



Bioinformatika Proteomik Genomik

www.elsevier.com/locate/gpb
www.sciencedirect.com



TINJAUAN

Pembelajaran Mesin untuk Diagnosis, Pengobatan, dan Prognosis Kanker Paru



Yawei Li¹, Xin Wu², Ping Yang³, Guoqian Jiang⁴, Yuan Luo^{1,*}

¹Departemen Pengobatan Pencegahan, Fakultas Kedokteran Feinberg, Universitas Northwestern, Chicago, IL 60611, AS

²Departemen Kedokteran, Universitas Illinois di Chicago, Chicago, IL 60612, AS

³Departemen Ilmu Kesehatan Kuantitatif, Mayo Clinic, Rochester, MN 55905 / Scottsdale, AZ 85259, AS

⁴Departemen Kecerdasan Buatan dan Informatika, Mayo Clinic, Rochester, MN 55905, AS

Diterima 4 Maret 2022; direvisi 3 Oktober 2022; diterima 17 November 2022

Tersedia online 1 Desember 2022

Ditangani oleh Feng Gao

KATA KUNCI

kumpulan data Omics;
 Kumpulan data pencitraan;
 Ekstraksi fitur;
 Ramalan;
 Imunoterapi

Abstrak Perkembangan teknologi pencitraan dan pengurutan terkini memungkinkan kemajuan sistematis dalam studi klinis kanker paru-paru. Sementara itu, pikiran manusia terbatas dalam menangani dan memanfaatkan sepenuhnya akumulasi data dalam jumlah besar secara efektif. Pendekatan berbasis pembelajaran mesin memainkan peran penting dalam mengintegrasikan dan menganalisis kumpulan data yang besar dan kompleks ini, yang telah secara luas mengkarakterisasi kanker paru-paru melalui penggunaan perspektif berbeda dari data yang diperoleh. Dalam ulasan ini, kami memberikan ikhtisar pendekatan berbasis pembelajaran mesin yang memperkuat berbagai aspek diagnosis dan terapi kanker paru-paru, termasuk deteksi dini, diagnosis tambahan, dan prognosis. ramalan, Dan imunoterapi praktik. Selain itu, kami menyoroti tantangan dan peluang penerapan pembelajaran mesin di masa depan pada kanker paru-paru.

Perkenalan

Kanker paru-paru adalah salah satu kanker yang paling sering didiagnosis dan penyebab utama kematian akibat kanker di seluruh dunia. Sekitar 2,20 juta pasien baru didiagnosis menderita kanker paru-paru setiap tahunnya[1], dan 75% di antaranya meninggal dalam waktu lima tahun setelah diagnosis[2]. Heterogenitas intra-tumor (ITH) yang tinggi dan kompleksitas sel kanker sehingga menimbulkan resistensi obat membuat pengobatan kanker menjadi lebih menantang[3]. Selama beberapa dekade terakhir, terus menerus

evolusi teknologi dalam penelitian kanker telah berkontribusi pada banyak proyek kanker kolaboratif besar, yang telah menghasilkan banyak database klinis, pencitraan medis, dan pengurutan[4–6]. Basis data ini memfasilitasi peneliti dalam menyelidiki pola kanker paru secara komprehensif mulai dari diagnosis, pengobatan, dan respons hingga hasil klinis[7]. Secara khusus, penelitian terkini tentang analisis -omics, seperti genomik, transkriptomik, proteomik, dan metabolomik, telah memperluas alat dan kemampuan kami untuk penelitian. Studi kanker sedang mengalami pergeseran ke arah integrasi berbagai tipe data dan ukuran besar. Namun, menggunakan tipe data yang beragam dan berdimensi tinggi untuk tugas klinis memerlukan waktu dan keahlian yang signifikan bahkan dengan bantuan metode reduksi dimensi seperti faktorisasi matriks dan tensor.[8–11], dan menganalisis penyakit terkait kanker yang tumbuh secara eksponensial

* Penulis yang sesuai.

Surel:yuan.luo@northwestern.edu (LuoY).

Tinjauan jawab di bawah tanggung jawab Institut Genomik Beijing, Akademi Ilmu Pengetahuan Tiongkok / Pusat Nasional Bioinformasi dan Masyarakat Genetika Tiongkok.

<https://doi.org/10.1016/j.gpb.2022.11.003>

1672-0229 - 2022 Para Penulis. Diterbitkan oleh Elsevier BV dan Science Press atas nama Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

Ini adalah artikel akses terbuka di bawah lisensi CC BY (<http://creativecommons.org/licenses/by/4.0/>).

database menimbulkan tantangan besar bagi para peneliti. Oleh karena itu, penggunaan model pembelajaran mesin (ML) untuk secara otomatis mempelajari karakteristik internal berbagai jenis data untuk membantu pengambilan keputusan dokter menjadi semakin penting.

ML merupakan subkelompok kecerdasan buatan (AI) yang berfokus pada pembuatan prediksi dengan mengidentifikasi pola dalam data menggunakan algoritma matematika[12]. Ini telah berfungsi sebagai alat bantu dalam fenotip dan terapi kanker selama beberapa dekade[13–19], dan telah diterapkan secara luas dalam pendekatan lanjutan untuk deteksi dini, klasifikasi jenis kanker, ekstraksi tanda tangan, dekonvolusi lingkungan mikro tumor (TME), prediksi prognosis, dan evaluasi respons obat.[20–27]. Di sini, kami menyajikan ikhtisar algoritma ML utama yang telah digunakan untuk mengintegrasikan data biomedis yang kompleks (misalnya, pencitraan atau pengurutan data) untuk berbagai aspek kanker paru-paru (Gambar 1; Tabel S1 dan S2), dan menguraikan tantangan dan peluang utama untuk penerapan ML di masa depan dalam penelitian dan praktik klinis kanker paru-paru. Kami berharap ulasan ini dapat meningkatkan pemahaman yang lebih baik tentang peran dan potensi ML di bidang ini.

Terapkan ML untuk deteksi dini dan diagnosis tambahan kanker paru-paru

ML tentang deteksi dini dan diagnosis menggunakan kumpulan data pencitraan medis

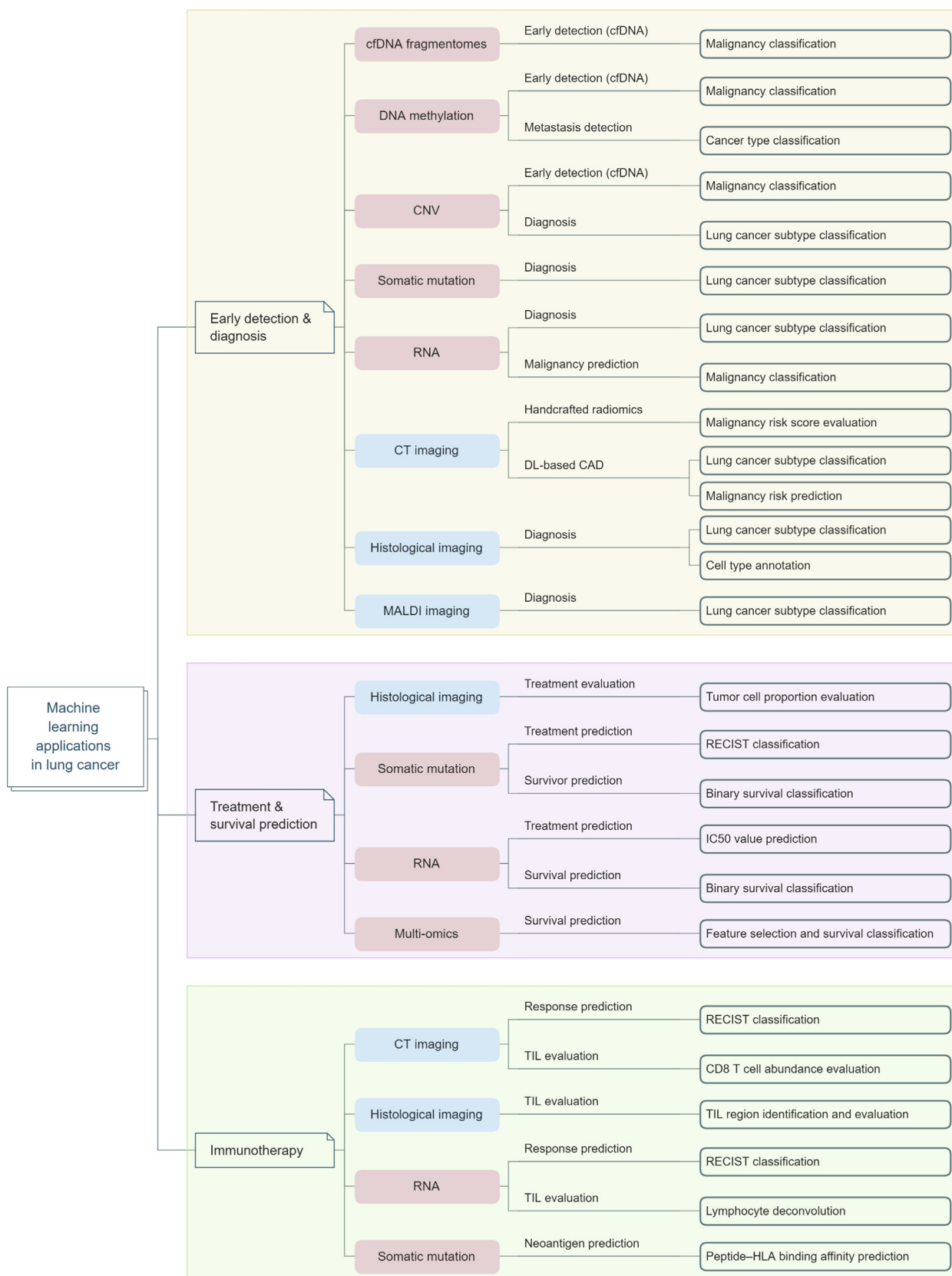
Diagnosis dini merupakan prosedur penting untuk mengurangi kematian terkait kanker paru-paru. Skrining dada menggunakan tomografi komputer dosis rendah (CT) adalah pendekatan utama untuk pengawasan orang-orang dengan peningkatan risiko kanker paru-paru. Untuk meningkatkan efisiensi diagnostik, sistem diagnosis berbantuan komputer (CAD) dikembangkan untuk membantu dokter dalam interpretasi data pencitraan medis[28,29], yang telah terbukti sebagai opini kedua yang berguna bagi para dokter[30]. Tugas CAD berbasis fitur tradisional dapat dibagi menjadi tiga langkah: segmentasi nodul, ekstraksi dan seleksi fitur, dan inferensi penilaian klinis (klasifikasi) (Gambar 2). Beberapa pendekatan menerapkan fitur tekstur terukur dari nodul tertentu pada gambar CT yang dikombinasikan dengan variabel klinis pasien sebagai fitur masukan untuk melatih pengklasifikasi ML, termasuk regresi logistik (LR)[31–33] atau analisis diskriminan linier (LDA)[34], untuk estimasi risiko keganasan. Biasanya, pengukuran ini mencakup ukuran nodul, jenis nodul, lokasi nodul, jumlah nodul, batas nodul, dan informasi emfisema pada gambar CT, dan variabel klinisnya meliputi usia pasien, jenis kelamin, waktu pengambilan spesimen, riwayat kanker paru-paru dalam keluarga, paparan merokok, dan banyak lagi. Namun, ciri-ciri ini sebagian besar bersifat subjektif dan ditentukan secara sewenang-wenang, dan biasanya gagal mencapai gambaran lengkap dan kuantitatif mengenai penampakan nodul ganas.

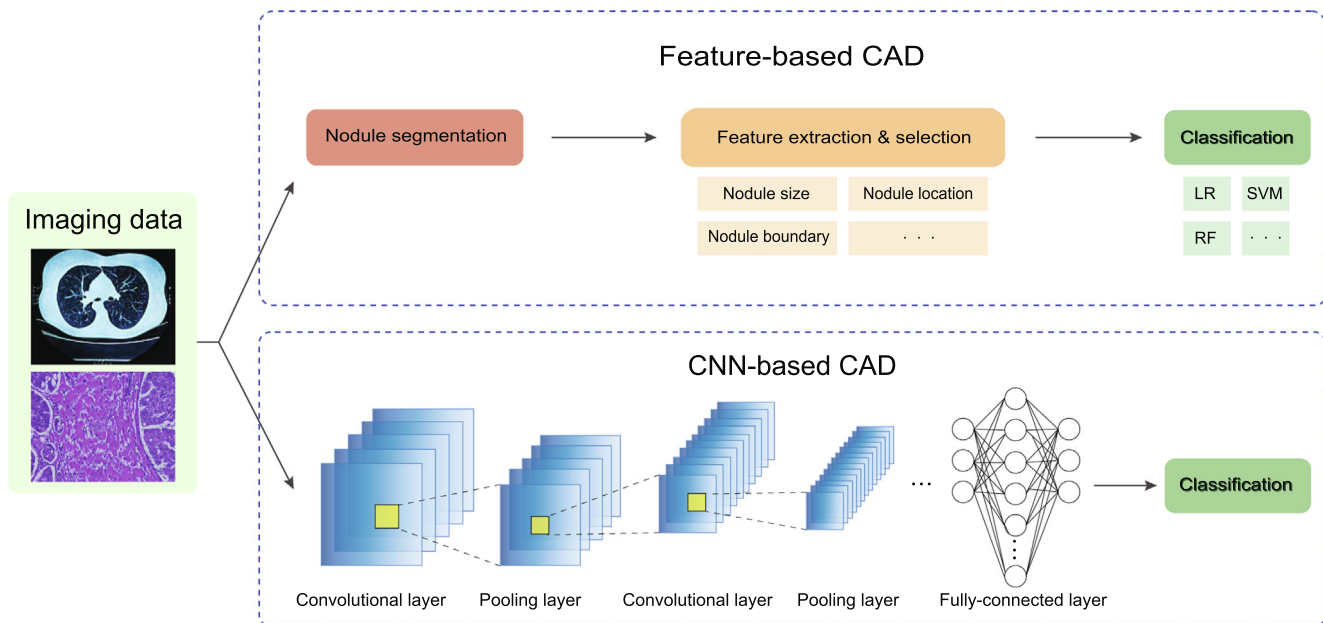
Dengan berkembangnya algoritma pembelajaran mendalam (DL), khususnya jaringan saraf konvolusional (CNN), lebih banyak penelitian telah dilakukan untuk menerapkan model berbasis DL dalam sistem CAD untuk meningkatkan akurasinya dan mengurangi tingkat positif palsu dan waktu eksekusi selama tumor paru-paru. deteksi (Tabel 1)[35,36]. Mirip dengan sistem CAD berbasis fitur, alur kerja model ini biasanya terdiri dari tiga langkah: deteksi dan segmentasi nodul, ekstraksi fitur nodul, dan inferensi penilaian klinis.[37]. Dibandingkan dengan tradisional

sistem CAD berbasis fitur, sistem CAD berbasis DL dapat secara otomatis mengambil dan mengekstrak fitur intrinsik dari nodul yang mencurigakan[38,39], dan dapat memodelkan bentuk 3D dari sebuah nodul (Gambar 2). Misalnya, Ciompi dkk.[40] merancang model berdasarkan OverFeat[41,42] dengan mengekstraksi tiga vektor fitur tampilan 2D (aksial, koronal, dan sagital) dari nodul dari CT scan. Model CNN yang baru-baru ini terintegrasi memfasilitasi pemeriksaan nodul secara global dan komprehensif untuk karakterisasi fitur dari gambar CT. Buty dkk.[37] merancang model CNN pelengkap, dimana model harmonik bola[43] untuk segmentasi nodul digunakan untuk mendapatkan deskripsi bentuk (fitur "bentuk") dari nodul tersegmentasi dan model berbasis jaringan saraf konvolusional dalam (DCNN)[41] untuk mengekstrak fitur tekstur dan intensitas (fitur "penampilan") bintil. Klasifikasi hilir mengandalkan kombinasi fitur "bentuk" dan "penampilan". Demikian pula, Venkadesh dkk.[44] menggunakan model ansambel dari dua model berbeda, berbasis 2D-ResNet50[45] dan 3D-Inception-V1[46], untuk masing-masing mengekstrak dua fitur nodul paru, lalu menggabungkan kedua fitur tersebut sebagai fitur masukan untuk klasifikasi. Keunggulan model ansambel CNN adalah model ini dapat secara akurat mengidentifikasi nodul ganas dari berbagai ukuran nodul menggunakan gambar CT mentah. Memanfaatkan fitur yang diambil dari model CNN yang canggih, inferensi penilaian klinis dapat diimplementasikan melalui teknik ML yang sering digunakan, termasuk LR, random forest (RF), support vector machine (SVM), dan neural network (NNs). Khususnya, beberapa penelitian juga menggunakan model CNN untuk kesimpulan penilaian klinis akhir. Ardila dkk.[47] mengusulkan pendekatan end-to-end untuk secara sistematis memodelkan tugas lokalisasi dan kategorisasi risiko kanker paru-paru hanya dengan menggunakan input data CT. Pendekatan mereka didasarkan pada kombinasi tiga model CNN: Mask-RCNN[48] model untuk segmentasi jaringan paru-paru, RetinaNet yang dimodifikasi[49] model untuk deteksi wilayah minat kanker (ROI), dan model volume penuh berdasarkan 3Dinflated Inception-V1[50,51] untuk prediksi risiko keganasan. Selain gambar CT, model berbasis CNN juga banyak digunakan dalam pencitraan histologis untuk membantu diagnosis kanker paru-paru. Dibandingkan dengan pencitraan CT, pencitraan histologis dapat memberikan lebih banyak informasi biologis tentang kanker pada tingkat sel. Untuk tujuan ini, Abduljabbar dkk.[52] menggunakan Micro-Net[53] model untuk mengidentifikasi batas jaringan diikuti oleh SC-CNN

[54] model untuk mensegmentasi sel individu dari gambar hematoxylin dan eosin (H&E) dan imunohistokimia (IHC). Sel yang tersegmentasi kemudian diterapkan untuk klasifikasi jenis sel guna mengevaluasi proporsi setiap jenis sel dalam gambar. Model ini membantu mengidentifikasi evolusi diferensial dan mekanisme penghindaran kekebalan antara adenokarsinoma paru (LUAD) dan karsinoma sel skuamosa paru (LUSC) dengan resolusi tinggi. Studi lain[55] memanfaatkan jaringan Inception-V3[51] untuk mengklasifikasikan apakah jaringan tersebut LUAD, LUSC, atau normal dari gambar seluruh slide histopatologi yang diwarnai dengan H&E. Hal yang menarik dari penelitian ini adalah bahwa model tersebut juga dapat memprediksi apakah jaringan tertentu memiliki mutasi somatik pada beberapa gen pemicu kanker paru-paru, termasuk STK11, EGFR, FAT1, SETBP1, KRAS, dan TP53. Perlu diperhatikan bahwa mengingat kompleksitas yang tinggi dan sumber daya yang besar dari kumpulan data, beberapa penelitian menggunakan pembelajaran transfer untuk meningkatkan efisiensi dan ketahanannya saat melatih model baru.[38,55].

Meskipun algoritma ML ini sudah banyak digunakan di CAD, tantangannya adalah jumlah gambarnya terbatas





Gambar 2 Sistem CAD berbasis fitur dan CAD berbasis DL

Perbedaan proses pengembangan sistem CAD berbasis fitur dan sistem CAD berbasis CNN. Dibandingkan dengan sistem CAD berbasis fitur, sistem CAD berbasis DL dapat secara otomatis mengambil dan mengekstrak fitur intrinsik dari nodul yang mencurigakan. CNN, jaringan saraf konvolusional; LR, regresi logistik; SVM, mendukung mesin vektor; RF, hutan acak.

diberi label. Melatih model CNN yang kompleks menggunakan set pelatihan dalam jumlah terbatas dapat mengakibatkan overfitting. Baru-baru ini, model berbasis generative adversarial network (GAN) telah digunakan untuk meningkatkan kinerja pengklasifikasi diskriminatif dengan menghasilkan gambar semu.[56]. Chuquicuma dkk.[57] pertama kali menggunakan GAN konvolusional mendalam (DCGAN)[58] model untuk menghasilkan CT scan nodul paru sintetis. Dengan hasil kerja mereka, penelitian terbaru telah mengintegrasikan model GAN dengan model CNN lainnya untuk mengatasi masalah overfitting dalam klasifikasi kanker paru-paru. Lin dkk.[59] menggunakan model dua langkah — DCGAN untuk menghasilkan gambar kanker paru-paru sintetis dan AlexNet[41] untuk klasifikasi kanker paru-paru menggunakan kumpulan data asli dan sintetis. Pekerjaan serupa juga dilakukan Ren dan rekannya[60]. Mereka juga menggunakan DCGAN[58] untuk augmentasi data. Untuk meningkatkan kinerja, mereka kemudian merancang model pembelajaran transfer yang ditingkatkan regularisasi yang disebut VGG-DF untuk diskriminasi data guna mencegah masalah overfitting dengan pemilihan otomatis model yang telah dilatih sebelumnya.

ML pada deteksi dini dan diagnosis menggunakan kumpulan data pengurutan -omics

Meskipun tes pencitraan medis berkala direkomendasikan untuk populasi berisiko tinggi, penerapannya masih rumit

dengan tingkat penemuan palsu yang tinggi[61,62]. Oleh karena itu, terdapat kebutuhan mendesak akan teknik baru dalam deteksi dini kanker paru-paru. Teknologi pengurutan terkini memungkinkan beragam metode untuk deteksi dini kanker paru-paru[63]. Sementara itu, pengklasifikasi subtipe kanker paru secara akurat sangat penting dalam memandu pengambilan keputusan terapi yang optimal. LUAD (~45%) dan LUSC (~25%) adalah dua subtipe kanker paru-paru yang paling umum namun sering kali ditangani dengan cara yang sama kecuali untuk terapi yang ditargetkan [64]. Namun, penelitian menunjukkan bahwa LUAD dan LUSC memiliki tanda biologis yang sangat berbeda, dan mereka menyarankan agar LUAD dan LUSC harus diklasifikasikan dan diperlakukan sebagai kanker yang berbeda.[65,66]. Dari perspektif komputasi, deteksi dini dan identifikasi subtipe adalah bagian dari tugas klasifikasi. Studi ML sebelumnya telah menunjukkan efisiensi dan kemajuan deteksi dini dan klasifikasi jenis kanker dalam kumpulan data pengurutan pan-kanker yang besar[67–75], yang dapat memberikan bukti untuk diagnosis kanker paru-paru. Diketahui bahwa sel-sel kanker dicirikan oleh banyak variasi genetik, dan akumulasi variasi genetik ini dapat menjadi tanda yang mendokumentasikan pola mutasi berbagai jenis kanker.[3,5,76,77]. Oleh karena itu, penelitian terbaru berkonsentrasi pada ekstraksi tanda tangan genom yang lebih baik sebagai fitur masukan untuk meningkatkan akurasi model ML mereka. Untuk deteksi dini, biopsi cair berbasis darah, termasuk

3

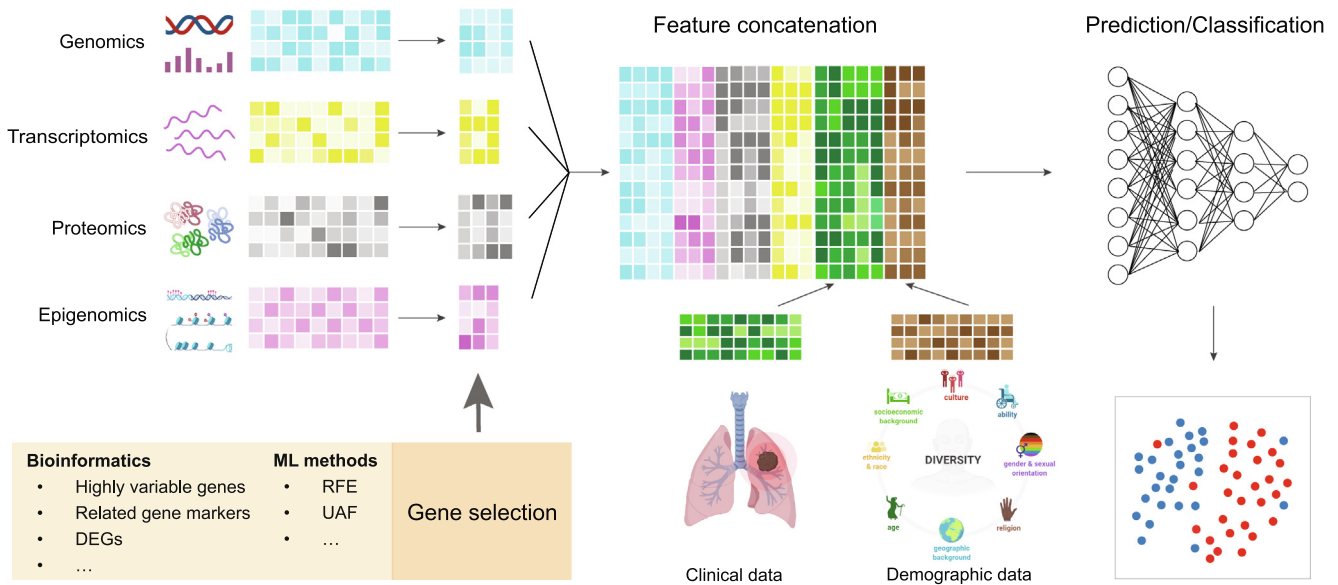
Gambar 1 Penerapan model ML pada kanker paru-paru

Kami menyajikan ikhtisar metodologi ML untuk berbagai aspek terapi kanker paru-paru, termasuk CAD dari kumpulan data pencitraan, deteksi dini kanker paru-paru berdasarkan teknologi pengurutan, integrasi data dan ekstraksi biomarker dari kumpulan data multi-omics, respons pengobatan dan prediksi prognosis, serta studi imunoterapi. . ML, pembelajaran mesin; IC50, konsentrasi penghambatan setengah maksimal; HLA, antigen leukosit manusia; CT, tomografi komputer; MALDI, desorpsi/ionisasi laser berbantuan matriks; DL, pembelajaran mendalam; cfDNA, DNA bebas sel; CAD, diagnosis dengan bantuan komputer; CNV, variasi nomor salinan; RECIST, Kriteria Evaluasi Respon pada Tumor Padat; TIL, limfosit yang menginfiltrasi tumor.

Tabel 1 Publikasi yang relevan dengan ML tentang deteksi dini dan diagnosis menggunakan data pencitraan

Publikasi	Ekstraksi fitur	Model klasifikasi	Ukuran sampel	Tipe data pencitraan	Pertunjukan	Metode validasi	Pemilihan/masukan fitur	Menyorot/Akeuntungan	Kekurangan
McWilliams dkk.[31]	TIDAK	LR	2961	gambar CT	AUC (0,907–0,960)	Tahan	Faktor risiko klinis + karakteristik nodul pada gambar CT	Dengan menggunakan fitur yang diekstraksi sebagai masukan, pengklasifikasi dapat mencapai AUC tinggi dalam nodul kecil (<10 mm)	Pemilihan karakteristik nodul mempengaruhi kinerja prediktif model
Riel dkk.[32]	TIDAK	LR	300	gambar CT	AUC (0,706–0,932)	Tahan	Faktor klinis + karakteristik nodul pada gambar CT	Pengklasifikasi dapat bekerja setara dengan pengamat manusia untuk klasifikasi ganas dan jinak	Kinerjanya sangat bergantung pada ukuran nodul diskriminator, dan tidak kuat pada bintil kecil
Kriegsmann dkk.[34]	TIDAK	LDA	326	MALDI	Akurasi (0,991)	Tahan	Spektrum massa dari ROI gambar MALDI	Model ini mempertahankan akurasi tinggi pada biopsi FFPE	Performanya mengandankan kualitas MALDI stratifikasi
Buty dkk.[37]	Harmonisa bola[44]; DCNN[41]	Federasi Rusia	1018	gambar CT	Akurasi (0,793–0,824)	Validasi silang 10 kali lipat	Patch pencitraan CT + segmentasi nodul biner ahli radiologi	Model ini mencapai akurasi prediksi yang lebih tinggi dengan mengintegrasikan fitur pencitraan nodul bentuk dan tampilan. Model ini mencapai akurasi yang lebih tinggi dibandingkan model tolak ukur lainnya	Tidak ada perbandingan pembandingan yang digunakan dalam penelitian ini
Husein dkk.[38]	Model multi-tugas berbasis CNN 3D	Model multi-tugas berbasis CNN 3D	1018	gambar CT	Akurasi (0,9126)	Validasi silang 10 kali lipat	Fitur volume CT 3D		Skor kebenaran dasar yang ditentukan oleh ahli radiologi untuk tolak ukurnya mungkin sewenang-wenang. Segmentasi mungkin gagal jika ROI berada di luar wilayah paru-paru
Khosravan dkk.[39]	Model multi-tugas berbasis CNN 3D	Model multi-tugas berbasis CNN 3D	6960	gambar CT	Segmentasi DSC (0,91); akurasi klasifikasi (0,97)	Validasi silang 10 kali lipat	Fitur volume CT 3D	Integrasi model algoritma pengelompokan dan sparsifikasi membantu mengekstraksi wilayah perhatian potensial secara akurat	
Ciampi dkk.[40]	Prestasi Berlebih[42]	100% Federasi Rusia	1729	gambar CT	AUC (0,868)	Validasi silang 10 kali lipat	Fitur volume CT 3D, koordinat posisi nodul, dan diameter maksimum	Ini adalah penelitian pertama yang mencoba mengklasifikasikan apakah nodul yang didiagnosis bersifat jinak atau ganas	Model ini memerlukan penentuan posisi dan diameter nodul sebagai masukan, namun banyak nodul yang tidak dapat ditemukan pada gambar CT.
Venkadesh dkk.[44]	Berbasis 2D-ResNet50[45]; 3D-Permulaan-V1[46]	Model ansambel berdasarkan dua model CNN	16.429	gambar CT	AUC (0,86–0,96)	Validasi silang 10 kali lipat	Fitur volume CT 3D dan koordinat nodul	Model ini mencapai AUC yang lebih tinggi dibandingkan model benchmark lainnya	Model ini memerlukan penentuan posisi nodul, namun banyak nodul yang tidak dapat ditemukan pada gambar CT. Kelompok pelatihan hanya berasal dari satu kumpulan data, meskipun ukuran sampelnya besar
Ardila dkk.[47]	Topeng-RCNN[48]; RetinaNet[49]; Inception-V1 yang digelembungkan 3D[50,51]	Topeng-RCNN[48]; RetinaNet[49]; meningkatkan 3D Awal-V1[50,51]	14.851	gambar CT	AUC (0,944)	Tahan	Fitur volume CT 3D pasien saat ini dan sebelumnya (jika tersedia).	Model ini mencapai AUC yang lebih tinggi dibandingkan ahli radiologi ketika sampel tidak memiliki gambar CT sebelumnya	
Abduljabbar dkk.[52]	Mikro-Net[53]; SC-CNN[54]	Model ansambel berdasarkan SC-CNN[54]	100	Gambar histologis	Akurasi (0,913)	Tahan	Fitur gambar slide histologis bagian tumor yang diwarnai H&E	Model ini dapat memberi anotasi pada tipe sel pada tingkat sel tunggal hanya dengan menggunakan gambar histologis	Akurasi anotasi dipengaruhi oleh kumpulan data referensi yang digunakan
Coudray dkk.[55]	Model CNN multi-tugas berdasarkan Inception-V3[51]	Model CNN multi-tugas berdasarkan Inception-V3 jaringan[51]	1634	Gambar histologis	AUC (0,733–0,856)	Tahan	Mengubah ubin 512 - 512 piksel dari 'tambalan' gambar keseluruhan slide yang tidak tumpang tindih	Model tersebut dapat memprediksi apakah jaringan tertentu memiliki mutasi gen somatik STK11, EGFR, FAT1, SETBP1, KRAS, DanTP53	Keakuratan prediksi mutasi gen tidak terlalu tinggi
Lin dkk.[59]	DCGAN[58]+ AlexNet[41]	DCGAN[58]+ AlexNet[41]	22.489	gambar CT	Akurasi (0,9986)	Tahan	Gambar CT awal + sintetik	Model ini menggunakan GAN untuk menghasilkan gambar kanker paru-paru sintetis guna mengurangi overfitting	Tidak ada perbandingan pembandingan yang digunakan
Ren dkk.[60]	DCGAN[58]+ VGG-DF	DCGAN[58]+ VGG-DF	15.000	Gambar histopatologi	Akurasi (0,9984); Skor F1 (99,84%)	Tahan	Gambaran histopatologi awal + sintetik	Model ini menggunakan GAN untuk menghasilkan gambar kanker paru-paru sintetis dan model yang disempurnakan dengan regularisasi untuk mengurangi overfitting	Dimensi gambar berdasarkan generator (64 - 64) tidak cukup untuk domain biomedis

Catatan:ML, pembelajaran mesin; TIDAK, tidak berlaku; LR, regresi logistik; AUC, area di bawah kurva; CT, tomografi komputer; LDA, analisis diskriminan linier; MALDI, desorpsi/ionisasi laser berbantuan matriksi; ROI, wilayah yang diminati; FFPE, tertanam parafin dengan formalin; CNN, jaringan saraf konvolusional; DSC, koefisien kesamaan dadu; SVM, mendukung mesin vektor; RF, hutan acak; DCNN, jaringan saraf konvolusional dalam; SC-CNN, jaringan saraf konvolusional yang dibatasi secara spasial; DCGAN, jaringan permusuhan generatif konvolusional yang mendalam; RCNN, Wilayah-CNN; H&E, hematoksilin dan eosin; 2D, dua dimensi; 3D, tiga dimensi. Dibandingkan dengan hold-out, validasi silang biasanya lebih kuat, dan memperhitungkan lebih banyak varian antara kemungkinan pemisahan data pelatihan, validasi, dan pengujian. Namun, validasi silang lebih memakan waktu dibandingkan menggunakan metode ketidaksepakatan sederhana.



Gambar 3 Analisis Omics dalam studi kanker paru-paru

Teknik pengurutan yang berbeda memungkinkan pengukuran beberapa fitur molekuler dari sampel biologis secara bersamaan. Untuk meningkatkan efisiensi dan mengurangi overfitting, alat statistik dan ML melakukan analisis diferensial atau pemilihan fitur. Model ML selanjutnya menggabungkan fitur omics yang diperoleh dengan fitur klinis sebagai masukan untuk prediksi diagnostik/prognostik kanker paru-paru. DEG, gen yang diekspresikan secara berbeda; RFE, penghapusan fitur rekursif; UAF, pemfilteran asosiasi univariat.

Fragmen DNA bebas sel (cfDNA), DNA tumor yang bersirkulasi (ctDNA), mikroRNA (miRNA), metilasi, eksosom, dan sel tumor yang bersirkulasi (CTC), untuk mengeksplorasi potensi tanda tangan tumor yang bersirkulasi dianggap sebagai metode yang dapat diandalkan[63](Gambar 3). Dengan mengintegrasikan tanda-tanda biopsi cair ini, banyak model diskriminatif (SVM, RF, dan LR) telah digunakan untuk mendeteksi tumor dengan tingkat penemuan yang tinggi.[78–81]. Untuk klasifikasi subtipe kanker paru-paru, mutasi somatik, termasuk varian nukleotida tunggal (SNV), penyisipan, dan penghapusan, biasanya memiliki profil jenis kanker tertentu.[82]. Dengan demikian, penelitian telah memanfaatkan mutasi somatik sebagai fitur masukan untuk melatih pengklasifikasi untuk klasifikasi LUAD-LUSC[83]. Banyak dari mutasi ini, terutama mutasi driver, dapat mengubah tingkat ekspresi, yang berdampak pada fungsi gen dan mengganggu proses sinyal seluler[82]. Akibatnya, jenis kanker yang berbeda menunjukkan tingkat ekspresi protein tertentu yang berbeda[84,85]. Dipengaruhi oleh profil ekspresi unik jenis kanker ini, model ML dapat memanfaatkan pengurutan RNA sebagai data masukan untuk mengkategorikan keganasan (jinak atau ganas) dan subtipe (LUAD atau LUSC) pasien[86–89]. Demikian pula, variasi jumlah salinan (CNV) dilaporkan sangat berkorelasi dengan ekspresi gen diferensial[90], dan dapat dideteksi di mana-mana pada sel kanker. Dengan demikian, CNV juga dapat digunakan untuk melatih model ML untuk klasifikasi jenis kanker dalam studi kanker paru-paru [81,91,92]. Perhatikan bahwa Daemen dkk.[92] mengusulkan model Markov tersembunyi berulang (HMM) untuk mengidentifikasi wilayah kromosom yang diperluas dari nomor salinan yang diubah, yang menawarkan akurasi tinggi untuk klasifikasi. Baru-baru ini, Jurmeister dkk.[93] menggunakan profil metilasi DNA sebagai fitur masukan untuk menentukan apakah nodul ganas yang terdeteksi adalah kanker paru primer atau metastasis dari kanker lain. Menggunakan semua gen yang dihasilkan secara langsung sebagai fitur masukan dapat mengakibatkan overfitting[94]. Jadi, banyak

penelitian menggunakan pendekatan komputasi yang berbeda untuk memilih beberapa gen terkait kanker untuk meningkatkan model ML mereka (Gambar 3). Beberapa penelitian menggunakan algoritma berbasis ML untuk pemilihan fitur. Misalnya, Liang dkk.[80] dan Whitney dkk.[86] menggunakan metode operator penyusutan dan seleksi absolut terkecil (LASSO) untuk memilih penanda optimal untuk pelatihan model; Aliferis dkk.[89] memanfaatkan penghapusan fitur rekursif (RFE)[95] dan model penyaringan asosiasi univariat (UAF) untuk memilih gen yang sangat terkait dengan kanker. Selain itu, penggunaan model tanpa pengawasan untuk pengelompokan subtipe populasi sampel, dan kemudian mengidentifikasi gen penanda setiap cluster juga terlihat dalam banyak penelitian.[96,97]. Selain model berbasis ML, beberapa penelitian menggunakan metode statistik untuk pemilihan fitur. Raman dkk.[81] merancang skor ketidaknormalan profil nomor salinan (CPA) untuk memperkuat fitur CNV yang lebih kuat dan tidak terlalu bergantung pada kualitas sampel variabel dibandingkan secara langsung menggunakan CNV sebagai fitur masukan. Daemen dkk.[92] mengintegrasikan beberapa uji statistik (perubahan lipatan biasa, biasa T-statistik, statistik SAM, dan dimoderasi T-statistik) untuk memilih kumpulan gen ekspresi diferensial yang kuat. Selain tanda tangan terukur tunggal ini, beberapa penelitian[81,86,88] menggabungkan tanda tangan -omics dengan tanda tangan klinis untuk mencapai hasil yang lebih baik. Dengan menggunakan tanda tangan -omics spesifik tipe tumor ini, banyak algoritma, K-nearest neighbours (KNN), naïf Bayes (NB), SVM, pohon keputusan (DT), LR, RF, LDA, peningkatan gradien, dan NN, telah mendemonstrasikannya kemampuan untuk secara akurat mendeteksi dan mengklasifikasikan pola kanker paru-paru yang berbeda (Meja 2). Perhatikan bahwa untuk meningkatkan akurasi model ML, Kobayashi dkk.[83] menambahkan penskalaan masukan berdasarkan elemen untuk model NN, yang memungkinkan model mempertahankan keakuratannya dengan sejumlah kecil parameter yang dapat dipelajari untuk pengoptimalan.

Meja 2 Publikasi yang relevan dengan ML tentang deteksi dini dan diagnosis menggunakan data pengurutan

Publikasi	metode ML	Ukuran sampel	Tipe data pengurutan	Pertunjukan	Metode validasi	Pemilihan fitur	Menyorot/Akeuntungan	Kekurangan
Mathios dkk.[78]	Model LR dengan penalti LASSO	799	fragmen cfDNA	AUC (0,98)	Validasi silang 10 kali lipat	fitur fragmen cfDNA, faktor risiko klinis, dan fitur pencitraan CT	Studi ini memberikan kerangka kerja untuk menggabungkan profil fragmentasi cfDNA dengan penanda lain untuk deteksi kanker paru-paru	Variasi DNA pada penyakit stadium akhir dapat mempengaruhi deteksi cfDNA
Paru-CLIP[79]	5-tetangga terdekat; 3-tetangga terdekat; Catatan; LR; DT	160	cfDNA	AUC (0,69–0,98)	Validasi silang tanpa satu kali keluar	Fitur SNV + CNV	Studi ini menetapkan kerangka ML untuk deteksi dini kanker paru-paru menggunakan cfDNA	Ada bias pengambilan sampel (sebagian besar adalah perokok) dalam kumpulan data pelatihan
Liang dkk.[80]	LR	296	ctDNA	AUC (0,816)	Validasi silang 10 kali lipat	Sembilan penanda metilasi DNA	Studi ini menetapkan kerangka kerja ML untuk deteksi dini kanker paru-paru menggunakan biomarker metilasi DNA sembilan, yang merupakan penanda metilasi	Fitur yang dipilih hanya terdiri dari keterbatasan kinerja pengujian Metode pemilihan fitur dapat digunakan untuk mengurangi overfitting dan mungkin berpotensi mencapai AUC yang lebih tinggi
Raman dkk.[81]	Federasi Rusia; SVM; LR dengan punggung, jaring elastis; Regularisasi LASSO	843	cfDNA	mAUC (0,896–0,936)	Validasi silang tinggalkan satu kali. Salin profil nomor cfDNA	Salin profil nomor cfDNA	Model ini memberikan kerangka kerja untuk menggunakan profil nomor salinan cfDNA sebagai biomarker dalam deteksi kanker paru-paru	EIS membantu menstabilkan proses pelatihan Diet Networks
Kobayashi dkk.[83]	Jaringan Diet dengan EIS	954	Mutasi somatik	Akurasi (0,8)	Validasi silang 5 kali lipat	SNV, penyisipan, dan penghapusan pada 1796 gen	Interpretasi tersembunyi yang dapat diinterpretasikan yang diperoleh dari EIS mungkin berbeda-beda di antara kumpulan data yang berbeda	Gen yang dipilih sangat bervariasi berdasarkan proses dan parameter pemilihan fitur yang berbeda. Metode pemilihan fitur dapat digunakan untuk mengurangi overfitting
Whitney dkk.[86]	LR	299	RNA-seq dari KBG	AUC (0,81)	Validasi silang 10 kali lipat	Penanda RNA kovariat klinis dan terkait kanker paru RNA-seq	Model ini menjaga sensitivitas untuk dugaan lesi kecil dan perifer	Gen yang dipilih sangat bervariasi antar kelompok pelatihan yang berbeda
Podolsky dkk.[87]	KNN; NB distribusi atribut normal; Distribusi NB melalui histogram; SVM; C4.5 DT	529	RNA-seq	AUC (0,91)	Tahan		Studi ini secara sistematis membandingkan berbagai model klasifikasi subtype kanker paru-paru di berbagai kumpulan data	Ukuran sampel dalam subkelompok tertentu kecil dan dapat menyebabkan pelatihan tidak seimbang
Choi dkk.[88]	Model ansambel berdasarkan jaring elastis LR; SVM; LR hierarkis	2285	RNA-seq bronkial sampel menyikat	AUC (0,74)	Validasi silang 5 kali lipat	RNA-seq dari 1232 gen dengan kovariat klinis	Model ini mengintegrasikan fitur RNA-seq dan informasi klinis untuk meningkatkan akurasi prediksi risiko	
Aliferis dkk.[89]	SVM linier; SVM kernel polinomial; KNN; NN 203		RNA-seq	AUC (0,8783–0,9980)	Validasi silang 5 kali lipat	RNA-seq dari gen yang dipilih menggunakan RFE dan UAF	Penelitian ini menggunakan algoritma seleksi gen yang berbeda untuk meningkatkan akurasi klasifikasi	Gen yang dipilih sangat bervariasi antar kelompok pelatihan yang berbeda
Aliferis dkk.[91]	DT; KNN; SVM linier; SVM kernel polinomial; SVM kernel RBF; NN	37	CNV diukur dengan Akurasi CGH (0,892)	Validasi silang satu kali keluar	Salin jumlah 80 gen yang dipilih berdasarkan SVM linier		Studi ini secara sistematis membandingkan berbagai model klasifikasi subtype kanker paru-paru. Penggunaan HMM berulang untuk deteksi CNV memberikan akurasi tinggi untuk klasifikasi kanker	Ukuran sampelnya kecil
Daemen dkk.[92]	HMM; LS-SVM tertimbang	89	CNV diukur dengan Akurasi CGH (0,880–0,955)	Validasi silang 10 kali lipat	CNV diukur dengan CGH		Studi ini memberikan kerangka kerja untuk menggunakan data metilasi DNA untuk memprediksi metastasis tumor	Perbandingan yang dijadikan patokan diperlukan untuk menunjukkan keunggulan penggunaan model HMM
Jurmeister dkk.[93]	tidak; SVM; Federasi Rusia	972	Metilasi DNA	Akurasi (0,878–0,964)	Validasi silang 5 kali lipat	2000 situs CpG variabel teratas		Model ini tidak dapat secara akurat memprediksi sampel dengan selularitas tumor yang rendah data metilasi

Catatan:LASSO, operator penyusutan dan seleksi absolut terkecil; cfDNA, DNA bebas sel; NB, Bayes yang naif; DT, pohon keputusan; SNV, varian nukleotida tunggal; CNV, variasi nomor salinan; ctDNA, DNA tumor yang bersirkulasi; mAUC, rata-rata area di bawah kurva; EIS, penskalaan masukan berdasarkan elemen; BEC, sel epitel bronkus; KNN, K-tetangga terdekat; NN, jaringan saraf; RFE, penghapusan fitur rekursif; UAF, pemfilteran asosiasi univariat; CGH, hibridisasi genom komparatif; HMM, model Markov tersembunyi; LS-SVM, kuadrat terkecil mendukung mesin vektor; RNA-seq, pengurutan RNA. Dibandingkan dengan hold-out, validasi silang biasanya lebih kuat, dan memperhitungkan lebih banyak varian antara kemungkinan pemisahan data pelatihan, validasi, dan pengujian. Namun, validasi silang lebih memakan waktu dibandingkan menggunakan metode ketidaksepakatan sederhana.

Menerapkan ML pada respons pengobatan kanker paru-paru dan prediksi kelangsungan hidup

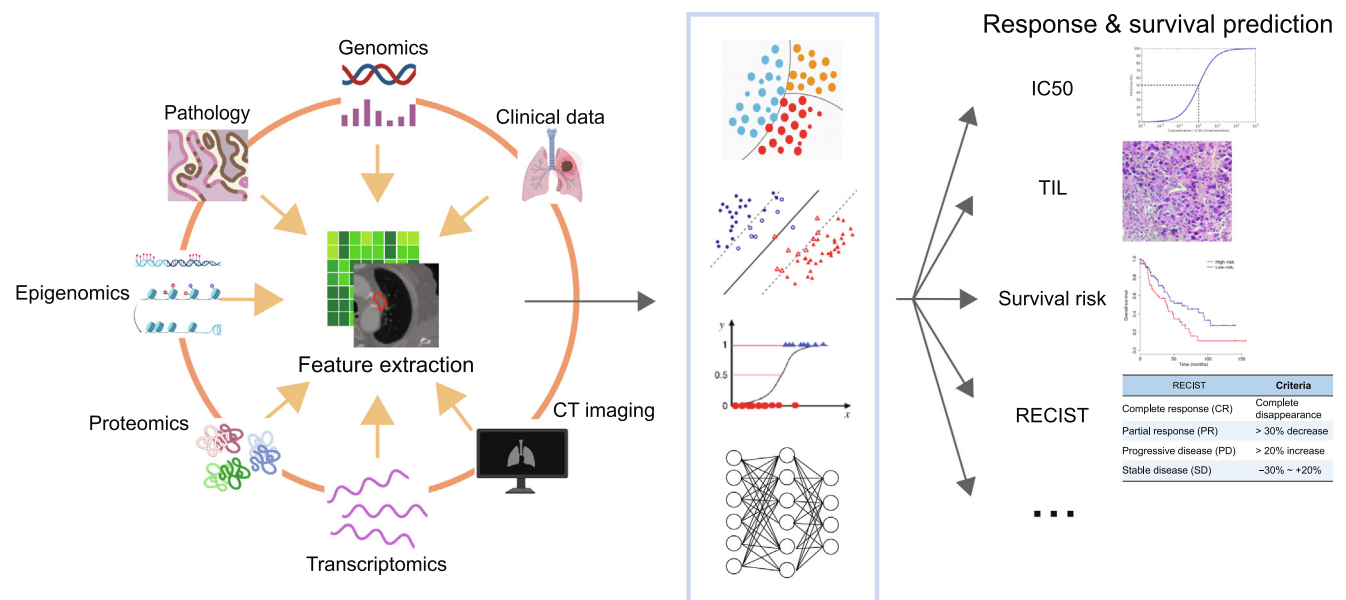
Prognosis dan prediksi respon terapi

Model ML yang canggih telah bertindak sebagai pelengkap untuk evaluasi dan prediksi respons intervensi kanker[98,99], dan telah menunjukkan kemajuan dalam mengoptimalkan keputusan terapi yang meningkatkan peluang keberhasilan pemulihan (Gambar 4; Tabel 3)[100,101]. Ada beberapa metrik yang tersedia untuk mengevaluasi respon terapi kanker, termasuk Response Evaluation Criteria in Solid Tumors (RECIST)[102]. Definisi RECIST bergantung pada data pencitraan, terutama CT dan magnetic resonance imaging (MRI), untuk menentukan bagaimana tumor tumbuh atau menyusut pada pasien.[103]. Untuk melacak perubahan volume tumor dari gambar CT, Jiang et al.[104] merancang model CNN terintegrasi. Model CNN mereka menggunakan dua jaringan dalam berdasarkan jaringan sisa resolusi penuh[105] membuat model dengan menambahkan beberapa aliran sisa dengan resolusi berbeda-beda, sehingga keduanya dapat secara bersamaan menggabungkan fitur pada resolusi berbeda untuk mensegmentasi tumor paru-paru. Menggunakan kriteria RECIST, Qureshi[106] menyiapkan model RF untuk memprediksi tingkat RECIST di bawah terapi EGFR tyrosine kinase inhibitor (TKI) dengan mempertimbangkan profil mutasi gen pasien EGFR. Untuk meningkatkan kinerja prediksi, model ini mengintegrasikan informasi klinis, fitur geometris, dan fitur energi yang diperoleh dari pasien EGFR kompleks obat mutan sebagai masukan untuk melatih pengklasifikasi. Dalam penelitian terbaru, penulis mendefinisikan metrik yang berbeda, skor proporsional tumor (TPS) yang dihitung sebagai persentase sel tumor dalam gambar patologi digital, untuk mengevaluasi respons pengobatan kanker paru-paru.[107]. Mereka menerapkan ambang batas Otsu[108] dengan model jaringan adversarial generatif pengklasifikasi tambahan (AC-GAN) untuk mengidentifikasi daerah sel tumor positif (TC+) dan daerah sel tumor negatif (TC-). Dan mereka akhirnya menggunakan rasio antara jumlah piksel TC+ wilayah dan jumlah piksel dari semua wilayah sel tumor yang terdeteksi untuk mengevaluasi jumlah TPS.

Studi lain dari Geeleher et al.[109] menggunakan konsentrasi penghambatan setengah maksimal (IC50) untuk mengevaluasi respon obat. Dalam modelnya, penulis menerapkan model regresi ridge [110] untuk memperkirakan nilai IC50 untuk garis sel yang berbeda dalam hal tingkat ekspresi seluruh genomnya. Baru-baru ini, Quiros dkk.[111] menetapkan pembelajaran representasi fenotip (PRL) melalui pembelajaran mandiri dan deteksi komunitas untuk anotasi tipe sel pengelompokan spasial pada gambar histopatologis. Hasil pengelompokannya selanjutnya dapat digunakan untuk melacak pola pertumbuhan tumor secara histologis dan mengidentifikasi kekambuhan tumor. Memang benar, model mereka juga menunjukkan kinerja yang baik dalam klasifikasi LUAD dan LUSC.

Prediksi kelangsungan hidup

Prognosis dan prediksi kelangsungan hidup sebagai bagian dari onkologi klinis merupakan tugas yang berat namun penting bagi dokter, karena mengetahui masa kelangsungan hidup dapat memberikan informasi dalam pengambilan keputusan pengobatan dan bermanfaat bagi pasien dalam mengelola biaya.[112–114]. Untuk sebagian besar riwayat kesehatan, prediksi terutama bergantung pada pengetahuan dan pengalaman dokter berdasarkan riwayat pasien dan catatan medis sebelumnya. Namun, penelitian menunjukkan bahwa dokter cenderung melakukan tindakan yang buruk dalam memprediksi prognosis dan harapan kelangsungan hidup, sering kali memprediksi waktu kelangsungan hidup secara berlebihan. [115–117]. Algoritme statistik, seperti model bahaya proporsional Cox[118], telah diterapkan untuk membantu prediksi dokter dalam banyak penelitian[119–122], tapi mereka tidak terlalu akurat[12]. Sebagai perbandingan, ML telah menunjukkan potensinya untuk memprediksi prognosis dan kelangsungan hidup pasien dalam kumpulan data genomik, transkriptomik, proteomik, radiomik, dan lainnya (Gambar 4; Tabel 3). Chen dkk.[123] menggunakan kelangsungan hidup 3 tahun sebagai ambang batas untuk membagi pasien menjadi kelompok risiko tinggi (waktu bertahan hidup <36 bulan) dan risiko rendah (waktu bertahan hidup > 36 bulan), dan kemudian membuat model NN untuk memprediksi secara biner risiko pasien menggunakan obatnya, data ekspresi gen dan variabel klinis. Dalam model mereka, mereka menguji empat



Gambar 4 Diagram penerapan ML dalam respons pengobatan dan prediksi kelangsungan hidup

Tabel 3 Publikasi yang relevan dengan ML tentang respons pengobatan dan prediksi kelangsungan hidup

Publikasi	Metode ekstraksi fitur	Model prediksi	Ukuran sampel	Tipe data	Pertunjukan	Metode validasi	Pemilihan fitur/Sayamasukan	Menyorot/Akeuntungan	Kekurangan
Jiang dkk.[104]		Model berbasis MRRN	1210	Gambar CT	DSC (0,68–0,75)	Validasi silang 5 kali lipat	Fitur gambar 3D	Model ini dapat secara akurat melacak perubahan volume tumor dari gambar CT di berbagai resolusi gambar	Model tersebut tidak memprediksi dengan cukup akurat ketika ukuran tumornya kecil
Qureshi[106]	TIDAK	Federasi Rusia; SVM; KNN; LDA; KERANJANG	201	Struktur molekul dan mutasi somatik EGFR	Akurasi (0,975)	Validasi silang 10 kali lipat	4 gambaran klinis + 4 gambaran interaksi obat protein + 5 gambaran geometris	Model ini mengintegrasikan beberapa fitur untuk pelatihan data, dan mencapai performa yang lebih baik dibandingkan model tolak ukur lainnya	Di antara kemungkinan 594 mutasi EGFR yang tersedia di database COSMIC, model ini hanya mempertimbangkan 33 mutasi EGFR yang paling umum untuk pelatihan model.
Kapil dkk.[107]	AC-GAN	AC-GAN	270	Gambar patologi digital	Lcc (0,94); Pcc (0,95); MAE (8,03)	Tahan	Slide histologis bagian tumor yang diwarnai PD-L1	Model ini sepenuhnya mencapai performa yang lebih baik dibandingkan model benchmark lainnya	Dalam percobaan, penggunaan pewarnaan PD-L1 untuk evaluasi TPS mungkin tidak cukup akurat
Geeleher dkk.[109]	TIDAK	Model regresi punggung	62	RNA-seq	Akurasi (0,89)	Tanda silang tinggalkan satu keluar validasi	Menghapus gen variabel rendah	Model ini dapat secara akurat memprediksi respons obat hanya dengan menggunakan profil RNA-seq	Ukuran sampel pelatihannya kecil
Chen dkk.[123]	Uji chi-kuadrat + NN	NN	440	RNA-seq	Akurasi (0,83)	Tahan	RNA-seq dari 5 gen	Model ini menggunakan beberapa kumpulan data laboratorium untuk pelatihan guna meningkatkan ketahanannya	Model ini tidak mempertimbangkan fitur demografi dan klinis, yang mungkin memengaruhi prediksi.
LUADpp[125]	Gen teras dengan mutasi paling signifikan perbedaan frekuensi	SVM	371	Mutasi somatik	Akurasi (0,81)	Validasi silang 5 kali lipat	Fitur mutasi somatik pada 85 gen	Frekuensi mutasi mungkin dipengaruhi oleh bias pengambilan sampel di seluruh kumpulan data; LD juga dapat mempengaruhi pemilihan fitur.	
Cho dkk.[126]	Perolehan informasi; Chl-tes kuadrat; minimum redundansi maksimum relevansi; korelasi algoritma	Catatan; KNN; SVM; DT	471	Mutasi somatik	Akurasi (0,68–0,88)	Validasi silang 5 kali lipat	Ciri mutasi somatik terdiri dari 19 gen	Model tersebut dapat memprediksi dengan akurasi tinggi hanya dengan tujuh fitur mutasi gen	Kelompok pelatihan hanya terdiri dari satu kumpulan data
Yu dkk.[128]	Rasio perolehan informasi; pengelompokan hierarki	Federasi Rusia	538	Multi-omics (histologi, laporan patologi, RNA, proteomik)	AUC (> 0,8)	tinggalkan satu keluar lintas validasi	15 fitur kumpulan gen	Untuk meningkatkan performa, model ini menggunakan empat metode berbeda untuk pemilihan fitur	
Asada dkk.[130]	Pembuat enkode otomatis + Cox-PH + K-means + ANOVA	SVM	364	Multi-omics (miRNA, mRNA)	Akurasi (0,81)	Tahan	20 miRNA + 25 mRNA	Penelitian ini menggunakan model omicspathology integratif untuk meningkatkan akurasi dalam memprediksi prognosis pasien	Model Cox mungkin dipasang secara berlebihan pada data multidimensi
Takahashi dkk.[131]	Pembuat enkode otomatis + Cox-PH + K-means + XGBoost/LightGBM	LR	483	Multi-omics (mRNA, mutasi somatik, CNV, mitos, RPPA)	AUC (0,43–0,99 di bawah data omics yang berbeda)	Tahan	12 mRNA, 3 miRNA, 3 metilasi, 5 CNV, 3 mutasi somatik, dan 3 RPPA	Studi ini menggunakan algoritma ML untuk memodelkan ekstraksi fitur secara sistematis dari kumpulan data multi-omics	Model ini tidak mempertimbangkan dampak varian klinis dan demografi dalam pelatihan data
Wiesweg dkk.[136]	Regresi laso	SVM	122	RNA-seq	Perbedaan rasio bahaya yang signifikan	Tahan	7 gen dari model seleksi fitur + 25 gen spesifik tipe sel	Studi ini menggunakan algoritma ML untuk memodelkan ekstraksi fitur secara sistematis dari kumpulan data multi-omics	Kumpulan data yang dikumpulkan dalam penelitian ini berisi sampel yang tidak umum di antara kumpulan data omics yang berbeda, sehingga dapat menyebabkan bias dalam evaluasi model
Trebeschi dkk.[137]	UK, Federasi Rusia	UK, Federasi Rusia	262	pencitraan CT	AUC (0,76–0,83)	Tahan	10 fitur radiografi	Model ekstraksi fitur berbasis ML berperforma lebih baik dibandingkan menggunakan penanda imun tunggal apa pun untuk memprediksi respons imunoterapi	Metrik yang digunakan dalam penelitian ini bukanlah intuisi persepsi. Menggunakan akurasi atau AUC mungkin lebih baik
Saltz dkk.[142]	CAE[143]	VGG16[144]+ DekonvNet[145]	4612 (13 jenis kanker)	Gambar histologis	AUC (0,9544)	Tahan	Fitur gambar slide histologis bagian tumor yang diwarnai H&E	Model ini dapat mengekstrak biomarker radiomik prediktif yang diturunkan dari CT untuk meningkatkan prediksi respons imunoterapi	Kinerja prediksi antara berbagai jenis kanker tidak kuat
							Fitur gambar slide histologis bagian tumor yang diwarnai H&E	Model ini mengungguli ahli patologi dan model patokan lainnya	Kinerja prediksi antara berbagai jenis kanker tidak kuat

Catatan:MRRN, resolusi jaringan yang terhubung secara sisa; CART, pohon klasifikasi dan regresi; AC-GAN, jaringan permusuhan generatif pengklasifikasi tambahan; Lcc, koefisien konkordansi Lin; Pcc, koefisien korelasi Pearson; MAE, berarti kesalahan mutlak; TPS, skor proporsional tumor; LD, ketidakseimbangan keterkaitan; Cox-PH, bahaya proporsional Cox; ANOVA, analisis varians; miRNA, mikroRNA; RPPA, susunan protein fase terbalik; CAE, autoencoder konvolusional; mRNA, RNA pembawa pesan; PD-L1, kematian sel terprogram 1 ligan 1; COSMIC, Katalog Mutasi Somatik Pada Kanker; EGFR, reseptor faktor pertumbuhan epidermal. Dibandingkan dengan hold-out, validasi silang biasanya lebih kuat, dan memperhitungkan lebih banyak varian antara kemungkinan pemisahan data pelatihan, validasi, dan pengujian. Namun, validasi silang lebih memakan waktu dibandingkan menggunakan metode ketidaksepakatan sederhana.

kumpulan data ekspresi gen microarray dan mencapai akurasi keseluruhan sebesar 83,0% dengan hanya lima gen teridentifikasi yang berkorelasi dengan waktu bertahan hidup. Liu dkk.[124]juga menggunakan data ekspresi gen untuk klasifikasi kelangsungan hidup 3 tahun. Berbeda dengan Chen dkk.[123], penulis mengintegrasikan tiga jenis data pengurutan — pengurutan RNA, metilasi DNA, dan mutasi DNA — untuk memilih total 22 gen guna meningkatkan stabilitas model mereka. Sedangkan LUADpp [125]dan Cho dkk.[126]menggunakan mutasi somatik sebagai fitur masukan untuk memodelkan klasifikasi risiko kelangsungan hidup 3 tahun. Untuk memilih gen yang terkait dengan angka kematian signifikan tertinggi, Cho et al.[126]menggunakan uji chi-kuadrat, dan LUADpp[125]menggunakan tes perbandingan tingkat genom yang dipublikasikan[127]yang mampu menyeimbangkan kekuatan statistik dan presisi untuk membandingkan tingkat mutasi gen. Karena kompleksitas prediksi kelangsungan hidup, data tumor multi-omics telah diintegrasikan untuk analisis dalam banyak penelitian. Dibandingkan dengan data single-omics, data multi-omics lebih sulit mengekstraksi gen paling signifikan untuk prediksi secara akurat. Untuk mengatasi masalah tersebut, beberapa penelitian[128–131] merancang alur kerja serupa. Mereka pertama-tama membuat matriks yang mewakili kesamaan antara pasien berdasarkan data multi-omics mereka. Dengan menggunakan matriks yang diperoleh, mereka kemudian menggunakan model pengelompokan tanpa pengawasan (biasanya autoencoder dengan pengelompokan K-means) untuk mengkategorikan pasien menjadi dua kelompok. Kedua klaster tersebut diberi label "risiko tinggi" dan "risiko rendah" dalam hal hasil kelangsungan hidup yang berbeda antara kedua klaster dalam analisis Kaplan – Meier. Mengikuti perbedaan hasil kelangsungan hidup, gen yang terkait dengan kematian diekstraksi menggunakan model statistik[128,129]atau model ML [130,131]untuk analisis hilir.

Terapkan ML untuk imunoterapi kanker paru-paru

Prediksi respon imunoterapi

Imunoterapi menjadi semakin penting dalam beberapa tahun terakhir. Hal ini memungkinkan sistem kekebalan tubuh pasien untuk melawan kanker, dalam banyak kasus, dengan menstimulasi sel T. Hingga saat ini, pengobatan imunoterapi baru yang berbeda sedang diuji untuk kanker paru-paru, dan beragam di antaranya telah menjadi bagian standar imunoterapi. Inhibitor pos pemeriksaan imun (ICI), terutama terapi blokade protein kematian sel terprogram 1 (PD-1)/protein kematian sel terprogram 1 (PD-L1) [132], telah terbukti bermanfaat dalam pengobatan pasien dengan kanker paru-paru non-sel kecil (NSCLC)[133,134]. Namun, imunoterapi belum banyak digunakan seperti pembedahan, kemoterapi, atau terapi radiasi. Salah satu interpretasinya adalah bahwa metode ini tidak berhasil pada semua pasien karena keunikan lingkungan mikro imun tumor (TIME) pasien. Oleh karena itu, memperkirakan apakah pasien akan merespons imunoterapi adalah penting dalam pengobatan kanker. Baru-baru ini, teknologi berbasis AI telah dikembangkan untuk memprediksi respons imunoterapi berdasarkan tanda pengurutan imun dan tanda pencitraan medis (Gambar 4;Tabel 3)[135]. Untuk memprediksi respon terhadap terapi blokade PD-1/PD-L1, Wiesweg et al.[136] menggunakan profil ekspresi gen dari 7 gen signifikan yang diekstraksi dari model ML ditambah 25 gen spesifik tipe sel sebagai fitur masukan untuk melatih pengklasifikasi SVM untuk klasifikasi RECIST. Selain mengurutkan data, fitur dari CT scan juga dapat digunakan untuk menilai tingkat RECIST pasien. Dua penelitian terbaru[137,138]menggunakan biomarker radiomik juga

sebagai fitur pencitraan lesi tumor lainnya dari pemindaian kontrasenhanced computer tomography (CE-CT) untuk melatih pengklasifikasi, termasuk LR dan RF, untuk klasifikasi RECIST.

Evaluasi limfosit yang menginfiltrasi tumor

Proporsi limfosit yang menginfiltrasi tumor (TILs) adalah metrik penting lainnya untuk evaluasi respon imunoterapi. Untuk tujuan ini, menggunakan data transkriptomik, DeepTIL[139] mengoptimalkan model dekonvolusi sel CIBERSORT[140]untuk secara otomatis menghitung kelimpahan subset leukosit (sel B, CD4+sel T, CD8+sel T,CD8sel T, sel Mo-Ma-DC, dan granulosit) dalam sampel tumor. Pendekatan yang berbeda[141] memanfaatkan total 84 fitur radiomik dari pemindaian CE-CT, bersama dengan pengurutan RNA dari 20.530 gen sebagai biomarker untuk melatih model regresi jaring elastis linier guna memprediksi kelimpahan CD8+sel T. Studi lain[142] membuat model DL untuk mengidentifikasi TIL dalam gambar H&E stained digital (Tabel 3). Metodologi ini terdiri dari dua modul CNN unik untuk mengevaluasi TIL pada skala berbeda: CNN klasifikasi infiltrasi limfosit (CNN limfosit) dan CNN segmentasi nekrosis (CNN nekrosis). "CNN limfosit" bertujuan untuk mengkategorikan gambar masukan ke dalam daerah infiltrasi limfosit dan tanpa limfosit. Ini terdiri dari dua langkah: autoencoder konvolusional (CAE)[143]untuk ekstraksi fitur, diikuti oleh jaringan 16 lapisan VGG[144] untuk klasifikasi wilayah TIL. "Necrosis CNN" bertujuan untuk mendeteksi TIL di wilayah nekrosis. Mereka menggunakan DeconvNet [145]model untuk segmentasi TIL di "necrosis CNN" sebagai model yang telah terbukti mencapai akurasi tinggi dengan beberapa kumpulan data pencitraan benchmark.

Prediksi neoantigen

Selain prediksi respons imunoterapi, algoritma ML telah menjelaskan prediksi neoantigen untuk imunoterapi. Neoantigen adalah peptida mutasi spesifik tumor yang dihasilkan oleh mutasi somatik pada sel tumor, yang dapat menginduksi respons imun antitumor.[146–148]. Penelitian terbaru menunjukkan bahwa neoantigen imunogenik bermanfaat bagi pengembangan dan optimalisasi terapi imun bertarget neoantigen[149–152]. Sesuai dengan studi neoantigen dalam uji klinis, pendekatan ML yang canggih telah diterapkan untuk mengidentifikasi neoantigen berdasarkan pemrosesan dan presentasi antigen leukosit manusia (HLA) kelas I dan II.[153–157]. Dengan menggunakan mutasi somatik yang teridentifikasi, model ML dapat memperkirakan afinitas pengikatan peptida bermutasi yang dikodekan terhadap alel HLA pasien (afinitas pengikatan peptida-HLA). Neoantigen dapat diprediksi lebih lanjut berdasarkan perkiraan afinitas pengikatan peptida-HLA. NetMHC[158,159]menggunakan kumpulan data reseptor-ligan yang terdiri dari 528 interaksi pengikatan peptida-HLA yang diukur oleh Buus et al.[160]untuk melatih kombinasi beberapa NN untuk prediksi afinitas neo-peptida. Untuk membuat prediksi lebih akurat, NetMHCpan[161,162]menggunakan kumpulan data yang lebih besar yang terdiri dari 37.384 interaksi peptida-HLA unik yang mencakup 24 alel HLA-A dan 18 alel HLA-B (masing-masing 26.503 dan 10.881 untuk alel A dan B) untuk melatih model NN mereka. Kedua alat tersebut telah diterapkan untuk mempelajari lanskap neoantigen pada kanker paru-paru[146,163–165].

Tantangan dan perspektif masa depan

Meskipun studi ML digunakan secara luas dalam praktik dan penelitian klinis kanker paru-paru, masih ada tantangan yang harus diatasi. Di sini, kami memposting beberapa contoh algoritma ML terbaru, terutama algoritma DL yang semakin populer dan penting dalam dekade terakhir, untuk memberikan pencerahan kepada mereka mengenai analisis terapi kanker paru-paru, serta tantangan untuk penelitian kanker paru-paru di masa depan.

Analisis data pencitraan

Mempelajari cara mengekstraksi nuansa secara efektif dari data pencitraan sangat penting untuk penggunaan klinis. Dalam sistem CAD berbasis ML sebelumnya, ekstraksi fitur biasanya didasarkan pada intensitas gambar, bentuk, dan tekstur wilayah yang mencurigakan serta variabel klinis lainnya.[166]. Namun, pendekatan ini didefinisikan secara sewenang-wenang dan mungkin tidak mengambil fitur intrinsik dari nodul yang mencurigakan. Untuk mencapai tujuan ini, sistem CAD berbasis DL dikembangkan dengan memanfaatkan model CNN untuk mengekstrak fitur langsung dari data pencitraan mentah dengan representasi bertingkat dan abstraksi hierarki.[167–169]. Berbeda dengan metode sebelumnya, fitur dari model CNN tidak dirancang oleh manusia, dan mencerminkan fitur intrinsik nodul secara obyektif dan komprehensif. Baru-baru ini, Vision Transformer (ViT) telah muncul sebagai teknologi terancang dalam visi komputer[170,171]. Dibandingkan dengan CNN, ViT mengungguli hampir 4- dalam hal efisiensi dan akurasi komputasi, dan lebih kuat ketika dilatih pada kumpulan data yang lebih kecil[172]. Meskipun, sepengetahuan kami, model ViT belum diterapkan dalam studi pencitraan kanker paru-paru, model tersebut telah menunjukkan potensinya sebagai alternatif kompetitif dibandingkan CNN dalam analisis data pencitraan.

Analisis kumpulan data Omics

DL adalah subbidang ML, yang menggunakan NN yang dapat diprogram untuk membuat keputusan yang akurat. Ini sangat menonjol ketika menyangkut masalah kompleks seperti klasifikasi gambar. Dalam studi ini, kami meninjau kegunaan model DL dalam kumpulan data pencitraan. Dibandingkan dengan kumpulan data pencitraan, algoritme DL lebih jarang digunakan dalam studi klinis kanker paru-paru yang menggunakan data omics. Namun, model DL telah diterapkan secara luas di bidang analisis omics lainnya. Misalnya, data genomik adalah rangkaian kontinu, sehingga model jaringan saraf berulang (RNN).[173] dan model CNN[174] adalah alat yang baik untuk analisis genetika populasi. Selain itu, mengingat dimensi input data omics biasanya sangat tinggi, untuk meningkatkan efisiensi dan mengurangi overfitting, banyak penelitian telah menggunakan autoencoder atau model generatif mendalam untuk ekstraksi fitur dan pengurangan dimensi.[175]. Sementara itu, model pembelajaran representasi mandiri dapat mengatasi kutukan dimensi dan mengintegrasikan data multi-omics untuk menggabungkan informasi tentang berbagai aspek dari sampel jaringan yang sama.[176]. Disertai dengan pengembangan yang berbasis sel tunggal [177] dan berbasis spasial[178] teknologi yang telah diterapkan dalam studi molekuler, banyak model DL menjadi lebih populer untuk analisis komputasi intensif. Untuk mengatasi kompleksitas data genomik yang besar, alat pengelompokan mendalam tanpa pengawasan telah dibuat untuk struktur populasi

identifikasi[179] atau anotasi sub tipe populasi sel [180–183]. Selain itu, untuk memproses struktur kompleks data multi-omics, model graph neural network (GNN) semakin populer dalam integrasi dataset[184], klasifikasi biomedis[185], prediksi prognosis[186], dan seterusnya. Meskipun studi-studi ini belum diterapkan secara langsung pada analisis klinis kanker paru-paru, studi-studi ini merupakan inspirasi yang baik untuk menggunakan alat DL untuk menangani kumpulan data omics kanker paru-paru yang kompleks.

Data multi-tampilan dan integrasi multi-database

Saat ini, mengakses sejumlah besar data pencitraan, data multiomik, dan catatan klinis dari satu pasien adalah hal yang umum. Mengintegrasikan data ini memberikan wawasan komprehensif mengenai fungsi molekuler penelitian kanker paru-paru. Namun, tipe data ini biasanya diperoleh dari platform yang berbeda, sehingga gangguan platform pasti ada di antara tipe data ini. Misalnya, analisis data pencitraan, khususnya radiomik, biasanya memiliki tantangan berupa normalisasi data yang rumit, penggabungan data, dan integrasi data. Untuk mengatasi keterbatasan ini, jaringan segmentasi medis multimodalitas telah dikembangkan untuk bersama-sama memproses citra medis multimodalitas[187]. Demikian pula, untuk pengurutan tipe data, gangguan batch juga terjadi di antara database yang berbeda (yaitu, efek kumpulan). Menghapus efek batch dan mengintegrasikan kumpulan data dari berbagai platform bersama-sama dalam kerangka kerja yang memungkinkan kita menganalisis lebih lanjut mekanisme resistensi dan kekambuhan obat kanker adalah hal yang penting untuk terapi kanker. Meskipun penelitian biomedis telah bereksperimen dan/atau membuat tolak ukur alat integratif[188–191], instrumen tersebut tidak komprehensif dan cukup diskriminatif untuk menjawab pilihan alat dalam konteks pertanyaan biologis yang menjadi perhatian.

Generalisasi dan ketahanan model

Dalam tinjauan ini, kami menemukan bahwa performa algoritme ML biasanya bervariasi di berbagai kumpulan data. Salah satu interpretasinya mungkin adanya efek batch database yang telah kita bahas sebelumnya. Namun, kurangnya kemampuan generalisasi dan ketahanan mungkin menjadi faktor lain yang menghambat model ML ini dalam studi klinis. Selain itu, untuk mengurangi overfitting, sebagian besar penelitian menggunakan model statistik atau model ML untuk memilih gen penanda sebelum klasifikasi. Namun, gen penanda ini biasanya sangat berbeda antar penelitian, yang menunjukkan bahwa gen penanda yang teridentifikasi kurang memiliki kemampuan generalisasi dan interpretasi biologis. Untuk meningkatkan kemampuan generalisasi dan ketahanan suatu model, penting untuk mengembangkan pemahaman yang lebih baik tentang masalah ketahanan dalam berbagai arsitektur ML dan menjembatani kesenjangan dalam teknik ketahanan di berbagai domain. Misalnya, penelitian terbaru telah menerapkan pembelajaran transfer untuk menggunakan model yang telah dilatih sebelumnya saat melatih kumpulan data mereka sendiri dalam analisis data pencitraan kanker paru-paru.[38,55,192], dan telah meningkatkan efisiensi dan ketahanan model berbasis CNN mereka. Untuk mengurutkan kumpulan data, pembelajaran transfer juga telah digunakan di NN mendalam untuk menyediakan pendekatan generalisasi[193], yang dapat menjadi contoh bagus dalam membangun model umum dan kuat untuk analisis data pengurutan kanker paru-paru. Selain itu, DL merupakan model kotak hitam yang kompleks. Memahami mekanisme sistem DL dalam studi klinis dapat membantu membangun kerangka DL yang terstandarisasi dan terpadu untuk meningkatkan kinerja dan ketahanannya. Itu

Model AI yang dapat dijelaskan (XAI) telah menyediakan alat untuk analisis model spesifik dan model-agnostik[194.195]. Metode-metode ini dapat memberikan penjelasan suatu model di tingkat lokal dan global, yang selanjutnya membantu para peneliti untuk menyempurnakan hyperparameter dari berbagai model dengan kemanjuran tinggi.[196.197].

Metrik untuk evaluasi kinerja

Studi biasanya fokus pada pengembangan algoritma untuk studi klinis. Namun, pemilihan metrik untuk penilaian performa algoritme ini biasanya diabaikan, meskipun biasanya metrik tersebut memainkan komponen penting dalam sistem ML[198]. Berdasarkan ulasan ini (Tabel 1–3), akurasi dan di bawah kurva (AUC) adalah dua metrik yang paling konvensional, sedangkan metrik ini tidak selalu mencerminkan kebutuhan klinis dan harus diterjemahkan ke dalam metrik yang dapat dijelaskan secara klinis. Dibandingkan dengan akurasi, sensitivitas atau spesifisitas mungkin lebih terkait dengan kebutuhan klinis dalam keadaan tertentu, misalnya, pasien berisiko tinggi mengunjungi unit gawat darurat[199].

Pengambilan keputusan klinis

Sebuah penelitian baru-baru ini memperkirakan bahwa biaya keseluruhan untuk terapi kanker paru-paru akan melebihi \$50.000[200] bagi sebagian besar pasien, dan biayanya akan mahal bagi sebagian besar keluarga. Oleh karena itu, prediksi prognosis yang akurat dan pengambilan keputusan akan membuka jalan bagi pengobatan yang dipersonalisasi. Model DL terbaru telah digunakan untuk memprediksi efektivitas terapi/obat dan mengoptimalkan kombinasi terapi/obat yang berbeda.[201.202]. Namun, sebagian besar model DL yang ada untuk pengambilan keputusan klinis mengalami kesulitan dalam mengikuti evolusi pengetahuan dan/atau perubahan data layanan kesehatan yang dinamis.[203]. Saat ini, sistem pendukung keputusan klinis, termasuk IBM Watson Health dan Google DeepMind Health, telah diterapkan dalam pengobatan kanker paru-paru dalam beberapa tahun terakhir.[204.205]. Meskipun efisiensi kerja klinis telah meningkat dengan bantuan sistem ini, sistem ini masih jauh dari sempurna dalam hal uji klinis, dan saat ini tidak dapat menggantikan dokter pada tahap ini.[205].

Kesimpulan

AI memberi kita perspektif berbeda tentang penelitian kanker paru-paru dan memungkinkan eksplorasi penerapan alat pendukung keputusan untuk memfasilitasi onkologi yang presisi. Dalam ulasan ini, kami mensurvei kemajuan terkini algoritma ML di berbagai bidang terapi kanker paru-paru, termasuk deteksi dini, keputusan diagnosis, prediksi prognosis, evaluasi respons obat, dan praktik imunoterapi. Untuk membantu pengembangan ML di masa depan dalam terapi kanker paru-paru, kami merangkum kumpulan data secara menyeluruh (Tabel S1), metode dasar (Tabel S2), dan karakteristik metode (Tabel 1–3). Terakhir, kami menyoroti tantangan saat ini yang perlu diatasi, seperti kurangnya kuantitas dan kualitas label data medis untuk pelatihan, pentingnya kekhawatiran terhadap metrik yang digunakan untuk kinerja, evaluasi, dan kebutuhan untuk integrasi data dan penghapusan batch. Seperti yang ditunjukkan oleh tinjauan ini, terapi kanker paru-paru di masa depan akan mencakup data pencitraan dan data omics, sehingga alat pengambilan keputusan klinis ML harus bersifat multi-modal.

sistem yang mempertimbangkan pengolahan data pencitraan dan data omics, serta integrasi beberapa tipe data. Pada akhirnya, kami berharap tantangan-tantangan ini dapat memotivasi penelitian lebih lanjut untuk fokus pada terapi kanker paru-paru.

Pernyataan penulis CRediT

Yawei Li:Konseptualisasi, Kurasi data, Penulisan - draf asli, Visualisasi.Xin Wu:Kurasi data, Penulisan - draf asli, Visualisasi. Pingyang:Menulis - mengulas & mengedit. Guoqian Jiang: Menulis - mengulas & mengedit.Yuan Luo:Konseptualisasi, Akuisisi pendanaan, Penulisan - review & editing. Semua penulis telah membaca dan menyetujui naskah akhir.

Kepentingan yang bersaing

Para penulis telah menyatakan tidak ada kepentingan yang bersaing.

Ucapan Terima Kasih

Penelitian ini sebagian didukung oleh National Institutes of Health, AS (Grant Nos. U01TR003528 dan R01LM013337).

Materi tambahan

Data tambahan untuk artikel ini dapat ditemukan online di <https://doi.org/10.1016/j.gpb.2022.11.003>.

ORCID

ORCID 0000-0001-9699-5118 (Yawei Li) ORCID
0000-0003-2386-6344 (Xin Wu) ORCID
0000-0002-8588-847X (Ping Yang) ORCID
0000-0003-2940-0019 (Guoqian Jiang) ORCID
0000 -0003-0195-7456 (Yuan Luo)

Referensi

- [1]AA Thailand, Solomon BJ, Sequist LV, Gainor JF, Heist RS. Kanker paru-paru. *Lancet* 2021;398:535–54.
- [2]Svoboda E. Kecerdasan buatan meningkatkan deteksi kanker paru-paru. *Alam* 2020;587:S20–2.
- [3]Ling S, Hu Z, Yang Z, Yang F, Li Y, Lin P, dkk. Keragaman genetik yang sangat tinggi pada satu tumor menunjukkan prevalensi evolusi sel non-darwinian. *Proc Natl Acad Sci AS* 2015;112: E6496–505.
- [4]Konsorsium Genom Kanker Internasional, Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, dkk. Jaringan internasional proyek genom kanker. *Alam* 2010;464:993–8.
- [5]Jaringan Penelitian Atlas Genom Kanker, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, dkk. Proyek analisis Cancer Genome Atlas Pan-Cancer. *Nat Genet* 2013;45:1113–20.
- [6]Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, dkk. The Cancer Imaging Archive (TCIA): memelihara dan mengoperasikan gudang informasi publik. *Pencitraan J Digit* 2013;26:1045–57.
- [7]Pavlopoulou A, Spandidos DA, Michalopoulos I. Database kanker manusia (ulasan). *Perwakilan Oncol* 2015;33:3–18.
- [8]Luo Y, Wang F, Szolovits P. Faktorisasi tensor menuju pengobatan presisi. *Bioinform Singkat* 2016;18:511–4.

- [9] Kolda TG, Bader BW. Dekomposisi dan aplikasi tensor. *SIAM Rev* 2009;51:455–500.
- [10] ChaoG, MaoC, WangF, ZhaoY, LuoY. Faktorisasi matriks non-negatif diawasi untuk memprediksi risiko kematian ICU. *Prosiding (IEEE Int Conf Bioinformatics Biomed)* 2018;2018:1189–94.
- [11] Chi EC, Kolda TG. Tentang tensor, ketersebaran, dan faktorisasi nonnegatif. *Aplikasi Anal Matriks SIAM J* 2012;33:1272–99.
- [12] Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Pembelajaran mendalam dalam diagnosis kanker, prognosis dan pemilihan pengobatan. *Genom Med* 2021;13:152.
- [13] Zeng Z, Yao L, Roy A, Li X, Espino S, Clare SE, dkk. Mengidentifikasi kekambuhan kanker payudara yang jauh dari catatan kesehatan elektronik menggunakan pembelajaran mesin. *J Healthc Informasikan Res* 2019;3:283–99.
- [14] Cruz JA, Wishart DS. Penerapan pembelajaran mesin dalam prediksi dan prognosis kanker. *Informasi Kanker* 2007;2:59–77.
- [15] Wang H, Li Y, Khan SA, Luo Y. Prediksi kekambuhan kanker payudara jauh menggunakan pemrosesan bahasa alami dan jaringan saraf konvolusional yang dipandu pengetahuan. *Artif Intell Med* 2020;110:101977.
- [16] Cochran AJ. Prediksi hasil untuk pasien dengan melanoma kulit. *Res Sel Pigmen* 1997;10:162–7.
- [17] Zeng Z, Li X, Espino S, Roy A, Kitsch K, Clare S, dkk. Deteksi kejadian kanker payudara kontralateral menggunakan pemrosesan bahasa alami. *Proc Sim AMIA Annu* 2018;2017:1885–92.
- [18] Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Aplikasi pembelajaran mesin dalam prognosis dan prediksi kanker. *Struktur Komputasi Bioteknologi J* 2015;13:8–17.
- [19] Luo Y, Xin Y, Hochberg E, Joshi R, Uzuner O, Szolovits P. Subgraph augmented faktorisasi tensor non-negatif (SANTF) untuk memodelkan teks narasi klinis. *J Am Med Informasikan Assoc* 2015;22:1009–19.
- [20] Benzekry S. Kecerdasan buatan dan pemodelan mekanistik untuk pengambilan keputusan klinis di bidang onkologi. *Clin Pharmacol Ada* 2020;108:471–86.
- [21] Li Y, Luo Y. Mengoptimalkan evaluasi panel yang ditargetkan gen untuk estimasi beban mutasi tumor. *Perwakilan Sains* 2021;11:21072.
- [22] Bhinder B, Gilvary C, Madhukar NS, Elemento O. Kecerdasan buatan dalam penelitian kanker dan pengobatan presisi. *Penemuan Kanker* 2021;11:900–15.
- [23] Zeng Z, Espino S, Roy A, Li X, Khan SA, Clare SE, dkk. Menggunakan pemrosesan bahasa alami dan pembelajaran mesin untuk mengidentifikasi kekambuhan lokal kanker payudara. *BMC Bioinformatika* 2018;19:498.
- [24] Luo Y, Sohani AR, Hochberg EP, Szolovits P. Klasifikasi limfoma otomatis dengan penambahan subgraf kalimat dari laporan patologi. *J Am Med Informasikan Assoc* 2014;21:824–32.
- [25] Luchini C, Pea A, Scarpa A. Kecerdasan buatan dalam onkologi: aplikasi saat ini dan perspektif masa depan. *Br J Kanker* 2022;126:4–9.
- [26] Zeng Z, Vo A, Li X, Shidfar A, Saldana P, Blanco L, dkk. Penyimpangan genetik somatik pada penyakit payudara jinak dan risiko kanker payudara selanjutnya. *NPJ Kanker Payudara* 2020;6:24.
- [27] Na J, Zong N, Wang C, Midthun DE, Luo Y, Yang P, dkk. Mengkarakterisasi kelainan fenotipik yang terkait dengan individu berisiko tinggi terkena kanker paru-paru menggunakan catatan kesehatan elektronik dari Kita semua kerja peneliti. *J Am Med Informasikan Assoc* 2021;28:2313–24.
- [28] Diagnosis berbantuan komputer berbasis Fujita H. AI (AI-CAD): ulasan terbaru untuk dibaca terlebih dahulu. *Radiol Phys Technol* 2020;13:6–19.
- [29] Yanase J, Triantaphyllou E. Sebuah survei sistematis diagnosis berbantuan komputer dalam kedokteran: perkembangan masa lalu dan sekarang. *Aplikasi Sistem Pakar* 2019;138:112821.
- [30] Abe Y, Hanai K, Nakano M, Ohkubo Y, Hasizume T, Kakizaki T, dkk. Sistem diagnosis berbantuan komputer (CAD) dalam skrining kanker paru-paru dengan tomografi komputer. *Res Antikanker* 2005;25:483–8.
- [31] McWilliams A, Tammemagi MC, Mayo JR, Roberts H, Liu G, Soghrati K, dkk. Kemungkinan kanker pada nodul paru terdeteksi pada pemeriksaan CT pertama. *N Engl J Med* 2013;369:910–9.
- [32] van Riel SJ, Ciompi F, Wille MMW, Dirksen A, Lam S, Scholten ET, dkk. Estimasi risiko keganasan nodul paru dalam skrining CT: perbandingan antara model komputer dan pengamat manusia. *PLoS Satu* 2017;12:e0185032.
- [33] Wille MMW, van Riel SJ, Saghir Z, Dirksen A, Pedersen JH, Jacobs C, dkk. Keakuratan prediktif model prediksi risiko kanker paru-paru pancan — validasi eksternal berdasarkan CT dari Uji Coba Skrining Kanker Paru-Paru Denmark. *Radio Euro* 2015;25:3093–9.
- [34] Kriegsmann M, Casadonte R, Kriegsmann J, Dienemann H, Schirmacher P, Kobarg JH, dkk. Subtipe entitas yang andal pada kanker paru-paru non-sel kecil dengan desorpsi laser berbantuan matriks/pencitraan ionisasi spektrometri massa pada spesimen jaringan yang tertanam parafin dengan formalin. *Proteomik Sel Mol* 2016;15:3081–9.
- [35] Mohammad BA, Brennan PC, Mello-Thoms C. Tinjauan skrining kanker paru-paru dan peran deteksi dengan bantuan komputer. *Klinik Radiol* 2017;72:433–42.
- [36] Armato 3rd SG, Li F, Giger ML, MacMahon H, Sone S, Doi K. Kanker paru-paru: kinerja deteksi nodul paru otomatis yang diterapkan pada kanker yang terlewat dalam program skrining CT. *Radiologi* 2002;225:685–92.
- [37] Buty M, Xu Z, Gao M, Bagci U, Wu A, Mollura D. Karakterisasi keganasan nodul paru menggunakan fitur bentuk dan penampilan hibrida. Dalam: Ourselin S, Joskowicz L, Sabuncu M, Unal G, Wells W, editor. *Komputasi citra medis dan intervensi bantuan komputer*. Cham: Peloncat; 2016, hal.662–70.
- [38] Hussein S, Cao K, Song Q, Bagci U. Stratifikasi risiko nodul paru menggunakan pembelajaran multi-tugas berbasis CNN 3D. Dalam: Niethammer M, Styner M, Aylward S, Zhu H, Oguz I, Yap PT, editor. *Pemrosesan informasi dalam pencitraan medis*. Cham: Peloncat; 2017, hal.249–60.
- [39] Khosravan N, Celik H, Turkbey B, Jones EC, Wood B, Bagci U. Sistem diagnosis berbantuan komputer kolaboratif (C-CAD) dengan pelacakan mata, model perhatian yang jarang, dan pembelajaran mendalam. *Anal Gambar Med* 2019;51:101–15.
- [40] Ciompi F, de Hoop B, van Riel SJ, Chung K, Scholten ET, Oudkerk M, dkk. Klasifikasi otomatis nodul perifissural paru dalam tomografi komputer menggunakan kumpulan tampilan 2D dan jaringan saraf konvolusional yang siap pakai. *Anal Gambar Med* 2015;26:195–202.
- [41] Krizhevsky A, Sutskever I, Hinton GE. Klasifikasi Imagenet dengan jaringan saraf konvolusional yang dalam. *Komunitas ACM* 2017;60:84–90.
- [42] Sermanet P, Eigen D, Zhang X, Mathieu M, Fergus R, LeCun Y. OverFeat: pengenalan, lokalisasi, dan deteksi terintegrasi menggunakan jaringan konvolusional. *arXiv* 2014;1312.6229.
- [43] Gu X, Wang Y, Chan TF, Thompson PM, Yau ST. Pemetaan konformal permukaan nol genus dan penerapannya pada pemetaan permukaan otak. *Pencitraan Med Proses Inf* 2003;18:172–84.
- [44] Venkadesh KV, Setio AAA, Schreuder A, Scholten ET, Chung KM, Wille MMW, dkk. Pembelajaran mendalam untuk estimasi risiko keganasan nodul paru yang terdeteksi pada CT skrining dosis rendah. *Radiologi* 2021;300:438–47.
- [45] He K, Zhang X, Ren S, Sun J. Pembelajaran sisa yang mendalam untuk pengenalan gambar. *IEEEConfComputVisPatternRecognit* 2016:770–8.
- [46] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, dkk. Lebih dalam dengan konvolusi. *Pengenalan Pola IEEE Conf Comput Vis* 2015:1–9.
- [47] Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, dkk. Skrining kanker paru menyeluruh dengan pembelajaran mendalam tiga dimensi pada tomografi komputer dada dosis rendah. *Nat Med* 2019;25:954–61.
- [48] He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. *Vis Komputasi Int Conf IEEE* 2017:2980–8.
- [49] Lin TY, Goyal P, Girshick R, He K, Dollár P. Kehilangan fokus untuk deteksi objek padat. *IEEE Trans Pola Anal Mach Intell* 2020;42:318–27.

- [50]Carreira J, Zisserman A. Quo vadis, pengenalan tindakan? Model baru dan kumpulan data kinetika. *Pengenalan Pola IEEE Conf Comput Vis* 2017;4724–33.
- [51]Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Memikirkan kembali arsitektur awal untuk visi komputer. *Pengenalan Pola IEEE Conf Comput Vis* 2016;2818–26.
- [52]AbdulJabbar K, Raza SEA, Rosenthal R, Jamal-Hanjani M, Veeriah S, Akarca A, dkk. Variabilitas imun geospasial menjelaskan evolusi diferensial adenokarsinoma paru. *Nat Med* 2020;26:1054–62.
- [53]Raza SEA, Cheung L, Shaban M, Graham S, Epstein D, Pelengaris S, dkk. Micro-Net: model terpadu untuk segmentasi berbagai objek dalam gambar mikroskop. *Anal Gambar Med* 2019;52:160–73.
- [54]Sirinukunwattana K, Raza SEA, Tsang YW, Snead DRJ, Cree IA, Rajpoot NM. Pembelajaran mendalam yang sensitif terhadap lokalitas untuk mendeteksi dan mengklasifikasikan inti dalam gambar histologi kanker usus besar rutin. *Pencitraan Med Trans IEEE* 2016;35:1196–206.
- [55]Ocampo P, Moreira A, Coudray N, Sakellariopoulos T, Narula N, Snuderl M, dkk. Klasifikasi dan prediksi mutasi dari gambar histopatologi kanker paru non-sel kecil menggunakan deep learning. *J Thorac Oncol* 2018;13:S562.
- [56]Yao Q, Xiao L, Liu P, Zhou SK. Segmentasi lesi COVID-19 tanpa label pada CT paru. *Pencitraan Med Trans IEEE* 2021;40:2808–19.
- [57]Chuquicusma MJM, Hussein S, Burt J, Bagci U. Bagaimana cara menipu ahli radiologi dengan jaringan permusuhan generatif? Tes turing visual untuk diagnosis kanker paru-paru. *Pencitraan Biomed Gejala Int IEEE* 2018:240–4.
- [58]Li J, Jia J, Xu D. Pembelajaran representasi tanpa pengawasan penyakit tanaman berbasis gambar dengan jaringan permusuhan generatif konvolusional yang dalam. *Konferensi Kontrol Tiongkok ke-37* 2018:9159–63.
- [59]Lin CH, Lin CJ, Li YC, Wang SH. Menggunakan jaringan permusuhan generatif dan optimalisasi parameter jaringan saraf konvolusional untuk klasifikasi tumor paru-paru. *Aplikasi Sains* 2021;11:480.
- [60]Ren Z, Zhang Y, Wang S. Kerangka kerja hibrida untuk klasifikasi kanker paru-paru. *Elektronik* 2022;11:1614.
- [61]Pinsky PF, Gierada DS, Black W, Munden R, Nath H, Aberle D, dkk. Kinerja RADS paru-paru dalam uji coba skrining paru-paru nasional: penilaian retrospektif. *Ann Magang Med* 2015;162:485–91.
- [62]Tim Peneliti Uji Coba Skrining Paru Nasional, Church TR, Black WC, Aberle DR, Berg CD, Clingan KL, dkk. Hasil skrining tomografi komputer dosis rendah awal untuk kanker paru-paru. *N Engl J Med* 2013;368:1980–91.
- [63]Herath S, Rad HS, Radfar P, Ladwa R, Warkiani M, O'Byrne K, dkk. Peran biomarker yang bersirkulasi pada kanker paru-paru. *Oncol Depan* 2021;11:801269.
- [64]Politi K, Herbst RS. Kanker paru-paru di era pengobatan presisi. *Klinik Kanker Res* 2015;21:2213–20.
- [65]Relli V, Trerotola M, Guerra E, Alberti S. Meninggalkan gagasan kanker paru-paru non-sel kecil. *Tren Mol Med* 2019;25:585–94.
- [66]Chen JW, Dhahbi J. Klasifikasi adenokarsinoma paru dan karsinoma sel skuamosa paru, identifikasi biomarker, dan analisis ekspresi gen menggunakan metode pemilihan fitur yang tumpang tindih. *Perwakilan Sains* 2021;11:13323.
- [67]Zeng Z, Vo AH, Mao C, Clare SE, Khan SA, Luo Y. Klasifikasi kanker dan penemuan jalur menggunakan faktorisasi matriks non-negatif. *J Biomed Informasikan* 2019;96:103247.
- [68]Jiao W, Atwal G, Polak P, Karlic R, Cuppen E, Danyi A, dkk. Sistem pembelajaran mendalam secara akurat mengklasifikasikan kanker primer dan metastasis menggunakan pola mutasi penumpang. *Nat Commun* 2020;11:728.
- [69]Li Y, Luo Y. Model pemungutan suara berbobot kinerja: metode pembelajaran mesin ansambel untuk klasifikasi jenis kanker menggunakan mutasi pengurutan seluruh exome. *Biol Kuantitas* 2020;8:347–58.
- [70]Eraslan G, Avsec Z, Gagneur J, Theis FJ. Pembelajaran mendalam: teknik pemodelan komputasi baru untuk genomik. *Nat Rev Genet* 2019;20:389–403.
- [71]Luo Y, Mao C. Panther: jalur menambah faktorisasi tensor non-negatif untuk pembelajaran fitur tingkat tinggi. *Konferensi Proc AAAI Artif Intell* 2021:37–180.
- [72]Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RCT, dkk. Prediksi hasil limfoma sel B besar yang menyebar dengan profil ekspresi gen dan pembelajaran mesin yang diawasi. *Nat Med* 2002;8:68–74.
- [73]Kononenko I. Pembelajaran mesin untuk diagnosis medis: sejarah, kecanggihan, dan perspektif. *Artif Intell Med* 2001;23:89–109.
- [74]Luo Y, Mao C. ScanMap: faktorisasi matriks non-negatif sadar perancu yang diawasi untuk pemodelan risiko poligenik. *Proc Mach Pelajari Res* 2020;126:27–45.
- [75]Zeng Z, Mao C, Vo A, Li X, Nugent JO, Khan SA, dkk. Pembelajaran mendalam untuk klasifikasi jenis kanker dan identifikasi gen penggerak. *BMC Bioinformatika* 2021;22:491.
- [76]Zhang Y, Li Y, Li T, Shen X, Zhu T, Tao Y, dkk. Beban genetik dan potensi kehancuran mutasi pada populasi sel kanker. *Mol Biol Evol* 2019;36:541–52.
- [77]Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, dkk. Tanda tangan proses mutasi pada kanker manusia. *Alam* 2013;500:415–21.
- [78]Mathios D, Johansen JS, Cristiano S, Medina JE, Phallen J, Larsen KR, dkk. Deteksi dan karakterisasi kanker paru-paru menggunakan fragmentom DNA bebas sel. *Nat Commun* 2021;12:5060.
- [79]Chabon JJ, Hamilton EG, Kurtz DM, Esfahani MS, Moding EJ, Stehr H, dkk. Mengintegrasikan fitur genom untuk deteksi dini kanker paru non-invasif. *Alam* 2020;580:245–51.
- [80]Liang W, Zhao Y, Huang W, Gao Y, Xu W, Tao J, dkk. Diagnosis noninvasif kanker paru stadium awal menggunakan sekuensing metilasi DNA bertarget throughput tinggi dari DNA tumor yang bersirkulasi (ctDNA). *Theranostik* 2019;9:2056–70.
- [81]Raman L, van der Linden M, van der Eecken K, Vermaelen K, Demedts I, Surmont V, dkk. Pengurutan seluruh genom dangkal DNA bebas sel plasma secara akurat membedakan karsinoma paru-paru kecil dan non-sel kecil. *Genom Med* 2020;12:35.
- [82]Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. *Biologi molekuler sel*. edisi ke-4. New York: Ilmu Karangan Bunga; 2002.
- [83]Kobayashi K, Bolakan A, Shiina S, Hamamoto R. Jaringan saraf yang terhubung sepenuhnya dengan parameterisasi yang dikurangi untuk memprediksi jenis kanker paru-paru histologis dari mutasi somatik. *Biomolekul* 2020;10:1249.
- [84]Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, dkk. Penemuan dan analisis saturasi gen kanker pada 21 jenis tumor. *Alam* 2014;505:495–501.
- [85]Uhlen M, Zhang C, Lee S, Sjostedt E, Fagerberg L, Bidkhorji G, dkk. Atlas patologi transkriptome kanker manusia. *Sains* 2017;357:eaan2507.
- [86]Whitney DH, Elashoff MR, Porta-Smith K, Gower AC, Vachani A, Ferguson JS, dkk. Penurunan pengklasifikasi genom bronkial untuk kanker paru-paru dalam studi prospektif pasien yang menjalani bronkoskopi diagnostik. *Genomik Med BMC* 2015;8:18.
- [87]Podolsky MD, Barchuk AA, Kuznetsov VI, Gusarova NF, Gaidukov VS, Tarakanov SA. Evaluasi pemanfaatan algoritma pembelajaran mesin untuk klasifikasi kanker paru berdasarkan tingkat ekspresi gen. *Kanker Pac J Asia Sebelumnya* 2016;17:835–8.
- [88]Choi Y, Qu J, Wu S, Hao Y, Zhang J, Ning J, dkk. Meningkatkan stratifikasi risiko kanker paru-paru dengan memanfaatkan pengurutan RNA transkriptome secara keseluruhan dan pembelajaran mesin di berbagai kelompok. *Genomik Med BMC* 2020;13:151.
- [89]Aliferis CF, Tsamardinos I, Massion PP, Statnikov A, Fananapazir N, Hardin D. Model pembelajaran mesin untuk klasifikasi kanker paru-paru dan pemilihan penanda genom menggunakan gen array

- data ekspresi. Prosiding Konferensi Masyarakat Penelitian Kecerdasan Buatan Florida Internasional ke-16 2003:67–71.
- [90] Shao X, Lv N, Liao J, Long JB, Xue R, Ai N, dkk. Variasi jumlah salinan sangat berkorelasi dengan ekspresi gen diferensial: studi pan-kanker. *BMC Med Genet* 2019;20:175.
- [91] Aliferis CF, Hardin D, Massion PP. Model pembelajaran mesin untuk klasifikasi kanker paru-paru menggunakan hibridisasi genom komparatif array. *Simpodium Proc AMIA* 2002:7–11.
- [92] Daemen A, Gevaert O, Leunen K, Legius E, Vergote I, De Moor B. Klasifikasi data array CGH yang diawasi dengan pemilihan fitur berbasis HMM. *Pac Symp Biocomput* 2009:468–79.
- [93] Jurmeister P, Bockmayr M, Seegerer P, Bockmayr T, Treue D, Montavon G, dkk. Analisis pembelajaran mesin terhadap profil metilasi DNA membedakan karsinoma sel skuamosa paru primer dari metastasis kepala dan leher. *Ilmu Terjemahan Sains* 2019;11: eaaw8513.
- [94] Luo Y, Riedlinger G, Szolovits P. Penambangan teks dalam gen kanker dan prioritas jalur. *Informasi Kanker* 2014;13:69–79.
- [95] Guyon I, Weston J, Barnhill S, Vapnik V. Seleksi gen untuk klasifikasi kanker menggunakan mesin vektor pendukung. *Mach Pelajari* 2002;46:389–422.
- [96] Mirhadi S, Tam S, Li Q, Moghal N, Pham NA, Tong J, dkk. Analisis integratif xenograft yang diturunkan dari pasien kanker paru-paru non-sel kecil mengidentifikasi proteotipe berbeda yang terkait dengan hasil akhir pasien. *Nat Commun* 2022;13:1811.
- [97] Xu JY, Zhang C, Wang X, Zhai L, Ma Y, Mao Y, dkk. Karakterisasi proteomik integratif adenokarsinoma paru manusia. *Sel* 2020;182:245–61.
- [98] El-Dereby W, Ashmore SM, Branston NM, Darling JL, Williams SR, Thomas DG. Prediksi pra-perawatan respon kemoterapi kultur sel glioma manusia menggunakan spektroskopi resonansi magnetik nuklir dan jaringan saraf tiruan. *Kanker Res* 1997;57:4196–9.
- [99] Zeng Z, Amin A, Roy A, Pulliam NE, Karavites LC, Espino S, dkk. Penggunaan pencitraan resonansi magnetik pra operasi dan hasil onkologis pada pasien kanker payudara pramenopause. *NPJ Kanker Payudara* 2020;6:49.
- [100] Chang Y, Park H, Yang HJ, Lee S, Lee KY, Kim TS, dkk. Pemindaian profil respons obat kanker (CDRscan): model pembelajaran mendalam yang memprediksi efektivitas obat dari tanda genom kanker. *Perwakilan Sains* 2018;8:8857.
- [101] Menden MP, Iorio F, Garnett M, McDermott U, Benes CH, Ballester PJ, dkk. Prediksi pembelajaran mesin sensitivitas sel kanker terhadap obat berdasarkan sifat genom dan kimia. *PLoS Satu* 2013;8:e61318.
- [102] Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, dkk. Kriteria evaluasi respons baru pada tumor padat: pedoman RECIST yang direvisi (versi 1.1). *Kanker Euro J* 2009;45:228–47.
- [103] Adam G, Rampasek L, Safikhani Z, Smirnov P, Haibe-Kains B, Goldenberg A. Pendekatan pembelajaran mesin untuk prediksi respons obat: tantangan dan kemajuan terkini. *NPJ Precis Oncol* 2020;4:19.
- [104] Jiang J, Hu YC, Liu CJ, Halpenny D, Hellmann MD, Deasy JO, dkk. Aliran fitur beresolusi ganda yang terhubung secara sisa untuk segmentasi tumor paru otomatis dari gambar CT. *Pencitraan Med Trans IEEE* 2019;38:134–44.
- [105] Pohlen T, Hermans A, Mathias M, Leibe B. Jaringan sisa resolusi penuh untuk segmentasi semantik dalam pemandangan jalanan. *Pengenalan Pola IEEE Conf Comput Vis* 2017:3309–18.
- [106] Qureshi R. Model prediksi respons obat yang dipersonalisasi untuk pasien kanker paru-paru menggunakan pembelajaran mesin. *TechRxiv* 2020; 13273319.v1.
- [107] Kapil A, Meier A, Zuraw A, Steele KE, Rebelatto MC, Schmidt G, dkk. Pembelajaran generatif semi diawasi secara mendalam untuk penilaian proporsi tumor otomatis pada biopsi jarum jaringan NSCLC. *Perwakilan Sains* 2018;8:17343.
- [108] Otsu N. Metode pemilihan ambang batas dari histogram tingkat abu-abu. *IEEE Trans Syst Man Cybern* 1979;9:62–6.
- [109] Geeleher P, Cox NJ, Huang RS. Respon obat klinis dapat diprediksi menggunakan tingkat ekspresi gen dasar dan secara in vitro sensitivitas obat dalam garis sel. *Biol Genom* 2014;15:R47.
- [110] Cule E, regresi De Iorio M. Ridge dalam masalah prediksi: pilihan otomatis parameter ridge. *Genet Epidemiol* 2013;37:704–14.
- [111] Quiros AC, Coudray N, Yeaton A, Yang X, Chiriboga L, Karimkhan A, dkk. Pembelajaran dengan pengawasan mandiri mengungkap kelompok morfologi di balik jenis dan prognosis kanker paru-paru. *arXiv* 2022;2205.01931.
- [112] Gensheimer MF, Aggarwal S, Benson KRK, Carter JN, Henry AS, Wood DJ, dkk. Model otomatis versus dokter yang merawat untuk memprediksi waktu kelangsungan hidup pasien dengan kanker metastatik. *J Am Med Informasikan Assoc* 2021;28:1108–16.
- [113] Doppalapudi S, Qiu RG, Badr Y. Prediksi dan pemahaman periode kelangsungan hidup kanker paru-paru: pendekatan pembelajaran mendalam. *Int J Med Informasikan* 2021;148:104371.
- [114] Nair M, Sandhu SS, Sharma AK. Biomarker prognostik dan prediktif pada kanker. *Target Obat Kanker Curr* 2014;14:477–504.
- [115] Chow E, Davis L, Panzarella T, Hayter C, Szumacher E, Loblaw A, dkk. Akurasi prediksi kelangsungan hidup oleh ahli onkologi radiasi paliatif. *Int J Radiat Oncol Biol Fisika* 2005;61:870–3.
- [116] Lakin JR, Robinson MG, Bernacki RE, Powers BW, Blok SD, Cunningham R, dkk. Memperkirakan angka kematian dalam 1 tahun untuk pasien layanan primer berisiko tinggi menggunakan pertanyaan "kejutan". *Kedokteran Magang JAMA* 2016;176:1863–5.
- [117] White N, Reid F, Harris A, Harries P, Stone P. Tinjauan sistematis terhadap prediksi kelangsungan hidup dalam perawatan paliatif: seberapa akurat dokter dan siapa ahlinya? *PLoS Satu* 2016;11: e0161407.
- [118] Cox DR. Model regresi dan tabel kehidupan. *JR Stat Soc B* 1972;34:187–220.
- [119] Wang X, Yao S, Xiao Z, Gong J, Liu Z, Han B, dkk. Pengembangan dan validasi model kelangsungan hidup adenokarsinoma paru berdasarkan gen terkait autophagy. *J Terjemahan Med* 2020;18:149.
- [120] Zhang YH, Lu Y, Lu H, Zhou YM. Pengembangan model prognostik kelangsungan hidup untuk kanker paru-paru non-sel kecil. *Oncol Depan* 2020;10:362.
- [121] Yu KH, Zhang C, Berry GJ, Altman RB, Re C, Rubin DL, dkk. Memprediksi prognosis kanker paru-paru non-sel kecil dengan fitur gambar patologi mikroskopis yang sepenuhnya otomatis. *Nat Commun* 2016;7:12474.
- [122] Hatlen P, Gronberg BH, Langhammer A, Carlsen SM, Amundsen T. Kelangsungan hidup yang berkepanjangan pada pasien kanker paru-paru dengan diabetes melitus. *J Thorac Oncol* 2011;6:1810–7.
- [123] Chen YC, Ke WC, Chiu HW. Klasifikasi risiko kelangsungan hidup kanker menggunakan ANN dengan data ekspresi gen dari berbagai laboratorium. *Hitung Biol Med* 2014;48:1–7.
- [124] Liu Y, Yang M, Sun W, Zhang M, Sun J, Wang W, dkk. Mengembangkan panel gen prognostik waktu bertahan hidup pada pasien adenokarsinoma paru menggunakan pembelajaran mesin. *Terjemahan Kanker Resolusi* 2020;9:3860–9.
- [125] Yu J, Hu Y, Xu Y, Wang J, Kuang J, Zhang W, dkk. LUADpp: model prediksi yang efektif pada prognosis adenokarsinoma paru berdasarkan fitur mutasi somatik. *Kanker BMC* 2019;19:263.
- [126] Cho HJ, Lee S, Ji YG, Lee DH. Asosiasi mutasi gen spesifik yang berasal dari pembelajaran mesin dengan kelangsungan hidup pada adenokarsinoma paru. *PLoS Satu* 2018;13:e0207204.
- [127] Hui X, Hu Y, Sun MA, Shu X, Han R, Ge Q, dkk. EBT: uji statistik yang mengidentifikasi fitur signifikan berukuran sedang dengan kekuatan dan presisi seimbang untuk perbandingan tingkat genom. *Bioinformatika* 2017;33:2631–41.

- [128] Yu KH, Berry GJ, Rubin DL, Re C, Altman RB, Snyder M. Asosiasi fitur omics dengan pola histopatologi pada adenokarsinoma paru. *Sistem Sel* 2017;5:620–7.
- [129] Ramazzotti D, Lal A, Wang B, Batzoglou S, Sidow A. Data tumor multiomik mengungkapkan keragaman mekanisme molekuler yang berkorelasi dengan kelangsungan hidup. *Nat Commun* 2018;9:4453.
- [130] Asada K, Kobayashi K, Joutard S, Tubaki M, Takahashi S, Takasawa K, dkk. Mengungkap gen dan jalur terkait prognosis dengan analisis multi-omics pada kanker paru-paru. *Biomolekul* 2020;10:524.
- [131] Takahashi S, Asada K, Takasawa K, Shimoyama R, Sakai A, Bolakan A, dkk. Memprediksi sub tipe kelangsungan hidup integrasi paralel multi-omics berbasis pembelajaran mendalam pada kanker paru-paru menggunakan data susunan protein fase terbalik. *Biomolekul* 2020;10:1460.
- [132] Xia L, Liu Y, Wang Y. Terapi blokade PD-1/PD-L1 pada kanker paru non-sel kecil stadium lanjut: status saat ini dan arah masa depan. *Ahli Onkologi* 2019;24:S31–41.
- [133] Doroshow DB, Sanmamed MF, Hastings K, Politi K, Rimm DL, Chen L, dkk. Imunoterapi pada kanker paru-paru non-sel kecil: fakta dan harapan. *Klinik Kanker Res* 2019;25:4592–602.
- [134] Lim SM, Hong MH, Kim HR. Imunoterapi untuk kanker paru-paru non-sel kecil: lanskap saat ini dan perspektif masa depan. *Jaringan Kekebalan Tubuh* 2020;20:e10.
- [135] Xu Z, Wang X, Zeng S, Ren X, Yan Y, Gong Z. Menerapkan kecerdasan buatan untuk imunoterapi kanker. *Acta Farmasi Dosa B* 2021;11:3393–405.
- [136] Wiesweg M, Mairinger F, Reis H, Goetz M, Kollmeier J, Misch D, dkk. Pembelajaran mesin mengungkapkan prediksi respons independen PD-L1 terhadap imunoterapi kanker paru-paru non-sel kecil berdasarkan konteks ekspresi gen. *Kanker Euro J* 2020;140:76–85.
- [137] Trebeschi S, Drago SG, Birnbak NJ, Kurilova I, Calin AM, Pizzi AD, dkk. Memprediksi respons terhadap imunoterapi kanker menggunakan biomarker radiomik noninvasif. *Ann Oncol* 2019;30:998–1004.
- [138] Coroller TP, Agrawal V, Narayan V, Hou Y, Grossmann P, Lee SW, dkk. Fitur fenotip radiomik memprediksi respon patologi pada kanker paru-paru non-sel kecil. *Radiother Oncol* 2016;119:480–6.
- [139] Tosolini M, Pont F, Poupot M, Vergez F, Nicolau-Travers ML, Vermijlen D, dkk. Penilaian TCRV yang menginfiltrasi tumor C9VD2 CDkelimpahan limfosit dengan dekonvolusi mikroarray kanker manusia. *Onkoimunologi* 2017;6:e1284723.
- [140] Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, dkk. Pencacahan subset sel yang kuat dari profil ekspresi jaringan. *Metode Nat* 2015;12:453–7.
- [141] Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, dkk. Pendekatan radiomik untuk menilai sel CD8 yang menginfiltrasi tumor dan respons terhadap imunoterapi anti-PD-1 atau anti-PD-L1: biomarker pencitraan, studi multikohort retrospektif. *Lancet Oncol* 2018;19:1180–91.
- [142] Saltz J, Gupta R, Hou L, Kurc T, Singh P, Nguyen V, dkk. Organisasi spasial dan korelasi molekuler limfosit yang menginfiltrasi tumor menggunakan pembelajaran mendalam pada gambar patologi. *Perwakilan Sel* 2018;23:181–93.
- [143] Hou L, Nguyen V, Kanevsky AB, Samaras D, Kurc TM, Zhao T, dkk. Autoencoder jarang untuk deteksi dan representasi nukleus tanpa pengawasan dalam gambar histopatologi. *Pengenalan Pola* 2019;86:188–200.
- [144] Xu Y, Jia ZP, Ai Y, Zhang F, Lai M, Chang EIC. Fitur aktivasi konvolusional mendalam untuk klasifikasi dan segmentasi gambar histopatologi tumor otak skala besar. *Proses Sinyal Spee Akust Conf IEEE Int* 2015:947–51.
- [145] Noh H, Hong S, Han B. Mempelajari jaringan dekonvolusi untuk segmentasi semantik. *IEEE Int Conf Comp Vis* 2015:1520–8.
- [146] De Mattos-Arruda L, Vazquez N, Finotello F, Lepore R, Porta E, Hundal J, dkk. Prediksi neoantigen dan perspektif komputasi terhadap manfaat klinis: Rekomendasi dari kelompok kerja pengobatan presisi ESMO. *Ann Oncol* 2020;31:978–90.
- [147] Roudko V, Greenbaum B, Bhardwaj N. Prediksi komputasi dan validasi neoantigen terkait tumor. *Imunol Depan* 2020;11:27.
- [148] Zhang Z, Lu M, Qin Y, Gao W, Tao L, Su W, dkk. Neoantigen: terobosan baru dalam imunoterapi tumor. *Imunol Depan* 2021;12:672356.
- [149] Hilf N, Kutruff-Coqui S, Frenzel K, Bukur V, Stevanovic S, Gouttefangeas C, dkk. Uji coba vaksinasi yang dipersonalisasi secara aktif untuk glioblastoma yang baru didiagnosis. *Alam* 2019;565:240–5.
- [150] Carreno BM, Magrini V, Becker-Hapak M, Kaabinejadian S, Hundal J, Petti AA, dkk. Vaksin sel dendritik meningkatkan luas dan keragaman sel T spesifik melanoma neoantigen. *Sains* 2015;348:803–8.
- [151] Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, dkk. Vaksin neoantigen pribadi imunogenik untuk pasien melanoma. *Alam* 2017;547:217–21.
- [152] Keskin DB, Anandappa AJ, Sun J, Tirosh I, Mathewson ND, Li S, dkk. Vaksin neoantigen menghasilkan respons sel T intratumoral dalam uji coba glioblastoma fase Ib. *Alam* 2019;565: 234–9.
- [153] Zhao W, Sher X. Membandingkan prediktor pengikatan peptida-MHC secara sistematis: dari epitop sintetis hingga epitop yang diproses secara alami. *Biol Komputasi PLoS* 2018;14:e1006457.
- [154] Racle J, Michaux J, Rockinger GA, Arnaud H, Bobisse S, Chong C, dkk. Prediksi kuat epitop HLA kelas II dengan dekonvolusi motif imunopeptidoma yang mendalam. *Nat Bioteknologi* 2019;37:1283–6.
- [155] Chen BB, Khodadoust MS, Olsson N, Wagar LE, Fast E, Liu CL, dkk. Memprediksi presentasi antigen HLA kelas II melalui deep learning terintegrasi. *Bioteknologi Alam* 2019;37:1332–43.
- [156] O'Donnell TJ, Rubinsteyn A, Bonsack M, Riemer AB, Laserson U, Hammerbacher J. MHCflurry: prediksi afinitas pengikatan MHC kelas I sumber terbuka. *Sistem Sel* 2018;7:129–32.
- [157] Bulik-Sullivan B, Busby J, Palmer CD, Davis MJ, Murphy T, Clark A, dkk. Pembelajaran mendalam menggunakan kumpulan data spektrometri massa peptida HLA tumor meningkatkan identifikasi neoantigen. *Bioteknologi Alam* 2019;37:55–63.
- [158] Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. NetMHC-3.0: prediksi akurat yang dapat diakses web tentang afinitas MHC kelas I manusia, tikus, dan monyet untuk peptida dengan panjang 8–11. *Asam Nukleat Res* 2008;36:W509–12.
- [159] Nielsen M, Lundegaard C, Wornig P, Lauemoller SL, Lamberth K, Buus S, dkk. Prediksi epitop sel T yang andal menggunakan jaringan saraf dengan representasi urutan baru. *Sains Protein* 2003;12:1007–17.
- [160] Buus S, Stryhn A, Winther K, Kirkby N, Pedersen LO. Interaksi reseptor-ligan diukur dengan teknik kromatografi kolom pinal yang ditingkatkan. metode pemisahan ukuran throughput yang tinggi dan efisiensi tinggi. *Biochim Biophys Acta* 1995;1243:453–60.
- [161] Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: Peningkatan prediksi interaksi peptida-MHC kelas I yang mengintegrasikan data afinitas pengikatan ligan dan peptida yang dielusi. *J Immunol* 2017;199:3360–8.
- [162] Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, dkk. NetMHCpan, metode prediksi kuantitatif pengikatan peptida pada protein lokus HLA-A dan -B dengan urutan yang diketahui. *PLoS Satu* 2007;2:e796.
- [163] Ye L, Creaney J, Redwood A, Robinson B. Lanskap neoantigen kanker paru-paru saat ini dan implikasinya terhadap terapi. *J Thorac Oncol* 2021;16:922–32.
- [164] Gong L, Dia R, Xu Y, Luo T, Jin K, Yuan W, dkk. Beban neoantigen sebagai penanda prognostik dan prediktif untuk kanker paru non sel kecil stadium II/III pada pasien Cina. *Kanker Thorac* 2021;12:2170–81.
- [165] Zhang W, Yin Q, Huang H, Lu J, Qin H, Chen S, dkk. Neoantigen pribadi dari pasien dengan NSCLC menginduksi respons antitumor yang efisien. *Oncol Depan* 2021;11:628456.

- [166] Zou L, Yu S, Meng T, Zhang Z, Liang X, Xie Y. Tinjauan teknis diagnosis kanker payudara mamografi berbasis jaringan saraf konvolusional. *Komputasi Metode Matematika Med* 2019;2019:6509357.
- [167] LeCun Y, Bengio Y, Hinton G. Pembelajaran mendalam. *Alam* 2015;521:436–44.
- [168] Mao C, Yao L, Luo Y. ImageGCN: jaringan konvolusional grafik gambar multi-relasional untuk identifikasi penyakit dengan rontgen dada. *Pencitraan Med Trans IEEE* 2022;41:1990–2003.
- [169] Mao C, Yao L, Pan Y, Zeng Z, Luo Y. Pengklasifikasi generatif mendalam untuk diagnosis penyakit toraks dengan gambar rontgen dada. *Prosiding (IEEE Int Conf Bioinformatics Biomed)* 2018;2018:1209–14.
- [170] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, dkk. Sebuah gambar bernilai 16 - 16 kata: transformator untuk pengenalan gambar dalam skala besar. *arXiv* 2020;2010.11929.
- [171] Khan S, Naseer M, Hayat M, Zamir SW, Khan F, Shah M. Transformers dalam visi: survei. *arXiv* 2022;2101.01169.
- [172] Transformator Boesch G. Vision (ViT) dalam pengenalan gambar - panduan 2022 [Internet]. <https://viso.ai/deep-learning/vision-transformervit/>.
- [173] Hie B, Zhong ED, Berger B, Bryson B. Mempelajari bahasa evolusi virus dan melarikan diri. *Sains* 2021;371:284–8.
- [174] Flage L, Brandvain Y, Schrider DR. Efektivitas jaringan saraf konvolusional yang tidak masuk akal dalam inferensi genetik populasi. *Mol Biol Evol* 2019;36:220–38.
- [175] Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Pemodelan generatif mendalam untuk transkriptomik sel tunggal. *Metode Nat* 2018;15:1053–8.
- [176] Hashim S, Ali M, Nandakumar K, Yaqub M. SubOmiEmbed: pembelajaran representasi data multi-omics yang diawasi sendiri untuk klasifikasi jenis kanker. *Biol Komputasi Bioinform Int Conf ke-10* 2022:66–72.
- [177] Lee J, Hyeon DY, Hwang D. Multiomik sel tunggal: teknologi dan metode analisis data. *Exp Mol Med* 2020;52: 1428–42.
- [178] Rao A, Barkley D, Franca GS, Yanai I. Menjelajahi arsitektur jaringan menggunakan transkriptomik spasial. *Alam* 2021;596:211–20.
- [179] Li Y, Liu Q, Zeng Z, Luo Y. Menggunakan model pengelompokan tanpa pengawasan untuk mendeteksi penyebaran awal SARS-CoV-2 di seluruh dunia. *Gen* 2022;13:648.
- [180] Tian T, Wan J, Song Q, Wei Z. Mengelompokkan data RNA-seq sel tunggal dengan pendekatan pembelajaran mendalam berbasis model. *Nat Mach Intell* 2019;1:191–8.
- [181] Hu J, Li X, Hu G, Lyu Y, Susztak K, Li M. Pembelajaran transfer berulang dengan jaringan saraf untuk pengelompokan dan klasifikasi tipe sel dalam analisis RNA-seq sel tunggal. *Nat Mach Intell* 2020;2:607–18.
- [182] Brbic M, Zitnik M, Wang S, Pisco AO, Altman RB, Darmanis S, dkk. MARS: menemukan tipe sel baru melalui eksperimen sel tunggal yang heterogen. *Metode Nat* 2020;17:1200–6.
- [183] Shen H, Li Y, Feng M, Shen X, Wu D, Zhang C, dkk. Miscell: pendekatan pembelajaran mandiri yang efisien untuk membedah transkriptom sel tunggal. *iScience* 2021;24:103200.
- [184] Song Q, Su J, Zhang W. scGCN adalah algoritma jaringan konvolusional grafik untuk transfer pengetahuan dalam omics sel tunggal. *Nat Commun* 2021;12:3826.
- [185] Wang T, Shao W, Huang Z, Tang H, Zhang J, Ding Z, dkk. Mogenet mengintegrasikan data multi-omics menggunakan jaringan konvolusional grafik yang memungkinkan klasifikasi pasien dan identifikasi biomarker. *Nat Commun* 2021;12:3445.
- [186] Wang Y, Zhang Z, Chai H, Yang Y. Analisis prognosis kanker multi-omics berdasarkan jaringan konvolusi grafik. *Prosiding (IEEE Int Conf Bioinformatics Biomed)* 2021;2021:1564–8.
- [187] Zhang YD, Dong Z, Wang SH, Yu X, Yao X, Zhou Q, dkk. Kemajuan dalam fusi data multimodal dalam neuroimaging: gambaran umum, tantangan, dan orientasi baru. *Inf Fusion* 2020;64:149–87.
- [188] Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck 3rd WM, dkk. Integrasi komprehensif data sel tunggal. *Sel* 2019;177:1888–902.
- [189] Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Integrasi data multiomics, interpretasi, dan penerapannya. *Bioinform Biol Wawasan* 2020;14:1177932219899051.
- [190] Welch JD, Kozareva V, Ferreira A, Vanderburg C, Martin C, Macosko EZ. Integrasi multi-omik sel tunggal membandingkan dan membedakan fitur identitas sel otak. *Sel* 2019;177:1873–87.
- [191] Luo Y, Eran A, Palmer N, Avillach P, Levy-Moonshine A, Szolovits P, dkk. Pendekatan pengobatan presisi multidimensi mengidentifikasi subtipe autisme yang ditandai dengan dislipidemia. *Nat Med* 2020;26:1375–9.
- [192] Diao L, Guo H, Zhou Y, He Y. Menjembatani kesenjangan antara keluaran: adaptasi domain untuk segmentasi IHC kanker paru-paru. *Proses Gambar IEEE Int Conf* 2021:6–10.
- [193] Lotfollahi M, Naghipourfar M, Luecken MD, Khajavi M, Buttner M, Wagenstetter M, dkk. Memetakan data sel tunggal ke atlas referensi dengan pembelajaran transfer. *Nat Bioteknologi* 2022;40:121–30.
- [194] Arrieta AB, Diaz-Rodriguez N, Ser JD, Benetot A, Tabik S, Barbado A, dkk. Kecerdasan buatan yang dapat dijelaskan (XAI): konsep, taksonomi, peluang dan tantangan menuju AI yang bertanggung jawab. *Dalam Penggabungan* 2020;58:82–115.
- [195] Salahuddin Z, Woodruff HC, Chatterjee A, Lambin P. Transparansi jaringan saraf dalam untuk analisis citra medis: tinjauan metode interpretabilitas. *Hitung Biol Med* 2021;140: 105111.
- [196] Antoniadis AM, Du Y, Guendouz Y, Wei L, Mazo C, Becker BA, dkk. Tantangan saat ini dan peluang masa depan untuk XAI dalam sistem pendukung keputusan klinis berbasis pembelajaran mesin: tinjauan sistematis. *Aplikasi Sains* 2021;11:5088.
- [197] Kourou K, Exarchos KP, Papaloukas C, Sakaloglou P, Exarchos T, Fotiadis DI. Pembelajaran mesin terapan dalam penelitian kanker: tinjauan sistematis untuk diagnosis, klasifikasi, dan prognosis pasien. *Struktur Komputasi Bioteknologi J* 2021;19:5546–55.
- [198] Maier-Hein L, Reinke A, Godau P, Tizabi MD, Christodoulou E, Glocker B, dkk. Metrik dimuat ulang: kendala dan rekomendasi untuk validasi analisis gambar. *arXiv* 2022;2206.01653.
- [199] Meropol NJ, Donegan J, Kaya AS. Kemajuan dalam penerapan algoritma pembelajaran mesin untuk penelitian dan perawatan kanker. *JAMA Netw Buka* 2021;4:e2116063.
- [200] Sheehan DF, Criss SD, Chen YF, Eckel A, Palazzo L, Tramontano AC, dkk. Biaya kanker paru-paru berdasarkan strategi pengobatan dan fase perawatan di antara pasien yang terdaftar di medicare. *Pengobatan Kanker* 2019;8:94–103.
- [201] Mao C, Yao L, Luo Y. MedGCN: rekomendasi pengobatan dan imputasi tes laboratorium melalui jaringan konvolusional grafik. *J Biomed Informasi* 2022;127:104000.
- [202] Zhu J, Wang J, Wang X, Gao M, Guo B, Gao M, dkk. Prediksi kemanjuran obat dari profil transkripsional dengan pembelajaran mendalam. *Nat Bioteknologi* 2021;39:1444–52.
- [203] Luo Y, Wunderink RG, Lloyd-Jones D. Pembelajaran mesin proaktif vs reaktif dalam layanan kesehatan: pelajaran dari pandemi COVID-19. *JAMA* 2022;327:623–4.
- [204] Anda HS, Gao CX, Wang HB, Luo SS, Chen SY, Dong YL, dkk. Kesesuaian rekomendasi pengobatan untuk kanker paru-paru non-sel kecil metastatik antara Watson untuk sistem onkologi dan tim medis. *Manajemen Kanker Res* 2020;12:1947–58.
- [205] Liu C, Liu X, Wu F, Xie M, Feng Y, Hu C. Menggunakan kecerdasan buatan (watson untuk onkologi) untuk rekomendasi pengobatan di antara pasien kanker paru-paru Tiongkok: studi kelayakan. *J Med Internet Res* 2018;20:e11087.