. Πηγή:

[https://www.researchgate.net/figure/Computational-analysis-commonly-used-in-cancer-research-on-genomics-AI-framework-can\\_fig3\\_357223392](https://www.researchgate.net/figure/Computational-analysis-commonly-used-in-cancer-research-on-genomics-AI-framework-can_fig3_357223392)

Πρόσφατα, έχει σημειωθεί μεγάλη πρόοδος στην τεχνητή νοημοσύνη στην έρευνα για τον καρκίνο χρησιμοποιώντας διάφορους τύπους γονιδιωματικών δεδομένων ως δεδομένα εισόδου σε μοντέλα. Morais-Rodrigues et al. [82] ανέπτυξε μια τροποποιημένη μέθοδο λογιστικής παλινδρόμησης για την ανάλυση της έκφρασης γονιδίου μικροσυστοιχίας για την εξέλιξη του καρκίνου του μαστού. Επιπλέον, οι Maros et al. [83] ανέπτυξε ροές εργασιών μηχανικής μάθησης για να εκτιμήσει τις πιθανότητες κατηγορίας για διαγνωστικά καρκίνου σε δεδομένα μικροσυστοιχίας μεθυλίωσης DNA.

Πιο πρόσφατα, οι Albaradei et al. [84] παρουσίασε ένα μοντέλο βασισμένο σε DL σε κατάσταση διαφοροποιημένης μετάστασης πανκαρκίνου που βασίζεται σε τρία ετερογενή στρώματα δεδομένων από το TCGA, συμπεριλαμβανομένων των δεδομένων μεθυλίωσης RNA-Seq, microRNA-Seq και DNA. Το μοντέλο χρησιμοποίησε συνελκτικό μεταβλητό

αυτόματο κωδικοποιητή για εξαγωγή χαρακτηριστικών και DNN για ταξινόμηση. Τα αποτελέσματα έδειξαν ότι η ενσωμάτωση δεδομένων μπορεί να βελτιώσει την απόδοση σε σύγκριση με τη χρήση μόνο δεδομένων mRNA. Σε άλλες μελέτες, τα μοντέλα AI έχουν επικεντρωθεί στην πρόβλεψη του βαθμού καρκίνου.

Σε μια μελέτη των Yamamoto et al. [17], ένας ταξινομητής SVM εκπαιδεύτηκε στη μορφολογική ταξινόμηση των μικροπεριβαλλοντικών μυοεπιθηλιακών κυττάρων για την ποσοτική διάγνωση όγκων του μαστού. Μέτρησαν ποσοτικά 11 661 πυρήνες σε τέσσερις ιστολογικούς τύπους: φυσιολογικές περιπτώσεις, συνήθη υπερπλασία του πόρου και χαμηλής/υψηλού βαθμού πόρων καρκίνωμα in situ (DCIS). Τουλάχιστον τρεις παθολόγοι διέγνωσαν και βαθμολόγησαν όλες τις περιπτώσεις ανεξάρτητα, και αυτό το μοντέλο ήταν σε θέση να ταξινομήσει τους τέσσερις ιστολογικούς τύπους με ακρίβεια 90,9%.

Συγκεκριμένα, οι βιοδείκτες που σχετίζονται με την ασθένεια μπορούν να αναγνωριστούν από γονιδιωματικά δεδομένα. Για παράδειγμα, οι Zeng et al. [85] χρησιμοποίησε βαθιά δάση σε συνδυασμό με μεθόδους εκμάθησης χωρίς σήμανση θετικών για την πρόβλεψη πιθανών circRNA που σχετίζονται με ασθένειες. Οι Radhakrishnan et al. [20] συνδύασε την απεικόνιση φθορισμού και τη βαθιά μάθηση για την ανίχνευση λεπτών αλλαγών στην πυρηνική μορφολογία σε ανάλυση ενός κυττάρου και άνοιξε νέους δρόμους για πρόωμη διάγνωση ασθενειών και ανακάλυψη φαρμάκων.

### **2.2.3 Προκλήσεις και μελλοντικές κατευθύνσεις**

Η τεχνητή νοημοσύνη έχει επιδείξει συγκρίσιμες επιδόσεις με αυτές ενός ειδικού σε κοινά πεδία εφαρμογών σε μια σειρά βιοϊατρικής. Ωστόσο, αν και ορισμένες λύσεις τεχνητής νοημοσύνης είναι ήδη διαθέσιμες, εξακολουθούν να υπάρχουν πολλές προκλήσεις για την τεχνητή νοημοσύνη να περάσει από τις θεωρητικές μελέτες στις εφαρμογές του πραγματικού κόσμου.

Η αποδοχή των τεχνολογιών τεχνητής νοημοσύνης στην ιατρική παρεμποδίζεται από την πανταχού παρούσα αναφερόμενη φύση του μηχανισμού «μαύρου κουτιού», ιδιαίτερα όταν εξετάζονται προσεγγίσεις που βασίζονται σε DL και νευρωνικά δίκτυα, οι οποίες βασίζονται σε περίπλοκα κρυφά επίπεδα αλληλεπίδρασης δεδομένων. Αν και οι



πρωτόγονοι αλγόριθμοι ML, όπως η γραμμική παλινδρόμηση, είναι πλήρως διαφανείς στη λειτουργία, πολλές σύγχρονες προσεγγίσεις χρησιμοποιούν στρατηγικές που περιλαμβάνουν τη δημιουργία πολλών χιλιάδων επικαλυπτόμενων δέντρων αποφάσεων με περίπλοκα συστήματα ενίσχυσης που δεν μπορούν να αναπαρασταθούν γραφικά σε οποιοδήποτε χρησιμοποιήσιμο βαθμό [100-102].

Η ερμηνευτικότητα περιπλέκεται περαιτέρω από την DL, η οποία βασίζεται σε κρυφά στρώματα αλληλεπίδρασης δεδομένων εμπνευσμένα από τη διασύνδεση των νευρώνων και των συνάψεων του εγκεφάλου. Η σημασία κάθε μεταβλητής που χρησιμοποιείται για τη μοντελοποίηση ενός αποτελέσματος μπορεί να εκτιμηθεί με αντίστροφη μηχανική αυτών των προσεγγίσεων AI μέσω της έννοιας των τιμών Shapley. Αυτές οι τιμές δημιουργούνται εξετάζοντας μία μόνο μεταβλητή και επεξεργάζοντας επαναληπτικά τα ίδια δεδομένα μέσω ενός μοντέλου, ενώ τροποποιούν μόνο την εν λόγω μεταβλητή για να εκτιμήσουν το μέγεθος της επίδρασης που έχει στο αποτέλεσμα. Οι τιμές Shapley μπορούν να δημιουργηθούν για ένα μοντέλο για την κατάταξη κάθε μεταβλητής κατά σπουδαιότητα, αλλά μπορούν επίσης να χρησιμοποιηθούν σε ατομικό επίπεδο. Αυτή η διάκριση επιτρέπει στον κλινικό ιατρό να ερμηνεύσει την αιτιότητα από την τεχνητή νοημοσύνη [103, 104].

Ένα άλλο εμπόδιο στην υιοθέτηση είναι η αντιληπτή δυσκολία πλοήγησης σε μια μελέτη που βασίζεται στην τεχνητή νοημοσύνη. Οι κύριοι ερευνητές δεν χρειάζεται να ασχοληθούν με την τεχνική εκτέλεση της επιστήμης, αλλά αντίθετα θα πρέπει να μάθουν τι είδους ερωτήσεις η τεχνητή νοημοσύνη είναι μοναδικά κατάλληλη για να απαντήσει. Οι παραδοσιακές στατιστικές μέθοδοι χρησιμοποιούνται για την αξιολόγηση των σχέσεων μεταξύ μεταβλητών και την παροχή κατευθυνόμενων δοκιμών υποθέσεων, ενώ η τεχνητή νοημοσύνη στοχεύει στη μοντελοποίηση ενός πολύπλοκου συστήματος και στην παροχή ακριβών προβλέψεων.

Μια κατάλληλη ερώτηση μπορεί να προσπαθήσει να ταξινομήσει ένα σύστημα σε δυαδικά αποτελέσματα, όπως η πρόβλεψη της επιτυχίας μιας μεθόδου θεραπείας. για την πρόβλεψη ενός συνεχούς αποτελέσματος, όπως η χρήση παλινδρόμησης Cox για την εκτίμηση της διάρκειας επιβίωσης. ή να ομαδοποιήσετε δεδομένα σε άγνωστο αριθμό κλάδων, όπως ο προσδιορισμός ομάδων χαρακτηριστικών που συσχετίζονται με υποτύπους ασθενειών. Περισσότερα πανεπιστήμια και ακαδημαϊκά ιατρικά κέντρα προσθέτουν πυρήνες AI (με αναλυτές δεδομένων, προγραμματιστές AI και υπολογιστικούς πόρους) παράλληλα με τους πυρήνες στατιστικής και γονιδιωματικής τους. Ακριβώς όπως οι

ερευνητές του καρκίνου εργάζονται με γονιδιωματικά δεδομένα παρά το γεγονός ότι λίγοι έχουν πρωτογενή εμπειρία βιοπληροφορικής, η έρευνα για τον καρκίνο που υποστηρίζεται περισσότερο προσιτή χωρίς λειτουργική γνώση του προγραμματισμού υπολογιστών όταν υποστηρίζεται από ισχυρές βασικές υπηρεσίες και πόρους AI.

Αν και έχει σημειωθεί σημαντική αύξηση στην ανάπτυξη αλγορίθμων τεχνητής νοημοσύνης που σχετίζονται με τον καρκίνο και προηγμένων CDSS, εξακολουθεί να υπάρχει έλλειψη έρευνας που σχετίζεται με την προοπτική επικύρωσή τους όταν εφαρμόζονται στη συνήθη κλινική πρακτική, είτε για να αντικαταστήσουν είτε για να αυξήσουν την ανθρώπινη νοημοσύνη. Συγκεκριμένα, υπάρχουν λίγες τυχαιοποιημένες ελεγχόμενες δοκιμές για να αποδειχθεί εάν αυτά τα εργαλεία AI βελτιώνουν τα αποτελέσματα των ασθενών, αυξάνουν την αποτελεσματικότητα του παρόχου ή / και είναι οικονομικά αποδοτικά [104,105].

Η υιοθέτηση μοντέλων τεχνητής νοημοσύνης στην πρακτική του καρκίνου θα πρέπει να βασίζεται σε αποδεικτικά στοιχεία, έτσι ώστε να οδηγούν σε μειωμένη νοσηρότητα και θνησιμότητα και/ή σε παρόμοια κλινικά αποτελέσματα που επιτυγχάνονται πιο αποτελεσματικά ή λιγότερο δαπανηρά. Απαιτούνται επενδύσεις για την υποστήριξη πολυκεντρικών, ρεαλιστικών κλινικών δοκιμών για την αξιολόγηση του αντίκτυπου των αλγορίθμων τεχνητής νοημοσύνης και στα πολυεπίπεδα αποτελέσματα των ασθενών, των παρόχων και των συστημάτων. Τα διδάγματα που αντλήθηκαν από τις προοπτικές μελέτες αναμένεται να οικοδομήσουν την εμπιστοσύνη των κλινικών ιατρών σε αυτά τα μοντέλα τεχνητής νοημοσύνης και τη χρήση τους στη φροντίδα των ασθενών, βοηθώντας έτσι να ξεπεραστούν ορισμένοι από τους προαναφερθέντες περιορισμούς [105].

Η εφαρμογή της τεχνητής νοημοσύνης στην πρακτική του καρκίνου περιλαμβάνει την παροχή υποστήριξης κλινικών αποφάσεων για τη διάγνωση και τον προσυμπτωματικό έλεγχο του καρκίνου, την επεξεργασία ιατρικών δεδομένων για την ανίχνευση του καρκίνου ή τον χαρακτηρισμό της πρόγνωσης των ασθενών και τη βελτιστοποίηση της παροχής φροντίδας και των κλινικών λειτουργιών αυξάνοντας την ικανότητα του συστήματος και κατανέμοντας πόρους. Παρόλο που οι αλγόριθμοι AI και τα προηγμένα CDSS υπόσχονται πολλά για την παροχή υγειονομικής περίθαλψης, υπάρχουν αρκετές προκλήσεις που πρέπει να αντιμετωπιστούν για τη βελτιστοποίηση της αναπαραγωγιότητάς τους, τη βελτίωση της απόδοσής τους, την ελαχιστοποίηση των

προκαταλήψεων, τη διευκόλυνση της δικαιοσύνης και τη διατήρηση της ακρίβειάς τους με την πάροδο του χρόνου [106].

Ένας σημαντικός περιορισμός στην ευρεία εφαρμογή των αλγορίθμων τεχνητής νοημοσύνης και στην παροχή φροντίδας για τον καρκίνο είναι η απαίτηση για ποικίλα και χωρίς αποκλεισμούς σύνολα δεδομένων για την κατάρτιση. Ο πληθυσμός των ασθενών για την εφαρμογή και τη χρήση αυτών των μοντέλων θα πρέπει να αντικατοπτρίζει τον πληθυσμό από τον οποίο ελήφθησαν τα δεδομένα εκπαίδευσης. Εάν αυτό δεν συμβεί και ορισμένοι πληθυσμοί ή σενάρια υπερεκπροσωπούνται ή υποεκπροσωπούνται, η μεροληπτική δειγματοληψία μπορεί να οδηγήσει σε κακή απόδοση του μοντέλου, ανακριβείς προβλέψεις, ακόμη και πιθανή βλάβη [106].

Αυτός ο περιορισμός είναι συγκρίσιμος με αυτόν που παρατηρείται με την πραγματική εφαρμογή των αποτελεσμάτων των κλινικών δοκιμών σε διαφορετικούς πληθυσμούς ασθενών, συμπεριλαμβανομένων κοορτών που υποεκπροσωπούνται σε αυτές τις δοκιμές (ηλικιωμένοι, έφηβοι και νεαροί ενήλικες, άτομα φυλετικών και εθνοτικών μειονοτήτων, αγροτικές, υποεξυπηρετούμενες και μη εξυπηρετούμενες κοινότητες). Για παράδειγμα, ένα πολυωμικό προγνωστικό μοντέλο που ενσωματώνει γονιδιωματικά δεδομένα μπορεί να μην έχει καλή απόδοση για άτομα φυλετικών και εθνοτικών μειονοτήτων, επειδή τα περισσότερα αποθετήρια αναφοράς σχολιασμένων γονιδιωματικών δεδομένων που χρησιμοποιούνται για την ανάπτυξη μοντέλων κατασκευάζονται κυρίως από άτομα βορειοευρωπαϊκής καταγωγής [106].

Επί του παρόντος, μια από τις μεγαλύτερες προκλήσεις που αντιμετωπίζει η τεχνητή νοημοσύνη, γενικά, είναι η ανάγκη για δεδομένα. Η απόκτηση επαρκούς μεγάλου, δημόσιου, καλά σχολιασμένου δεδομένων καρκίνου είναι μια συνεχής ανάγκη για την τεχνητή νοημοσύνη. Αν και η συμπερίληψη εικόνων, γονιδιωματικών δεδομένων και κλινικών αποτελεσμάτων σε ορισμένες ανοιχτές βάσεις δεδομένων είχε σημαντικό αντίκτυπο στην ενίσχυση της υπολογιστικής κλινικής έρευνας.

Η κλίμακα, η ποιότητα και η ποικιλία των τύπων δεδομένων, όπως το ιστορικό ασθενών από προηγούμενες αναφορές, είναι δυνητικά σχετικές με τον κίνδυνο και την εξέλιξη του καρκίνου, αλλά είναι χρονοβόρα η συλλογή τους. Η συμφωνία κοινής χρήσης δεδομένων μπορεί να διαδραματίσει σημαντικό ρόλο στην αντιμετώπιση της παραπάνω πρόκλησης. Η κοινή χρήση μεγάλων συνόλων δεδομένων με την κοινότητα μπορεί να

ενεργοποιηθεί με το cloud computing και την προηγμένη ανάπτυξη της επόμενης γενιάς προγνωστικών μοντέλων καρκίνου.

Επιπλέον, η επιτυχής ανάπτυξη ενός μοντέλου AI εξαρτάται από τα υψηλής ποιότητας δεδομένα. Παρά το γεγονός ότι ο όγκος και η ποικιλία των διαθέσιμων δεδομένων αυξάνεται, η αξιολόγηση της ποιότητας των δεδομένων δεν είναι τυποποιημένη.

Επιπλέον, ορισμένες κλινικές εργασίες, όπως η πρόβλεψη της πρόγνωσης, είναι πιο αδόμητες από τις παραδοσιακές εργασίες βαθιάς μάθησης [107]. Μερικές φορές, πρέπει να δώσουμε ακριβείς προβλέψεις (π.χ. χρόνους επιβίωσης) από έναν συνδυασμό κλινικοπαθολογικών, γονιδιωματικών δεικτών και εικόνων που έχουν πολύ υψηλότερη ανάλυση. Επιπλέον, οι ασθενείς καλύπτουν μια μεγάλη ποικιλία τύπων καρκίνου και συχνά λείπουν κάποια μορφή κλινικών, απεικονιστικών ή γονιδιωματικών δεδομένων, γεγονός που καθιστά δύσκολη την εφαρμογή της τεχνητής νοημοσύνης.

Παρά το γεγονός ότι η τεχνητή νοημοσύνη επιτυγχάνει τακτικά υψηλές επιδόσεις στην ιατρική έρευνα, η υιοθέτηση της τεχνητής νοημοσύνης σε πραγματικές περιπτώσεις είναι περιορισμένη λόγω της κάπως αδιαφάνειας του μοντέλου. Το μηχάνημα δεν μπορούσε να εξηγήσει πώς ήξερε και γιατί πήρε αυτό το αποτέλεσμα. Αυτό αναφέρεται συχνά ως πρόβλημα του «μαύρου κουτιού». Είναι δύσκολο να παρουσιαστεί ποια χαρακτηριστικά των δεδομένων εισόδου συμβάλλουν στην έξοδο. Για παράδειγμα, η τεχνητή νοημοσύνη μπορεί να προβλέψει τη βέλτιστη θεραπεία για έναν ασθενή, αλλά δεν παρέχει το σκεπτικό που χρησιμοποίησε για να κάνει αυτή την πρόβλεψη. Το ερμηνεύσιμο DL είναι μια τάση για την άμβλυνση αυτού του περιορισμού [107].

Επιπλέον, το χάσμα γνώσης μεταξύ των ειδικών της κλινικής και της επιστήμης δεδομένων εξακολουθεί να παρουσιάζει σημαντικές προκλήσεις. Οι γιατροί έχουν μεγάλη εμπειρία με την ογκολογική επεξεργασία και διαχείριση, ενώ οι επιστήμονες δεδομένων διαθέτουν γνώση υψηλού επιπέδου στην επιστήμη δεδομένων για την κατανόηση των μηχανισμών τεχνητής νοημοσύνης. Θα πρέπει να επιδιωχθεί περαιτέρω συνεργασία μεταξύ ειδικών της κλινικής και της επιστήμης δεδομένων για να γεφυρωθεί το χάσμα μεταξύ τους [107].

Επεξεργασία φυσικής γλώσσας- και οι αλγόριθμοι ML χρησιμοποιούνται όλο και περισσότερο για την αυτοματοποίηση του CDSS και η ενσωμάτωσή τους στο EHR αναμένεται να ενισχύσει την υιοθέτησή τους σε κλινικές ροές εργασίας. Δεδομένης της

ευρείας δυναμικής εφαρμογής της τεχνητής νοημοσύνης στην ογκολογία, η παρούσα ανασκόπηση θα επικεντρωθεί στις τεχνολογίες και τους αλγόριθμους που υποστηρίζουν άμεσα τη φροντίδα των ασθενών με καρκίνο.

Ο καρκίνος βασίζεται σε μεγάλο βαθμό σε τεκμηριωμένα συστήματα βαθμολόγησης φαρμάκων για την αξιολόγηση του κινδύνου καρκίνου και τη διάγνωση ασθενειών, την προγνωστική σταδιοποίηση, τη θεραπεία και την παρακολούθηση της επιτήρησης. Αυτά τα συστήματα συχνά προέρχονται από απλές παρατηρήσεις χρησιμοποιώντας μικροσκοπία φωτός και αυξήθηκαν σε αποτελεσματικότητα με την εισαγωγή πιο προηγμένων δοκιμών, όπως δοκιμές γονιδιακής έκφρασης και αλληλούχιση επόμενης γενιάς σωματικών και γεννητικών γονιδιωμάτων.

Το αποτέλεσμα αυτού του εκσυγχρονισμού είναι ένας συνεχώς διευρυνόμενος κατάλογος προγνωστικών και προγνωστικών παραγόντων που σχετίζονται με μια συγκεκριμένη ασθένεια, όπως φαίνεται από τον αυξανόμενο επιπολασμό κλινικών μοντέλων που ενημερώνονται για το γονιδίωμα. Ωστόσο, κάθε πρόσθετος προγνωστικός παράγοντας ανυψώνει την πολυπλοκότητα ενός μοντέλου και δημιουργεί γρήγορα ένα πλέγμα αλληλεπιδράσεων μεταξύ αναδυόμενων και καθιερωμένων παραγόντων ασθένειας που είναι ακατανόητο χρησιμοποιώντας παραδοσιακές προσεγγίσεις, αν και αυξάνει τη δυναμική δύναμη επίλυσης των σχετικών αλγορίθμων μοντελοποίησης.

Για να προωθήσουν την ογκολογία ακριβείας και να παρέχουν μια ακριβή ερμηνεία της κατάστασης του καρκίνου ενός ατόμου, είναι απαραίτητο τόσο οι ερευνητές όσο και οι κλινικοί γιατροί να χρησιμοποιήσουν όλες τις διαθέσιμες πληροφορίες που θα επιτρέψουν στην πολυπλοκότητα του υπολογιστικού μοντέλου να προσεγγίσει την πολυπλοκότητα του βιολογικού συστήματος. Ο μόνος εφικτός τρόπος σύνθεσης του μεγέθους και της αλληλεξάρτησης τέτοιων πολυτροπικών δεδομένων είναι η τεχνητή νοημοσύνη, αξιοποιώντας υπολογιστική υψηλών επιδόσεων και πρωτοποριακές τεχνικές DL.

Μια αναδυόμενη στρατηγική προσυμπτωματικού ελέγχου του καρκίνου είναι η ανάπτυξη ανίχνευσης του καρκίνου του ολικού αίματος από βαθιά αλληλούχιση. Το ολικό αίμα είναι ελκυστικό για ανάλυση δεδομένης της άμεσης προσβασιμότητάς του και του γεγονότος ότι όλα τα κύτταρα του σώματος, είτε άμεσα είτε έμμεσα, έχουν πρόσβαση στο κυκλοφορικό σύστημα.

Οι επαγγελματίες – ιατροί του καρκίνου επωφελούνται τώρα από τα πρώιμα εργαλεία τεχνητής νοημοσύνης που μόλις γίνονται διαθέσιμα στο EHR και δεν υπάρχει έλλειψη κραυγής για μελλοντικές επιπτώσεις. Η χρησιμότητα της τεχνητής νοημοσύνης στην ογκολογία μπορεί να εκτιμηθεί με την απεικόνιση της υγείας κάθε ασθενούς ως ψηφιακή φωτογραφία. Η ιατρική επιχειρεί να αποκρυπτογραφήσει αυτή την εικόνα σε βάση pixel-by-pixel, χρησιμοποιώντας μελέτες που έχουν τις ρίζες τους στην εξηγήσιμη παθοφυσιολογία για να αποσαφηνίσει με ακρίβεια μεμονωμένα εικονοστοιχεία που είναι διάσπαρτα παντού. Εν τω μεταξύ, η τεχνητή νοημοσύνη υιοθετεί μια προσέγγιση ολόκληρης της εικόνας, ξεκινώντας ως μια κακώς εστιασμένη προσέγγιση, αλλά με αυξανόμενο ορισμό καθώς συνεχίζονται οι βελτιώσεις στις βάσεις δεδομένων, τους αλγόριθμους και την υπολογιστική ισχύ.

Ένα άλλο σημαντικό ζήτημα για την τεχνητή νοημοσύνη είναι ο ρόλος της. Είναι σχεδόν αδύνατο να εκτελέσετε μια τεχνητή νοημοσύνη χωρίς ειδικούς. Η τεχνητή νοημοσύνη δεν πρέπει να θεωρείται ως μια αυτόνομη λύση σε ένα εντελώς μη εποπτευόμενο περιβάλλον. Αντιθέτως, είναι ένας βοηθητικός βοηθός στους ειδικούς, καθώς και ένα εργαλείο που μπορεί να βοηθήσει σε τομείς όπου οι ανθρώπινες δυνατότητες παραμένουν περιορισμένες.

Στο μέλλον, θεωρείται ότι η τεχνητή νοημοσύνη θα συμμετάσχει κλινικά στη θεραπεία του καρκίνου και θα χρησιμοποιηθεί για την επίσπευση της διάγνωσης, της θεραπείας και ακόμη και της θεραπείας. Επιπλέον, αναμένεται ότι η τεχνολογία AI θα είναι ευρύτερα διαθέσιμη και θα εφαρμοστεί για την ενίσχυση των ποσοστών επιβίωσης, τη βελτίωση των ανταποκρίσεων στη θεραπεία και τη μείωση των παρενεργειών.

## Συμπεράσματα

---

Σκοπός της εργασίας είναι να αναδείξει την εφαρμογή της τεχνητής νοημοσύνης στον τομέα της ιατρικής, καθώς η λειτουργία της τεχνητής νοημοσύνης στοχεύει στη διαμόρφωση ενός πλαισίου που είναι κατάλληλα εκπαιδευμένο για να καθοδηγεί με ακρίβεια εάν ένας ασθενής θα χρειαστεί θεραπεία ή όχι. Η τεχνητή νοημοσύνη είναι ικανή να προτείνει ποιος ασθενής χρειάζεται περισσότερες εξετάσεις, όπως το WGS, και να υπολογίσει ποιο φάρμακο θεραπείας θα βοηθήσει περισσότερο έναν ασθενή στην ανάρρωση.

Η τεχνητή νοημοσύνη μπορεί να διαχειριστεί μη γραμμικές πολύπλοκες σχέσεις, πολυεπεξεργασία, ανοχή σφαλμάτων και μάθηση. Δεδομένου ότι μπορεί να αλλάξει από μόνη της καθώς και να ασχοληθεί τόσο με ποσοτικές όσο και με ποιοτικές πληροφορίες ταυτόχρονα. Τα διάφορα κλινικά δεδομένα σε διάφορους τομείς έχουν δείξει τη λειτουργική εφαρμογή του. Η τεχνητή νοημοσύνη εφαρμόζεται επί του παρόντος σε μορφές διακύμανσης στην κλινική και υγειονομική περίθαλψη.

Στην εποχή της ταχέως αναπτυσσόμενης τεχνολογίας, το μέλλον των ιατρικών επιστημών, ειδικά στη διάγνωση και τη θεραπεία του καρκίνου, είναι ασφαλές λόγω των πιθανών εφαρμογών της τεχνητής νοημοσύνης και της νομιμοποίησης εσόδων από παράνομες δραστηριότητες στις κλινικές διαδικασίες, ως εκ τούτου αντιμετώπισε προβλήματα και περιορισμούς που αφορούν στο ιατρικό μέρος της. Με στόχο η τεχνητή νοημοσύνη να βοηθήσει στον τομέα της διάγνωσης και κατηγοριοποίησης του καρκίνου με ταχύτερα τρόπο και με διαρκώς αυξανόμενο ποσοστό επιτυχίας αποτελεσμάτων, η τεχνητή νοημοσύνη παρέχει ταχύτερους χάρτες καθοδήγησης για τη θεραπεία των ατόμων που αποτελούν βασική ανάγκη του μέλλοντος με αυξανόμενο πληθυσμό. Λόγω της ακρίβειας της τεχνητής νοημοσύνης, οι επιστήμονες και οι τεχνολόγοι απέκτησαν την εμπιστοσύνη να βελτιώσουν τις παραδοσιακές μεθόδους προόδου. Συγκεκριμένα, η χρήση της τεχνολογίας τεχνητής νοημοσύνης στον προσυμπτωματικό έλεγχο του καρκίνου επιτυγχάνει ειδικότητα 89% και ευαισθησία 76%, οι οποίες είναι σημαντικά υψηλότερες σε σύγκριση με τα ποσοστά για τα τυπικά συστήματα διάγνωσης με τη βοήθεια υπολογιστή (CAD), τα οποία είναι περίπου 50%.

Οι προκλήσεις που επισημάνθηκαν στην ανάπτυξη, την εφαρμογή και τη συντήρηση των μοντέλων τεχνητής νοημοσύνης είναι σημαντικές αλλά όχι ανυπέρβλητες. Συνοπτικά,

περιλαμβάνουν την έλλειψη τυποποίησης, συλλογής και διαχείρισης δεδομένων. εγγενείς προκαταλήψεις στα σύνολα δεδομένων κατάρτισης· έλλειψη προτύπων αναφοράς για τον πηγαίο κώδικα και τις συνθήκες κατάρτισης· περιορισμένη προοπτική κλινική επικύρωση μοντέλων· ανάγκη για απρόσκοπτη ενσωμάτωση στις κλινικές ροές εργασίας χωρίς προσθήκη υπαλληλικής ή γνωστικής επιβάρυνσης για τους παρόχους· έλλειψη κανονιστικών και νομικών πλαισίων· και περιορισμένη αναπαραγωγιμότητα όταν μεταφέρεται μεταξύ των συστημάτων υγειονομικής περίθαλψης και των πληθυσμών, δεδομένης της περιορισμένης διαλειτουργικότητας του EHR και της δυναμικής φύσης των δεδομένων και των κατευθυντήριων γραμμών βέλτιστων πρακτικών. Οι ταυτόχρονες επενδύσεις στο ανθρώπινο δυναμικό και τους πόρους που απαιτούνται για την αντιμετώπιση αυτών των προκλήσεων θα πρέπει να πραγματοποιούνται παράλληλα με τις επενδύσεις που προωθούν τις τεχνολογικές καινοτομίες τεχνητής νοημοσύνης.



## Βιβλιογραφία

---

1. Levine AB, Schlosser C, Grewal J, et al. Rise of the machines: advances in deep learning for cancer diagnosis. *Trends Cancer* 2019;5:157–69.
2. Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov* 2019;18 :463 –77.
3. Grace K, Salvatier J, Dafoe A, et al. When will AI exceed human performance? Evidence from AI experts. *J Artif Intell Res* 2018;62:729–54.
4. Tan AC, Gilbert D. Ensemble machine learning on gene expression data for cancer classification. *Appl Bioinformatics* 2003;2(3 Suppl):S75–83.
5. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inform Process Syst* 2012;25:1097–105.
6. Larochelle H, Bengio Y, Louradour J, et al. Exploring strategies for training deep neural networks. *J Mach Learn Res* 2009;10: 1–40. Gu J, Wang Z, Kuen J, et al. Recent advances in convolutional neural networks. *Pattern Recognition* 2018;77:354–77. Medsker LR, Jain LC. Recurrent neural networks: design and applications. Los Angeles: CRC Press, 1999.
7. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 2008;455:1061–8.
8. Strickland E. IBM Watson, heal thyself: how IBM overpromised and underdelivered on AI health care. *IEEE Spectrum* 2019;56: 24–31.
9. Linn A. How Microsoft computer scientists and researchers are working to ‘solve’ cancer [Internet]. *News.microsoft.com* 2028. Available from: <https://news.microsoft.com/stories/computingcancer/> (Ημερομηνία Πρόσβασης 20/09/2022)
10. Singireddy S, Alkhateeb A, Rezaeian I, et al. Identifying differentially expressed transcripts associated with prostate cancer progression using RNA-Seq and machine learning techniques. 2015 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), IEEE 2015;1–5.
11. Wang D, Li JR, Zhang YH, et al. Identification of differentially expressed genes between original breast cancer and xenograft using machine learning algorithms. *Genes* 2018;9:155.
12. Zhang Y, Zhang XF, Lane AN, et al. TFmeta: a machine learning approach to uncover transcription factors governing metabolic reprogramming. In *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* ACM 2018; 351–9.
13. Koelsche C, Schrimpf D, Stichel D, et al. Sarcoma classification by DNA methylation profiling. *Nat Commun* 2021;12:498.
14. Ming F, He T, Peng Z, et al. Diffusion-weighted imaging features of breast tumours and the surrounding stroma reflect intrinsic heterogeneous characteristics of molecular subtypes in breast cancer. *NMR Biomed* 2018;31:e0189302.

15. Yamamoto Y, Saito A, Tateishi A, et al. Quantitative diagnosis of breast tumors by morphometric classification of microenvironmental myoepithelial cells using a machine learning approach. *Sci Rep* 2017;7:46732.
16. Ko J, Bhagwat N, Yee SS, et al. Combining machine learning and nanofluidic technology to diagnose pancreatic cancer using exosomes. *ACS Nano* 2017;11:11182–93.
17. Yuan Y, Shi Y, Li C, et al. DeepGene: an advanced cancer type classifier based on deep learning and somatic point mutations. *BMC Bioinformatics* 2016;17:243–56.
18. Radhakrishnan A, Damodaran K, Soylemezoglu AC, et al. Machine learning for nuclear mechano-morphometric biomarkers in cancer diagnosis. *Sci Rep* 2017;7:17946.
19. Hollon TC, Pandian B, Adapa AR, et al. Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks. *Nat Med* 2020;26:52–8.
20. Guillen P, Ebalunode J. Cancer classification based on microarray gene expression data using deep learning. *International Conference on Computational Science and Computational Intelligence (CSCI), IEEE* 2016; 1403–5.
21. Couture HD, Williams LA, Geradts J, et al. Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. *NPJ Breast Cancer* 2018;4:30.
22. Fan M, Yuan W, Zhao W, et al. Joint prediction of breast cancer histological grade and Ki-67 expression level based on DCEMRI and DWI radiomics. *IEEE J Biomed Health Inform* 2019;24: 1632–42.
23. Fan M, Liu ZH, Xie SD, et al. Integration of dynamic contrast enhanced magnetic resonance imaging and T2-weighted imaging radiomic features by a canonical correlation analysis-based feature fusion method to predict histological grade in ductal breast carcinoma. *Phys Med Biol* 2019;64:215001.
24. Trebeschi S, van Griethuysen JJM, Lambregts DMJ, et al. Deep learning for fully-automated localization and segmentation of rectal cancer on multiparametric MR. *Sci Rep* 2017;7:5301.
25. Schwyzer M, Ferraro DA, Muehlematter UJ, et al. Automated detection of lung cancer at ultralow dose PET/CT by deep neural networks—initial results. *Lung Cancer* 2018;126:170–3.
26. Li H, Giger ML, Huynh BQ, et al. Deep learning in breast cancer risk assessment: evaluation of convolutional neural networks on a clinical dataset of full-field digital mammograms. *J Med Imaging* 2017;4:041304.
27. Behravan H, Hartikainen JM, Tengström M, et al. Machine learning identifies interacting genetic variants contributing to breast cancer risk: a case study in Finnish cases and controls. *Sci Rep* 2018;8:13149.
28. Varghese B, Chen F, Hwang D, et al. Objective risk stratification of prostate cancer using machine learning and radiomics applied to multiparametric magnetic resonance images. *Sci Rep* 2019;9:1570.

29. Choi J, Park S, Ahn J. RefDNN: a reference drug based neural network for more accurate prediction of anticancer drug resistance. *Sci Rep* 2020;10:1861.
30. Chang Y, Park H, Yang HJ, et al. Cancer drug response profile scan (CDRscan): a deep learning model that predicts drug effectiveness from cancer genomic signature. *Sci Rep* 2018;8: 8857.
31. Ing N, Huang F, Conley A, et al. A novel machine learning approach reveals latent vascular phenotypes predictive of renal cancer outcome. *Sci Rep* 2017;7:13190.
32. Fan M, Cheng H, Zhang P, et al. DCE-MRI texture analysis with tumor subregion partitioning for predicting Ki-67 status of estrogen receptor-positive breast cancers. *J Magn Reson Imaging* 2018;48:237–47.
33. Tseng CJ, Lu CJ, Chang CC, et al. Application of machine learning to predict the recurrence-proneness for cervical cancer. *Neural Comput Applic* 2014;24:1311–6.
34. Wang S, Liu Z, Rong Y, et al. Deep learning provides a new computed tomography-based prognostic biomarker for recurrence prediction in high-grade serous ovarian cancer. *Radiother Oncol* 2019;132:171–7.
35. Cha KH, Hadjiiski L, Chan HP, et al. Bladder cancer treatment response assessment in CT using radiomics with deep learning. *Sci Rep* 2017;7:8738.
36. Xu Y, Hosny A, Zeleznik R, et al. Deep learning predicts lung cancer treatment response from serial medical imaging. *Clin Cancer Res* 2019;25:3266–75.
37. Bibault JE, Giraud P, Housset M, et al. Deep learning and radiomics predict complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer. *Sci Rep* 2018;8:12611.
38. Fan M, Chen H, You C, et al. Radiomics of tumor heterogeneity in longitudinal dynamic contrast-enhanced magnetic resonance imaging for predicting response to neoadjuvant chemotherapy in breast cancer. *Front Mol Biosci* 2021;8:622219.
39. Kann BH, Aneja S, Loganadane GV, et al. Pretreatment identification of head and neck cancer nodal metastasis and extra nodal extension using deep learning neural networks. *Sci Rep* 2018;8:14036.
40. Nagpal K, Foote D, Liu Y, et al. Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. *NPJ Digit Med* 2019;2:48.
41. Matheny M, Israni T, Ahmed M, et al. *Artificial Intelligence in Health Care: The Hope, The Hype, The Promise, The Peril*. Washington, DC: National Academy of Medicine Press, 2019.
42. Lao J, Chen Y, Li ZC, et al. A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Sci Rep* 2017;7:10353.
43. Fan M, Xia P, Liu B, et al. Tumour heterogeneity revealed by unsupervised decomposition of dynamic contrast-enhanced magnetic resonance imaging is associated with underlying gene expression patterns and poor survival in breast cancer patients. *Breast Cancer Res* 2019;21:112.
44. Wang JY, Wang XL, Gao X. Non-negative matrix factorization by maximizing correntropy for cancer clustering. *BMC Bioinformatics* 2013;14:107.

45. Huang L, Shao D, Wang Y, et al. Human body-f luid proteome: quantitative profiling and computational prediction. *Brief Bioinform* 2021;22:315–33.
46. Wang S, Chen A, Yang L, et al. Comprehensive analysis of lung cancer pathology images to discover tumor shape and boundary features that predict survival outcome. *Sci Rep* 2018;8:10393.
47. Fan M, Liu ZH, Xu MS, et al. Generative adversarial network based super-resolution of diffusion-weighted imaging: application to tumour radiomics in breast cancer. *NMR Biomed* 2020;33:e4345.
48. Nguyen D, Long T, Jia X, et al. A feasibility study for predicting optimal radiation therapy dose distributions of prostate cancer patients from patient anatomy using deep learning. *Sci Rep* 2019;9:1076.
49. Mobadersany P, Yousefi S, Amgad M, et al. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc Natl Acad Sci USA* 2018;15:E2970–9.
50. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115–8.
51. Camacho DM, Collins KM, Powers RK, et al. Next-generation machine learning for biological networks. *Cell* 2018;173: 1581–92.
52. Rhee S, Seo S, Kim S. Hybrid approach of relation network and localized graph convolutional filtering for breast cancer subtype classification. 2017; arXiv:1711.05859.
53. Li R, Yao J, Zhu X, et al. Graph CNN for survival analysis on whole slide pathological images. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham 2018; 11071:174–82.
54. Tandel GS, Biswas M, Kakde OG, et al. A review on a deep learning perspective in brain cancer classification. *Cancers (Basel)* 2019;11:111.
55. Dashtban M, Balafar M. Gene selection for microarray cancer classification using a new evolutionary method employing artificial intelligence concepts. *Genomics* 2017;109:91–107.
56. Khan J, Wei JS, Ringnér M, et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 2001;7:673–9.
57. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000;403:503–11.
58. Vo JN, Cieslik M, Zhang Y, et al. The landscape of circular RNA in cancer. *Cell* 2019; 176:869–81.e13.
59. Mottini C, Napolitano F, Li ZX, et al. Computer-aided drug repurposing for cancer therapy: approaches and opportunities to challenge anticancer targets. *Semin Cancer Biol* 2021;68:59–74.
60. Tsherniak A, Vazquez F, Montgomery PG, et al. Defining a cancer dependency map. *Cell* 2017;170:564–76.
61. Iorio F, Knijnenburg TA, Vis DJ, et al. A landscape of pharmacogenomics interactions in cancer. *Cell* 2016;166:740–54.

62. Coker EA, Mitsopoulos C, Tym JE, et al. canSAR: update to the cancer translational research and drug discovery knowledgebase. *Nucleic Acids Res* 2018;47:D917–22.
63. Koscielny G, An P, Carvalho-Silva D, et al. Open targets: a platform for therapeutic target identification and validation. *Nucleic Acids Res* 2017;45:D985–94.
64. Igarashi Y, Nakatsu N, Yamashita T, et al. Open TG-GATEs: a large-scale toxicogenomics database. *Nucleic Acids Res* 2015;43:D921–7.
65. Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;46:D1074–82.
66. Aubreville M, Knipfer C, Oetter N, et al. Automatic classification of cancerous tissue in laserendomicroscopy images of the oral cavity using deep learning. *Sci Rep* 2017;7:11979.
67. Granter SR, Beck AH, Papke DJ. Alphago, deep learning, and the future of the human microscopist. *Arch Pathol Lab Med* 2017;141:619–21.
68. Vang YS, Chen Z, Xie X. Deep learning framework for multiclass breast cancer histology image classification. 2018; arXiv: 1802.00931.
69. Wang X, Yang W, Weinreb J, et al. Searching for prostate cancer by fully automated magnetic resonance imaging classification: deep learning versus non-deep learning. *Sci Rep* 2017;7:15415.
70. Han Z, Wei B, Zheng Y, et al. Breast cancer multi-classification from histopathological images with structured deep learning model. *Sci Rep* 2017;7:4172.
71. Fan M, Zhang P, Wang Y, et al. Radiomic analysis of imaging heterogeneity in tumours and the surrounding parenchyma based on unsupervised decomposition of DCE-MRI for predicting molecular subtypes of breast cancer. *Eur Radiol* 2019;29:4456–67.
72. Geras KJ, Wolfson S, Shen Y, et al. High-resolution breast cancer screening with multi-view deep convolutional neural networks. 2017; arXiv: 1703.07047.
73. Li H, Weng J, Shi Y, et al. An improved deep learning approach for detection of thyroid papillary cancer in ultrasound images. *Sci Rep* 2018;8:6600.
74. Fan M, Zheng HZ, Zheng S, et al. Mass detection and segmentation in digital breast tomosynthesis using 3D-mask region based convolutional neural network: a comparative analysis. *Front Mol Biosci* 2020;7:599333.
75. Gurcan MN, Boucheron LE, Can A, et al. Histopathological image analysis: a review. *IEEE Rev Biomed Eng* 2009;2:147–71.
76. Eraslan G, Avsec Z, Gagneur J, et al. Deep learning: new computational modelling techniques for genomics. *Nat Rev Genet* 2019;20:389–403.
77. Dumbrava EI, Meric-Bernstam F. Personalized cancer therapy leveraging a knowledge base for clinical decision-making. *Cold Spring Harb Mol Case Stud* 2018;4:a001578.
78. Fan J, Slowikowski K, Zhang F. Single-cell transcriptomics in cancer: computational challenges and opportunities. *Exp Mol Med* 2020;52:1452–65.
79. Shao D, Huang L, Wang Y, et al. DeepSec: a deep learning framework for secreted protein discovery in human body fluids. *Bioinformatics* 2021;2021:btab545.

80. Morais-Rodrigues F, Silv'erio-Machado R, Kato RB, et al. Analysis of the microarray gene expression for breast cancer progression after the application modified logistic regression. *Gene* 2020;726:144168.
81. Maros ME, Capper D, Jones DTW, et al. Machine learning workflows to estimate class probabilities for precision cancer diagnostics on DNA methylation microarray data. *Nat Protoc* 2020;15:479–512.
82. Albaradei S, Napolitano F, Thafar MA, et al. MetaCancer: a deep learning-based pan-cancer metastasis prediction model developed using multi-omics data. *Comput Struct Biotechnol J* 2021;19:4404–11.
83. Zeng X, Zhong Y, Lin W, et al. Predicting disease-associated circular RNAs using deep forests combined with positive unlabeled learning methods. *Brief Bioinform* 2020;21:1425–36.
84. Kann BH, Hosny A, Aerts HJWL. Artificial intelligence for clinical oncology. *Cancer Cell* 2021;39:916–27.
85. Xu H, Anderson K, Grann VR, et al. Facilitating cancer research using natural language processing of pathology reports. *Stud Health Technol Inform* 2004;107:565–72.
86. Karimi YH, Blayney DW, Kurian AW, et al. Development and use of natural language processing for identification of distant cancer recurrence and sites of distant recurrence using unstructured electronic health record data. *JCO Clin Cancer Info* 2021;5:469–78.
87. Zeng J, Banerjee I, Henry AS, et al. Natural language processing to identify cancer treatments with electronic medical records. *JCO Clin Cancer Info* 2021;5:379–93.
88. Chan HCS, Shan H, Dahoun T, et al. Advancing drug discovery via artificial intelligence. *Trends Pharmacol Sci* 2019;40: 592–604.
89. Workman P, Antolin AA, Al-Lazikani B. Transforming cancer drug discovery with big data and AI. *Expert Opin Drug Discovery* 2019;14:1089–95.
90. Paul D, Sanap G, Shenoy S, et al. Artificial intelligence in drug discovery and development. *Drug Discov Today* 2021;26: 80–93.
91. Huang C, Clayton EA, Matyunina LV, et al. Machine learning predicts individual cancer patient responses to therapeutic drugs with high accuracy. *Sci Rep* 2018;8:16444.
92. Borisov N, Tkachev V, Suntsova M, et al. A method of gene expression data transfer from cell lines to cancer patients for machine-learning prediction of drug efficiency. *Cell Cycle* 2018;17:486–91.
93. Yanagisawa K, Toratani M, Asai A, et al. Convolutional neural network can recognize drug resistance of single cancer cells. *Int J Mol Sci* 2020;21:3166.
94. Ramsundar B, Eastman P, Walters P, et al. *Deep Learning for the Life Sciences: Applying Deep Learning to Genomics, Microscopy, Drug Discovery and More*. Sebastopol, CA: O'Reilly Media, 2019.

95. Mayr A, Klambauer G, Untertiner T, et al. DeepTox: toxicity prediction using deep learning. *Front Environ Sci* 2016;3:80.
96. Napolitano F, Carrella D, Mandriani B, et al. gene2drug: a computational tool for pathway-based rational drug repositioning. *Bioinformatics* 2017;34:1498–505.
97. Kuhn M, Szklarczyk D, Pletscher-Frankild S, et al. STITCH 4: integration of protein–chemical interactions with user data. *Nucleic Acids Res* 2014;42:D401–7.
98. Senior AW, Evans R, Jumper J, et al. Improved protein structure prediction using potentials from deep learning. *Nature* 2020;577:706–10.
99. Aberle DR, Berg CD, Black WC, et al. The national lung screening trial: overview and study design. *Radiology* 2011;258: 243–53.
100. Jensen MA, Ferretti V, Grossman RL, et al. The NCI genomic data commons as an engine for precision medicine. *Blood* 2017;130: 453–9.
101. Kurnit KC, Bailey AM, Zeng J, et al. "Personalized cancer therapy": a publicly available precision oncology resource. *Cancer Res* 2017;77:e123–6.
102. Yu Y, Wang Y, Xia Z, et al. PreMedKB: an integrated precision medicine knowledgebase for interpreting relationships between diseases, genes, variants and drugs. *Nucleic Acids Res* 2019;47:D1090–101.
103. Griffith M, Spies NC, Krysiak K, et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat Genet* 2017;49:170–4.
104. Shao D, Huang L, Wang Y, et al. HBFP: a new repository for human body fluid proteome. *Database* 2021;2021:baab065. 10.1093/database/baab065.
105. Cheerla A, Gevaert O. Deep learning with multimodal representation for pancancer prognosis prediction. *Bioinformatics* 2019;35:i446–54.
106. Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access* 2018;6: 52138–60.
107. Fortelny N, Bock C. Knowledge-primed neural networks enable biologically interpretable deep learning on single-cell sequencing data. *Genome Biol* 2020;21:190