

Aplikasi Pembelajaran Mesin dalam Pengolahan Data Citra untuk Bidang Medis: Sebuah Kajian Pustaka

Stanislaus Jiwandana Pinasthika¹
Universitas Negeri Yogyakarta¹
E-mail: stanislausjp@uny.ac.id

ABSTRAK

Dalam dunia medis, penggunaan citra sangat penting untuk mengidentifikasi kemunculan dan menentukan tindakan pencegahan agar penyakit tidak bertambah parah. Citra medis berisi data atau fitur yang merepresentasikan kondisi kesehatan pasien. Menganalisa citra medis membutuhkan ketelitian supaya pengambilan keputusan terkait pencegahan suatu penyakit menjadi akurat dan tepat. Namun, mengamati citra medis tidak mudah karena terkadang fitur citra yang dicari tidak terlalu jelas atau sulit dibedakan dengan yang objek lainnya, misalnya noise di citra tersebut. Sehingga, proses analisa memakan waktu yang lama dan keputusan tindakan medis menjadi lama untuk diambil. Beberapa penelitian pembelajaran mesin untuk pengolahan citra medis digital telah dikembangkan untuk mempercepat proses analisa. Kajian pustaka ini hadir untuk meninjau beberapa penelitian-penelitian tersebut. Kajian ini berisi pendekatan apa yang sering digunakan, tujuan penggunaan yang sering dicapai, serta performa model pembelajaran mesin yang dicapai dalam penelitian-penelitian tersebut dengan cara membandingkan hasil dari 10 artikel ilmiah berdasarkan langkah penggunaan dan objek penelitiannya. Dalam kajian ini dapat disimpulkan bahwa pendekatan *Convolutional Neural Network* (CNN) dan U-Net beserta pendekatan turunannya menjadi pendekatan yang favorit untuk digunakan karena keandalannya dalam mengekstrak fitur-fitur citra. Penerapan pembelajaran mesin untuk data citra lebih sering digunakan untuk memberi segmentasi pada objek citra tertentu yang ingin dianalisa. Performa yang diperoleh masing-masing penelitian cukup variatif mengingat tujuan dan objek penelitian juga bervariasi.

Kata kunci: *citra medis, klasifikasi, deteksi objek, ekstraksi fitur, segmentasi*

ABSTRACT

Analyzing medical images is a crucial phase before diagnosing a patient and deciding further medical treatment for the patient. Medical images contain important features that represent a patient's health condition. The analysis requires thoroughness in order to increase medical decision-making accuracy. However, it is complicated since the features are uncertainly clear to be observed. For instance, the observed object often shared similar features with image noises so the decision will take too long to be decided. Various machine learning studies have been developed to accelerate the analysis process and increase the accuracy of diagnosis. We have reviewed previous studies related to the utilization of machine learning for medical purposes. In this study, we obtained ten studies and reviewed the machine learning approaches, development purposes, and performance of the developed model. We concluded the most favorite ML approaches are Convolutional Neural Network (CNN) and U-Net and the descendants of their robustness and reliability to overcome the required tasks. The most assigned task is medical object segmentation. Despite medical objects demanding a high standard of model performance, the obtained performance varies depending on the object and the chosen model.

Keywords: *medical images, classification, object detection, feature extraction, segmentation*

PENDAHULUAN

Dalam dunia medis, penggunaan citra sangat penting untuk mengidentifikasi kemunculan atau menentukan tindakan

pencegahan agar penyakit tidak bertambah parah (Afiahayati et al., 2020; I Md. Dendi Maysanjaya, 2020; Shamim et al., 2019). Citra medis merupakan data penting yang membawa representasi kondisi kesehatan

pasien secara keseluruhan. Sebagai contoh, citra comet assay yang diambil dari sel *buccal mucosa* mulut manusia bagian dalam merepresentasikan kerusakan sel yang terjadi di dalam mulut pasien sehingga dapat menjadi bahan pertimbangan dalam melakukan diagnosa terkait kesehatan mulut pasien (Afiahayati et al., 2020).

Umumnya, citra medis diambil menggunakan proses radiologi (Cai et al., 2020; I Md. Dendi Maysanjaya, 2020) atau melalui citra hasil pengamatan menggunakan mikroskop jenis khusus (Afiahayati et al., 2020). Namun, tidak menutup kemungkinan citra tersebut diambil menggunakan kamera biasa, karena ciri-ciri fisik dari gejala penyakit tertentu dapat terlihat jelas dengan mata, seperti pengamatan permukaan lidah untuk mendeteksi gejala kanker mulut (Shamim et al., 2019).

Analisa yang tepat dan akurat untuk mendeteksi gejala awal merupakan dasar yang tepat untuk mencegah berkembangnya penyakit ke tahapan yang lebih parah, seperti kanker dan tumor (Cai et al., 2020; Kurnianingsih et al., 2019; Shamim et al., 2019; Wang et al., 2020). Namun, karena keterbatasan fisik manusia dan objek pengamatan yang sulit dibedakan dengan objek lainnya di dalam citra medis, pengambilan tindakan medis selanjutnya menjadi kurang cepat dan tepat (Linfeng Jiang et al., 2023; Wang et al., 2020).

Untuk menjawab permasalahan tersebut, berbagai penelitian mengenai pengolahan citra digital sudah dikembangkan untuk membantu menganalisa citra medis. Dalam pengolahan citra digital, langkah ekstraksi fitur menjadi penting untuk mengambil informasi sebanyak-banyaknya dari dalam citra tersebut (Mutlag et al., 2020).

Selain itu, ekstraksi fitur berguna untuk meningkatkan kualitas citra dengan mengurangi dimensi citra dan mengurangi redundansi informasi (Kumar et al., 2021).

Konsep pemrosesan citra medis pada umumnya sama seperti pemrosesan citra digital biasa. Jenis fitur yang dapat diambil juga sama mulai dari fitur geometris, fitur statistik, fitur tekstur, dan fitur warna (Mutlag et al., 2020). Untuk jenis data medis, lebih sering menggunakan pengambilan fitur geometris, fitur tekstur, dan fitur warna (Afiahayati et al., 2020; Nikhil Mehandru et al., 2020). Proses pencarian fitur inilah yang diotomatisasi dalam pendekatan pembelajaran mesin.

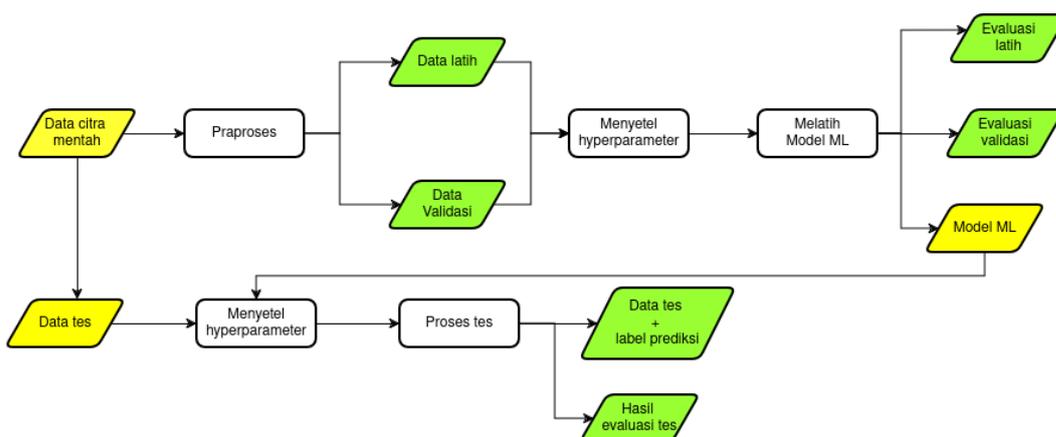
Sama seperti metode pengolahan citra digital biasa, metode pembelajaran mesin juga terdapat langkah untuk memproses fitur. Langkah ini bertujuan untuk memperjelas citra dan membersihkan citra dari *noise*, sehingga model pembelajaran mesin dapat mengambil fitur dan mengenali objek dengan mudah (Rana & Bhushan, 2023). Model yang pembelajaran mesin yang dikembangkan diharapkan mampu melakukan tugas-tugas seperti deteksi objek (Nikhil Mehandru et al., 2020; Shamim et al., 2019), segmentasi (Cai et al., 2020; Linfeng Jiang et al., 2023; Nogovitsyn et al., 2019; Pacheco et al., 2023; Robin Brügger et al., 2019; Vaidyanathan et al., 2021; Yan et al., 2019), klasifikasi (Afiahayati et al., 2020; I Md. Dendi Maysanjaya, 2020; Kurnianingsih et al., 2019), atau bahkan gabungan dari tugas-tugas tersebut untuk melakukan prediksi terkait kondisi kesehatan pasien (Wang et al., 2020).

Dalam pengembangan model pembelajaran mesin, dibutuhkan dua jenis data yang berbeda, yaitu data mentah yang sudah diketahui label serta letak dari objek

deteksinya dan data yang tidak diketahui untuk melakukan pengujian terhadap performa deteksi dan segmentasi yang dimiliki oleh model pembelajaran mesin tersebut (Gambar 1) (Cai et al., 2020; Nogovitsyn et al., 2019; Rana & Bhushan, 2023). Proses pelabelan tersebut bisa berupa pemberian *bounding box/ground-truth* per objek deteksi atau segmentasi manual yang menutupi objek deteksi dengan bidang transparan berukuran sama dengan objek dan dengan warna yang berbeda-beda sesuai dengan kelasnya.

Selain membersihkan dari noise, tindakan pra-proses dapat dilakukan tergantung kebutuhannya. Tindakan tersebut antara lain: mengubah citra ke citra biner atau *binary images* (Kurnianingsih et al., 2019), pengubahan ukuran atau *resize* (Kurnianingsih et al., 2019), mengubah jangkauan nilai intensitas piksel/normalisasi (Wang et al., 2020), atau interpolasi atau pengubahan resolusi citra sehingga citra tidak hanya selaras secara spasial, namun juga *voxel-wise* (Pacheco et al., 2023).

Penggunaan citra medis untuk penelitian sangat terbatas karena menyangkut privasi pasien dan biasanya hanya dapat diberikan dalam jumlah yang terbatas. Sementara, penelitian pembelajaran mesin membutuhkan data dalam jumlah besar untuk menghindari *overfitting* dan untuk meningkatkan keandalan model (Linfeng Jiang et al., 2023). Untuk mengatasi kondisi tersebut, dibutuhkan augmentasi data atau manipulasi letak atau dimensi citra untuk memperoleh seakan-akan data citra yang baru (Kurnianingsih et al., 2019). Contoh teknik augmentasi adalah rotasi, translasi atau menggeser citra ke arah tertentu, modifikasi ukuran atau *scale*, deformasi elastis, pembalikan citra atau mengubah dari sisi kanan ke sisi kiri dan sebaliknya yang disebut juga *mirroring* atau *flipping*, atau dapat juga dilakukan konversi warna dari RGB ke grayscale (Afiahayati et al., 2020; Cai et al., 2020; Kurnianingsih et al., 2019; Linfeng Jiang et al., 2023; Pacheco et al., 2023; Robin Brügger et al., 2019; Shamim et al., 2019; Wang et al., 2020).



Gambar 1. Alur pembelajaran mesin secara umum menggunakan data citra

Data yang telah melewati tindakan pra-proses dan dibuat augmentasinya telah menjadi satu dataset. Dataset ini kemudian

dibagi menjadi data latih, data validasi, dan data tes. Strategi pembagian dataset ini bertujuan untuk menghasilkan model yang

berkinerja baik (Xu & Goodacre, 2018). Dataset latih digunakan untuk melatih model pembelajaran mesin dan dievaluasi performanya menggunakan data validasi (Genç & Tunc, 2019). Proses latih akan diulang dengan parameter baru, apabila hasil proses latih yang sangat bagus berbeda jauh dengan hasil proses validasi atau yang disebut dengan *overfitting* (Genç & Tunc, 2019; Xu & Goodacre, 2018).

Jika hasil antara proses latih dan validasi sudah fit, maka langkah selanjutnya adalah proses tes menggunakan data tes (Genç & Tunc, 2019). Berbeda dari data latih dan data validasi, data tes merupakan data yang benar-benar baru dan asing bagi model pembelajaran mesin.

Idealnya, sebanyak 70% data dialokasikan untuk data latih dan masing-masing 15% untuk data validasi dan data tes (Genç & Tunc, 2019). Semakin banyak jumlah data dan semakin bervariasi kelas data yang dipelajari oleh model pembelajaran mesin, maka keandalan model akan semakin meningkat.

Salah satu metode paling umum untuk membagi data adalah *cross-validation* (CV). Dataset dibagi menjadi k bagian. Satu bagian dijadikan data validasi, sementara sebanyak $k-1$ digunakan sebagai data latih. Pelatihan model dilakukan sebanyak k kali dan model terbaik dihasilkan dengan memiliki error validasi terkecil.

Berbagai algoritma dan pendekatan telah dilakukan untuk melatih model pembelajaran mesin. Akhir-akhir ini, model *deep learning* sering digunakan untuk membuat model karena keandalannya dalam melakukan klasifikasi dan prediksi (Shamim et al., 2019). Untuk data citra, pendekatan *Convolutional Neural Network* (CNN) dan U-Net beserta masing-masing turunannya

menjadi pendekatan yang disukai karena memiliki arsitektur dapat dimodifikasi dan parameter yang dapat disetel sesuai dengan kebutuhan dan permasalahan yang dihadapi.

		Prediksi Kelas	
		Positif	Negatif
Kelas Aktual	Positif	TP	FN
	Negatif	FP	TN

(a)

		Prediksi Kelas			
		C1	C2	...	C _n
Kelas Aktual	C1	C1,1	FP	...	C1,n
	C2	FN	TP	...	FN

	C _n	C _n ,1	FP	...	C _n ,n

(b)

Gambar 2. Confusion matrix. (a) Untuk klasifikasi biner, (b) Untuk klasifikasi multi-kelas (Markoulidakis)

Tahap terakhir setelah model didapatkan adalah tahap evaluasi. Pada tahap ini, performa model akan diketahui melalui jumlah objek yang dideteksi dengan benar. Cara evaluasi yang paling digunakan adalah menggunakan *confusion matrix* (Gambar 2). Untuk sejumlah n kelas, maka dimensi *confusion matrix* yang digunakan adalah $n \times n$.

Komponen *confusion matrix* terdiri dari *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN). Pada Gambar 1b, ketika ingin mendeteksi objek

kelas C_2 dan model memberikan label C_2 secara tepat pada objek tersebut, maka hasil deteksi tersebut tergolong TP. Sementara, jika ada objek C_2 lain namun tidak terklasifikasi dalam C_2 maka termasuk FN. Jika ada objek non C_2 namun digolongkan ke dalam C_2 maka disebut sebagai FP. Sedangkan sisanya yang bukan objek C_2 dan tidak digolongkan ke dalam kelas C_2 adalah TN.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2)$$

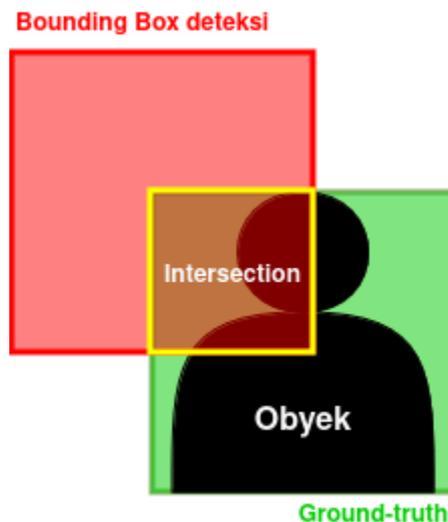
$$\text{Specificity} = \frac{TN}{TN+FP} \quad (3)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (4)$$

$$F1 - \text{Score} = \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (5)$$

Untuk klasifikasi, matriks yang biasa digunakan adalah matriks akurasi, matriks recall, matriks spesifisitas, matriks presisi dan matriks F1-Score. Matriks akurasi digunakan untuk klasifikasi biner, sementara matriks F1-Score digunakan untuk mengevaluasi performa model klasifikasi multi kelas (Markoulidakis et al., 2021).

Akurasi seperti terlihat di Persamaan 1, menghitung performa model dari total objek yang diprediksi secara benar dibanding dengan total objek telah diprediksi. Recall merupakan banyaknya objek yang diprediksi secara benar dibandingkan dengan seluruh objek yang menjadi anggota kelompok kelas tersebut. Sebaliknya, spesifisitas merupakan banyaknya objek yang tidak termasuk objek dalam penghitungan recall baik yang



Gambar 3. Intersection over Union (IoU) diprediksi secara benar maupun yang salah prediksi. Sementara presisi merupakan jumlah objek dalam kelas yang diprediksi secara tepat dibanding dengan seluruh prediksi yang telah dibuat baik yang tepat maupun yang keliru. Sementara F1-Score merupakan rata-rata harmonik dari presisi dan recall. Berdasarkan definisi tersebut, maka F1-Score lebih tepat digunakan untuk klasifikasi multi kelas karena dapat melihat sebaran data secara objektif. Akurasi kurang tepat digunakan karena melibatkan TN yang berdefinisi relatif tergantung kelas yang disorot (Grandini et al., 2020).

Untuk deteksi dan segmentasi, evaluasi dilakukan dengan matriks *average precision* (AP). Konsep ini menghitung semua presisi dan recall pada satu citra sebelum menghitung rata-rata dengan seluruh AP di seluruh dataset.

Setiap objek data latih label berupa *bounding box ground-truth* yang diberikan dengan bantuan ahli. Ketika dilakukan deteksi dengan model pembelajaran mesin, *bounding box* yang dihasilkan oleh model selalu tidak pas menindih pada *ground-truth* (Gambar 3). Oleh karena itu, dibutuhkan

batas toleransi terhadap hasil deteksi yang disebut dengan *Intersection over Union* (IoU). Interseksi adalah area tumpang tindih antara *ground-truth* dan *bounding box* prediksi, sedangkan *union* adalah area gabungan antara keduanya. Sehingga, IoU adalah luas area yang saling bertindih dibagi dengan luas total kedua area. Nilai IoU yang ditoleransi dan dianggap sebagai performa deteksi adalah 50%. Jika nilai IoU lebih dari batas maka dianggap TP, sebaliknya IoU yang kurang akan dianggap FP. Sementara *ground-truth* yang tidak dideteksi sama sekali akan masuk dalam kelompok FN. Selanjutnya akan dihitung presisi dan recall yang kemudian dipetakan pada kurva *precision-recall* yang berisi 11 poin. Nilai AP didapatkan dengan menjumlahkan kesebelas nilai presisi maksimal dan membaginya dengan sebelas. Semakin tinggi nilai AP, maka semakin baik performa deteksi atau segmentasi yang dilakukan oleh model.

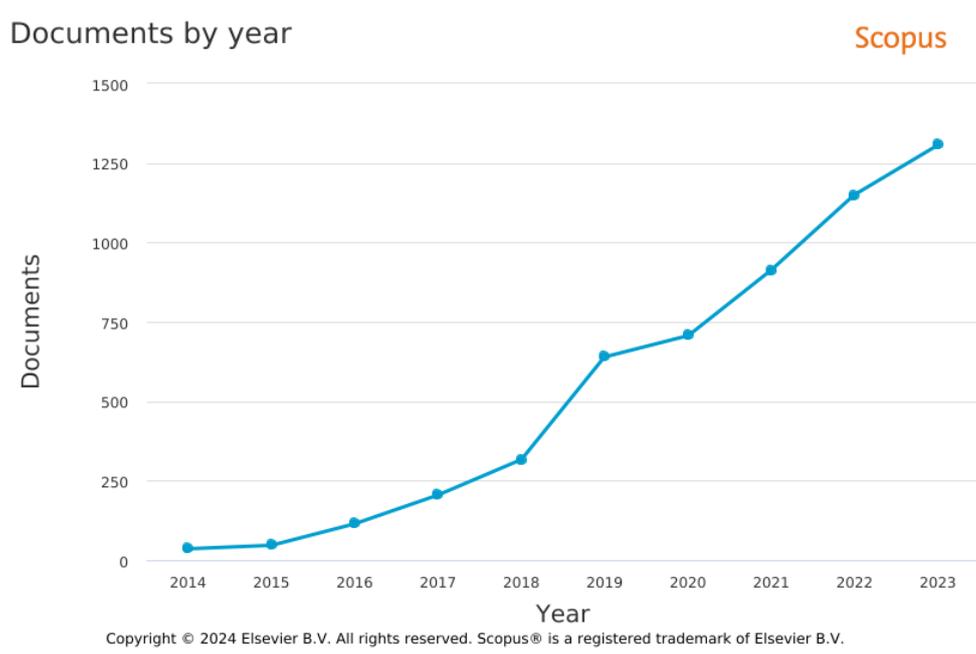
Penelitian ini bertujuan untuk mengkaji penelitian-penelitian terdahulu terkait pengembangan aplikasi pembelajaran mesin untuk pengolahan citra medis digital. Penelitian ini berisi paparan tentang pendekatan pembelajaran mesin yang sering diambil, tujuan pengembangan model, serta performa model tersebut.

ANALISIS PEMECAHAN MASALAH

Penelitian ini menggunakan jenis penelitian kajian pustaka. Penelitian yang dikaji adalah penelitian pengembangan model pembelajaran mesin dengan objek data citra medis.

Langkah pertama adalah melakukan pencarian artikel ilmiah yang tepat. Pencarian dilakukan dengan kata kunci "*Neural networks for medical image processing*" dengan membatasi tahun publikasi dari tahun 2014 hingga 2023. Selanjutnya konsentrasi bidang studi dibatasi pada ilmu komputer/*computer science* dan teknik/*engineering*. Batasan lain juga diberikan pada tipe dokumen, yaitu artikel publikasi ilmiah dan artikel konferensi ilmiah (Gambar 4). Kueri tersebut menghasilkan total 5.438 dokumen dengan 2.506 dokumen menggunakan *Convolutional Neural Network* (CNN) sebagai pendekatan pembelajaran mesin utama.

Dari total data yang diberikan, segmentasi citra menjadi tugas yang paling sering diselesaikan oleh model pembelajaran mesin secara spesifik dengan jumlah total 2.084 data. Berikutnya ada pemrosesan citra secara umum dengan 2.053 data. Sehingga, metode CNN dan turunannya dapat diambil sebagai topik kajian pustaka.



Gambar 4. Tren penelitian pembelajaran mesin untuk citra medis. Sumber: Scopus

Analisa citra medis tidak hanya menjadi ajang pengembangan penelitian antar bidang, namun juga sebagai ajang kompetisi. Dari 150 kompetisi analisa data citra biomedis, sebanyak 70% kompetisi memberikan tantangan kepada peserta untuk membuat segmentasi dan 10% kompetisi menantang peserta untuk membuat model klasifikasi (Maier-Hein et al., 2018). Organisasi yang banyak menyelenggarakan kompetisi segmentasi atau klasifikasi tersebut adalah *Medical Image Computing and Computer Assisted Interventions (MICCAI)* dan *IEEE International Symposium on Biomedical Imaging (ISBI)*.

Salah satu cabang kompetisi yang paling terkenal dari MICCAI adalah *Brain Tumor Segmentation (BraTS)*. Dalam kompetisi ini peserta diberikan kesempatan untuk menyelesaikan salah satu atau seluruhnya dari segmentasi tumor otak atau prediksi status promotor metilasi MGMT.

Untuk tantangan segmentasi tumor otak, hampir seluruh peserta mengembangkan model berbasis U-Net. Hal ini dikarenakan konsep *encode* dan *decode* dalam U-Net yang mampu menyimpan fitur penting dalam citra (Wang et al., 2020). Sehingga, metode yang dikaji di penelitian ini adalah turunan dari CNN dan U-Net.

Pencarian artikel ilmiah dilakukan secara acak melalui portal SCOPUS dan IEEE dengan mempertimbangkan adanya bahasan terkait pendekatan pembelajaran mesin yang sering digunakan, tujuan pengembangan model, serta performa model. Dari proses pencarian tersebut, didapatkan 10 artikel ilmiah yang terdiri dari 5 artikel membahas tentang pengembangan CNN dan 5 artikel membahas tentang penggunaan U-Net dan turunannya, dengan rentang waktu publikasi tahun 2019 hingga 2023. Informasi terkait pendekatan, sumber daya yang digunakan, jumlah sebaran data,

hingga faktor pengaruh artikel terhadap penelitian lain dapat dilihat di Tabel 1.

HASIL DAN PEMBAHASAN

A) Pendekatan Pembelajaran Mesin

Penggunaan pendekatan pembelajaran mesin yang digunakan tergantung pada tugas yang dikerjakan dan objek yang akan ditangani. Pada Tabel 1, CNN dengan arsitektur biasa dapat digunakan sebagai klasifikator, namun untuk tugas yang lebih rumit seperti memberikan segmentasi pada objek, peneliti harus mengembangkan arsitektur CNN atau mengimplementasikan model turunan dari CNN yang memiliki arsitektur yang lebih rumit seperti DCNN dan Mask R-CNN.

Dilihat dari objek penelitiannya (Tabel 1), semakin sulit citra tersebut diambil fiturnya, maka semakin rumit juga arsitektur model pembelajaran mesin yang dibutuhkan. Sebagai contoh, citra tumor otak cenderung dapat diatasi dengan pendekatan U-Net dan turunannya, sedangkan untuk citra MRI untuk objek lain cenderung menggunakan turunan dari CNN. Hal ini disebabkan karena

U-Net memiliki arsitektur yang lebih rumit daripada CNN dan karakter citra MRI tumor otak memiliki kontras yang cenderung rendah serta objek benjolan tumornya terlihat samar (Linfeng Jiang et al., 2023). Anatomi bagian dalam telinga juga lebih kecil dan lebih rumit, sehingga perlu diterapkan turunan dari U-Net yang mampu mengambil fitur citra dari objek organ bagian dalam telinga (Vaidyanathan et al., 2021).

Pemilihan model pengembangan mesin dapat juga disebabkan oleh faktor di luar objek deteksi, seperti jumlah dataset yang terbatas serta sumber daya yang terbatas. Sebagai contoh, model DCNN diterapkan untuk deteksi dan klasifikasi kondisi lidah karena dataset dikumpulkan secara mandiri oleh peneliti melalui citra foto biasa (Shamim et al., 2019).

Keterbatasan data dan sumber daya juga mendorong adanya cara baru dalam menerapkan model pembelajaran mesin. Sebagai contoh, teknik *transfer learning* atau menggunakan model yang sudah dilatihkan pada dataset yang sejenis sebelumnya, digunakan untuk melakukan klasifikasi citra comet assay (Afiahayati et al., 2020).

Tabel 1. Pendekatan, objek, dan sebaran data yang digunakan dalam aplikasi

Aplikasi	Pendekatan	Sumber data	Sebaran data	Pengaruh
Deteksi dan klasifikasi tonjolan dan kondisi lidah dari foto biasa. (Shamim et al., 2019)	DCNN	Pengumpulan data secara mandiri.	Tidak diketahui.	Disitasi 51 kali, diakses 541 kali.
Deteksi, segmentasi, visualisasi 3D bintil pada	Mask R-CNN	LUNA 16 dan Alibaba Cloud Tianchi Medical Competition	Tidak dijelaskan. LUNA 16: 888 kasus.	Disitasi 62 kali, diakses 4.073 kali.

Aplikasi	Pendekatan	Sumber data	Sebaran data	Pengaruh
paru-paru (Cai et al., 2020).			Alibaba Cloud Tianchi Medical Competition: 800 kasus.	
Segmentasi objek hippocampus dari citra MRI [12].	DCNN	Canadian Biomarker Integration Network in Depression (CAN-BIND) 1, 3, 4	Total: 452. CAN-BIND-1: 309 CAN-BIND-3: 99 CAN-BIND-4: 44	Disitasi 27 kali.
Segmentasi objek tumor otak dari citra MRI [13].	nnU-Net	BraTS challenge, subset data TCGA-LGG dan GBM.	Data latih: BraTS 2020: 202 BraTS-TCGA: 93 Data validasi: BraTS-TCGA: 74 Data tes: BraTS 2018: 76	Disitasi 5 kali, diakses 9 kali.
Segmentasi objek tumor otak dari citra MRI [14].	Partially Reversible U-Net	BraTS 2018.	Data latih: 80% Data validasi: 20%	Disitasi 42 kali, diakses oleh 10.000 dinasti.
Segmentasi dan klasifikasi sel servik pada citra pap smear [5].	Mask R-CNN untuk segmentasi, VGG-like Net untuk klasifikasi	Herlev University Hospital dan Technical University of Denmark.	Tidak dijelaskan. Total: 917 sel servik. Sel yang ditangkap oleh citra dikelompokkan menjadi 7 kelas. Tiga kelas normal berisi 242 sel, empat kelas berisi 675 sel.	Disitasi 82 kali, diakses 5.320 kali.
Segmentasi bintil pada paru-paru dari citra MRI [16].	Mask R-CNN	Lung Image Database Consortium – Infectious Disease Reasearce Institute (LIDC-IDRI)	Total: 6.889 Data latih: 5564 Data tes: 1325	Disitasi 6 kali, diakses 518 kali.
Segmentasi tumor otak dari citra mpMRI [6].	U-Net AT	BraTS 2019.	Data latih: 335 kasus. Data validasi: 125 kasus.	Disitasi 12 kali, diakses 1.637 kali.
Segmentasi objek tumor	RMAU-Net	Liver tumor segmentation challenge (LiTS) 2017 untuk	LiTS: Data latih: 131	Disitasi 4 kali, diakses 12 kali.

Aplikasi	Pendekatan	Sumber data	Sebaran data	Pengaruh
liver dari citra MRI [7].		data latih dan data tes dan 3d Image Reconstruction for Comparison of Algorithm Database (3D-IRCADb) untuk data validasi.	Data tes: 70 3D-IRCADb: tidak diketahui	
Segmentasi objek organ dalam telinga dari citra MRI [15].	3D U-Net	Maastricht Medical University center.	Data latih: 944 Data validasi: 99 Data tes: 177	Disitasi 35 kali, diakses 8.453 kali.
Klasifikasi objek pneumonia pada paru-paru dari citra MRI [1].	CNN	Kaggle	Data latih: Total: 5.216 Paru-paru normal: 1.341 paru-paru pneumonia: 3.875. Data tes: Total: 624 paru-paru normal: 234 paru-paru pneumonia: 390 .	Disitasi 6 kali.
Segmentasi dan klasifikasi objek comet assay dari citra mikroskop (Afiahayati et al., 2020).	Transfer learning berbasis CNN VGG16 dan VGG19.	Pengumpulan data secara mandiri.	Data latih: Tanpa augmentasi: 73 Data tes: Tanpa augmentasi: 61 Dengan augmentasi: 250	Diakses 811 orang kali, disitasi 1 kali.

B) Tujuan Pengembangan Model

Pengaplikasian pembelajaran mesin pada analisa citra medis bertujuan untuk membantu proses diagnosa dengan mendeteksi atau memberi segmentasi pada objek citra medis tertentu. Segmentasi dibagi menjadi dua, yaitu segmentasi 2D dan segmentasi 3D, tergantung karakteristik objek yang ingin disegmentasi.

Proses diagnosa dapat dilakukan dengan menganalisa hasil deteksi atau

segmentasi tersebut. Namun, hasil deteksi tersebut dapat digunakan sebagai dasar pengolahan data citra medis lanjutan agar proses diagnosa menjadi lebih mudah. Sebagai contoh, pengklasifikasian objek deteksi (Afiahayati et al., 2020; I Md. Dendi Maysanjaya, 2020; Kurnianingsih et al., 2019; Shamim et al., 2019), prediksi kondisi kesehatan pasien (Wang et al., 2020), atau melakukan visualisasi 3D berdasarkan hasil deteksi dan segmentasi untuk

mempertimbangkan tindakan medis selanjutnya (Cai et al., 2020).

Selain itu, tujuan pengembangan dapat didorong oleh faktor non-medis. Sebagai contoh pengembangan Partially Reversible U-Net untuk efisiensi penggunaan memori saat proses latih (Robin Brügger et al., 2019), pengembangan RMAU-Net agar dapat menutup area objek seluas mungkin dengan segmentasi (Linfeng Jiang et al., 2023), atau mengevaluasi manfaat dari penerapan salah satu teknik pra proses seperti *Brain Extraction* (BE) (Pacheco et al., 2023).

C) Performa Model

Pemilihan model pembelajaran mesin yang tepat dan pengenalan karakteristik objek yang diteliti menjadi kunci sukses dalam mencapai performa model yang diinginkan. Selain itu, penggunaan arsitektur yang tepat juga mempengaruhi performa tugas pembelajaran mesin.

Secara ideal, model pembelajaran mesin membutuhkan dataset latih dalam jumlah besar. Seperti contoh klasifikasi pneumonia yang dapat mencapai akurasi 89,58% setelah model dilatihkan pada 5.216 citra MRI paru-paru (I Md. Dendi Maysanjaya, 2020).

Memiliki karakteristik citra yang beragam, pengembangan model DCNN untuk kondisi lidah pasien menerapkan basis arsitektur yang berbeda (Shamim et al., 2019). Hasilnya DCNN dengan arsitektur VGG19 meraih akurasi sebesar 98% untuk klasifikasi kondisi lidah sehat dan kondisi prakanker. Sementara arsitektur ResNet50 yang digunakan untuk mengklasifikasikan lidah berdasarkan tipe permukaannya meraih akurasi tertinggi sebesar 97%.

Pengembangan model berbasis DCNN cukup baik dibandingkan dengan performa dari aplikasi segmentasi lain (Nogovitsyn et al., 2019). Pemilihan dataset untuk melakukan uji dan keberagaman karakter dataset latih membuat Hippodeep menuntaskan segmentasi hampir seluruh bagian *hemisphere* kanan dengan nilai Pearson correlation (r) sebesar 99%. Mengalahkan program FreeSurfer yang hanya mencapai 97%.

Penggunaan teknik penerapan model yang tepat dapat memberikan performa yang cukup memuaskan meskipun data yang dimiliki terbatas (Afiahayati et al., 2020). Terbukti, penerapan transfer learning CNN dengan arsitektur VGG16 mampu mencapai akurasi sebesar 70,5%.

Pendekatan CNN terus mengalami perkembangan sesuai dengan kebutuhan. Munculnya kebutuhan untuk bisa mendeteksi objek yang berbeda dan tumpang tindih serta memberi segmentasi yang presisi membuat Mask R-CNN muncul (He et al., 2018; Yan et al., 2019). Model Mask R-CNN dapat digunakan untuk jumlah dataset yang terbatas karena fitur RoI Align yang presisi dalam mendeteksi letak dan ukuran objek. Model ini berhasil mencapai nilai AP tertinggi sebesar 78% untuk batas IoU 50%, dan nilai AP 50% untuk batas IoU 70% (Yan et al., 2019).

Meskipun tujuan utamanya untuk deteksi dan segmentasi, Mask R-CNN pernah dimodifikasi untuk dapat menyelesaikan tugas klasifikasi sel servik (Kurnianingsih et al., 2019). Untuk klasifikasi biner, Mask R-CNN hanya menentukan sel tersebut tergolong normal atau abnormal dan mencapai nilai F1-Score 96,5% dan akurasi 98,1%.

Keakuratan segmentasi dari model Mask R-CNN dapat digunakan sebagai dasar bagi tugas lainnya seperti visualisasi 3D dari kondisi paru-paru (Cai et al., 2020). Keakuratan visualisasi benjolan paru-paru tersebut disebabkan oleh nilai AP performa segmentasinya yang mencapai 88,2% untuk batas IoU 50%.

Permasalahan yang ingin dipecahkan menggunakan metode U-Net adalah masalah segmentasi 3D atau objek bervolume dan ingin diamati perkembangannya untuk menentukan tindakan medis selanjutnya (Robin Brügger et al., 2019). Untuk mengevaluasi

performa segmentasi menggunakan U-Net, biasa digunakan matriks *Dice similarity coefficient* (DSC) dan jarak Hausdorff (HD95). Dice digunakan untuk membandingkan perpotongan dari segmentasi model (P) dan segmentasi *ground-truth* (T) terhadap jumlah positif P dan T. Sementara itu, jarak Hausdorff digunakan untuk mencari perbedaan terbesar di antara permukaan segmentasi model dan permukaan segmentasi *ground-truth*. DSC mencari nilai tertinggi, sementara Hausdorff mencari nilai terendah.

Table 2. Hasil penelitian menggunakan metode U-Net

Penelitian	Pendekatan	Objek	Dice	HD95
Brügger, Robin., et al. (Robin Brügger et al., 2019)	Partially Reversible U-Net	ET	0,80	2,58
		WT	0,91	4,58
		TC	0,86	6,84
Wang, Shuo., et al. (Wang et al., 2020)	U-Net AT	ET	0,75	4,99
		WT	0,90	4,70
		TC	0,81	7,11
Pacheco, Bruno Machado., et al. (Pacheco et al., 2023)	nnU-Net	HD-BET 2D	0,81	6,0
		HD-BET 3D	0,81	4,9
Jiang, Linfeng., et al. (Linfeng Jiang et al.,	RMAU-Net	Liver data et LiTS	0,96	-
		Tumor	0,76	-

2023)		data et LiTS		
		Liver data et 3D-IRCA Db	0,97	-
		Tumor data et 3D-IRCA Db	0,83	-
Vaidyanathan, Akshaya., et al. (Vaidyanathan et al., 2021)	3D U-Net	Organ dalam telinga	0,67	-

Bagian tumor, terutama untuk tumor otak yang perlu dideteksi adalah besar/*tumor core* (TC), *enhancing tumor* (ET), dan keseluruhan tumor/*whole tumor* (WT). Untuk organ yang berukuran kecil seperti organ dalam telinga perlu

dilakukan pencarian metode segmentasi yang sangat presisi.

Jika ingin melakukan segmentasi yang akurat namun tetap memperhatikan efisiensi memori, dapat digunakan Partially Reversible U-Net yang mampu menghemat memori hingga 50% untuk kebutuhan *encoder* (Robin Brügger et al., 2019). Dalam tabel 2, pencapaian terbaiknya dapat mencapai nilai Dice tertinggi 91,01% untuk WT dan nilai HD terendah untuk ET 2,58%, dengan menggunakan 3 *encoder*, 1 *decoder*, dengan arsitektur reversibel, menggunakan ukuran patch penuh, dan hanya mengonsumsi 9.620 MB memori.

Untuk memberikan prediksi yang akurat terkait keberlangsungan hidup pasien, 3D U-Net dimodifikasi dengan menambahkan *attention block* untuk menyimpan komponen segmentasi yang memiliki banyak informasi fitur (Wang et al., 2020). Pendekatan tersebut disebut dengan U-Net AT. Performa U-Net AT mencapai nilai Dice tertinggi 81% untuk segmentasi TC. Hasil segmentasi tersebut digunakan untuk mengambil bentuk, tekstur, serta lokasi tumor untuk digabung bersama biofisik menjadi dasar prediksi kelangsungan hidup pasien. Namun, sayangnya akurasi hanya mendapat skor 59% akibat kesalahan yang disebabkan oleh prediksi terhadap pasien yang dapat bertahan lebih dari 1.000 hari.

Sementara untuk objek yang memiliki ukuran dan bentuk variatif serta batasan yang kabur antara objek penyakit dan organnya, dapat digunakan metode Residual Multi-scale Attention U-Net (RMAU-Net) (Linfeng Jiang et al., 2023). Hal itu dikarenakan RMAU-Net terdiri dari dua modul Res-SE-Block dan Multi-scale Attention Block (MAB). Res-SE-Block mengatasi masalah hilang gradien dan meningkatkan kualitas representasi dengan menambahkan parameter di tiap *channel block convolution* (Hu et al., 2018). Sehingga, bagian yang paling tidak

jelas dari suatu tumor liver dapat tetap tersegmentasi dengan baik dan dapat dibuktikan dengan perolehan nilai Dice 95,52%.

Menggunakan nnU-Net, Pacheco dkk mencoba mengevaluasi pra-proses BE. Hasilnya, tanpa mengikutsertakan BE, nilai Dice hanya unggul pada segmentasi 3D senilai 81,5%. Sementara metode HD-BET unggul di nilai Dice untuk segmentasi 2D yaitu 80,8% dan mencapai nilai paling rendah dan yang paling baik untuk HD95 yaitu 4,896. Penelitian tersebut menyimpulkan bahwa pra proses BE tetap dibutuhkan, namun diperlukan pemilihan metode BE yang tepat.

Kelemahan U-Net dalam segmentasi adalah jika bertemu dengan objek dengan karakteristik berikut: anatomi situasi patologis yang variatif, terdapat struktur anatomi yang hilang, ukuran anatomi yang kecil, dan sudut pengambilan citra MRI yang tidak biasa (Vaidyanathan et al., 2021). Permasalahan tersebut dapat diatasi dengan 3D U-Net yang menambahkan beberapa gerbang *attention* untuk menyimpan informasi yang berguna pada segmentasi. Model yang dikembangkan cukup berhasil dengan memperoleh skor Dice 87,6%.

SIMPULAN

Penelitian ini mengkaji penggunaan pendekatan pembelajaran mesin untuk proses analisa objek dalam citra medis, tujuan penggunaan yang sering dicapai, serta performa model yang dicapai oleh penelitian-penelitian sebelumnya. Kajian ini dilakukan dengan membandingkan ketiga topik tersebut.

Metode CNN dan U-Net serta turunannya menjadi yang paling populer untuk digunakan. Model CNN sering digunakan karena keandalannya dalam

mengerjakan beberapa tugas seperti deteksi, segmentasi, dan klasifikasi secara sekaligus. Namun, penggunaan CNN membutuhkan sumber daya yang tidak sedikit sehingga dibutuhkan modifikasi terhadap parameter CNN atau metode lain yang memiliki keandalan yang sama dan mereduksi penggunaan sumber daya. Sementara model U-Net dan turunannya lebih sering digunakan untuk mendeteksi dan memberi segmentasi objek yang memiliki ukuran yang lebih kecil serta penampakan yang lebih lembut dan kabur dibandingkan dengan objek lain.

Tujuan penerapan pendekatan tersebut ada bermacam-macam, namun yang paling sering adalah untuk melakukan segmentasi terhadap objek di citra medis. Tujuan-tujuan lain dalam pengembangan model pembelajaran mesin adalah untuk mengklasifikasikan objek, memprediksi kondisi kesehatan pasien, menghemat penggunaan sumber daya, atau mengatasi objek dengan fitur citra yang sulit diambil.

Performa kedua pendekatan untuk membangun model cukup variatif tergantung objek dan tujuan yang dihadapi. Secara keseluruhan, performa model yang digunakan di setiap penelitian cukup memuaskan untuk menyelesaikan tugas yang diberikan.

Mengingat cepatnya perkembangan dalam dunia medis, maka pengembangan model pembelajaran mesin sangat dibutuhkan untuk memberikan layanan kesehatan yang optimal untuk umat manusia.

DAFTAR RUJUKAN

- Afiahayati, Anarossi, E., Yanuaryska, R. D., Nuha, F. U., & Mulyana, S. (2020). Comet Assay Classification for Buccal Mucosa's DNA Damage Measurement with Super Tiny Dataset Using Transfer Learning. *Intelligent Information and Database Systems: Recent Developments*.
- Cai, L., Long, T., Dai, Y., & Huang, Y. (2020). Mask R-CNN-Based Detection and Segmentation for Pulmonary Nodule 3D Visualization Diagnosis. *IEEE Access*, 8, 44400–44409. <https://doi.org/10.1109/ACCESS.2020.2976432>
- Genç, B., & Tunc, H. (2019). Optimal training and test sets design for machine learning. *TURKISH JOURNAL OF ELECTRICAL ENGINEERING & COMPUTER SCIENCES*, 27, 1534–1545. <https://doi.org/10.3906/elk-1807-212>
- Grandini, M., Bagli, E., & Visani, G. (2020). *Metrics for Multi-Class Classification: An Overview*. 1–17.
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2018). Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2), 386–397. <https://doi.org/10.1109/TPAMI.2018.2844175>
- Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-Excitation Networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7132–7141. <https://doi.org/10.1109/CVPR.2018.00745>
- I Md. Dendi Maysanjaya. (2020). Klasifikasi Pneumonia pada Citra X-rays Paru-paru dengan Convolutional Neural Network. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 9(2), 190–195. <https://doi.org/10.22146/jnteti.v9i2.66>
- Kumar, K. K., Chaduvula, K., & Rao, B. (2021). *A Detailed Survey On Feature Extraction Techniques In Image Processing For Medical Image Analysis*. <https://api.semanticscholar.org/CorpusID:236493971>
- Kurnianingsih, Allehaibi, K. H. S., Nugroho, L. E., Widyawan, Lazuardi, L., Prabuwo, A. S., & Mantoro, T. (2019). Segmentation and classification of cervical cells using

- deep learning. *IEEE Access*, 7, 116925–116941.
<https://doi.org/10.1109/ACCESS.2019.2936017>
- Lin Feng Jiang, Jiajie Ou, Ruihua Liu, Yangyang Zou, Ting Xie, Hanguang Xiao, & Ting Bai. (2023). RMAU-Net: Residual Multi-Scale Attention U-Net For liver and tumor segmentation in CT images. *Computers in Biology and Medicine*, 158, 1–11.
<https://doi.org/10.1016/j.compbiomed.2023.106838>
- Maier-Hein, L., Eisenmann, M., Reinke, A., Onogur, S., Stankovic, M., Scholz, P., Arbel, T., Bogunovic, H., Bradley, A. P., Carass, A., Feldmann, C., Frangi, A. F., Full, P. M., van Ginneken, B., Hanbury, A., Honauer, K., Kozubek, M., Landman, B. A., März, K., ... Kopp-Schneider, A. (2018). Why rankings of biomedical image analysis competitions should be interpreted with care. *Nature Communications*, 9(1), 5217.
<https://doi.org/10.1038/s41467-018-07619-7>
- Markoulidakis, I., Kopsiaftis, G., Rallis, I., & Georgoulas, I. (2021). Multi-Class Confusion Matrix Reduction method and its application on Net Promoter Score classification problem. *Proceedings of the 14th Pervasive Technologies Related to Assistive Environments Conference*, 412–419.
<https://doi.org/10.1145/3453892.3461323>
- Mutlag, W. K., Ali, S. K., Aydam, Z. M., & Taher, B. H. (2020). Feature Extraction Methods: A Review. *Journal of Physics: Conference Series*, 1591(1), 012028.
<https://doi.org/10.1088/1742-6596/1591/1/012028>
- Nikhil Mehandru, Wesley L. Hicks Jr, Anurag K. Singh, & Michael R. Markiewicz. (2020). Machine Learning for Identification of Craniomaxillofacial Radiographic Lesions. *American Association of Oral and Maxillofacial Surgeons*.
<https://doi.org/10.1016/j.joms.2020.07.010>
- Nogovitsyn, N., Souza, R., Muller, M., Srajer, A., Hassel, S., Arnott, S. R., Davis, A. D., Hall, G. B., Harris, J. K., Zamyadi, M., Metzack, P. D., Ismail, Z., Bray, S. L., Lebel, C., Addington, J. M., Milev, R., Harkness, K. L., Frey, B. N., Lam, R. W., ... MacQueen, G. M. (2019). Testing a deep convolutional neural network for automated hippocampus segmentation in a longitudinal sample of healthy participants. *NeuroImage*, 197, 589–597.
<https://doi.org/10.1016/j.neuroimage.2019.05.017>
- Pacheco, B. M., Cassia, G. de S. e, & Silva, D. (2023). Towards fully automated deep-learning-based brain tumor segmentation: Is brain extraction still necessary? *Biomedical Signal Processing and Control*, 82, 104514.
<https://doi.org/10.1016/j.bspc.2022.104514>
- Rana, M., & Bhushan, M. (2023). Machine learning and deep learning approach for medical image analysis: Diagnosis to detection. *Multimedia Tools and Applications*, 82(17), 26731–26769.
<https://doi.org/10.1007/s11042-022-14305-w>
- Robin Brügger, Christian F. Baumgartner, & Ender Konukoglu. (2019). A Partially Reversible U-Net for Memory-Efficient Volumetric Image Segmentation. *Medical Image Computing and Computer Assisted Intervention*, 429–437.
- Shamim, M., Syed, S., Shiblee, M., Usman, M., & Ali, S. (2019). Automated detection of oral pre-cancerous tongue lesions using deep learning for early diagnosis of oral cavity cancer. *arXiv*, 1–25.
<https://doi.org/10.1093/comjnl/bxaa136>

- Vaidyanathan, A., van der Lubbe, M. F. J. A., Leijenaar, R. T. H., van Hoof, M., Zerka, F., Miraglio, B., Primakov, S., Postma, A. A., Brintjes, T. D., Bilderbeek, M. A. L., Sebastiaan, H., Dammeijer, P. F. M., van Rompaey, V., Woodruff, H. C., Vos, W., Walsh, S., van de Berg, R., & Lambin, P. (2021). Deep learning for the fully automated segmentation of the inner ear on MRI. *Scientific Reports*, *11*(1), 2885. <https://doi.org/10.1038/s41598-021-82289-y>
- Wang, S., Dai, C., Mo, Y., Angelini, E., Guo, Y., & Bai, W. (2020). Automatic Brain Tumour Segmentation and Biophysics-Guided Survival Prediction. In A. Crimi & S. Bakas (Eds.), *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (pp. 61–72). Springer International Publishing.
- Xu, Y., & Goodacre, R. (2018). On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. *Journal of Analysis and Testing*, *2*(3), 249–262. <https://doi.org/10.1007/s41664-018-0068-2>
- Yan, H., Lu, H., Ye, M., Yan, K., Xu, Y., & Jin, Q. (2019). Improved mask R-CNN for lung nodule segmentation. *Proceedings - 10th International Conference on Information Technology in Medicine and Education, ITME 2019*, 137–141. <https://doi.org/10.1109/ITME.2019.00041>