

Implementasi *Data Science* dalam Ranah Kesehatan: Pendeteksi Kanker Paru-paru Berdasarkan Analisis Citra *CT-scan*

Kanker, salah satu penyakit paling ditakuti oleh orang-orang, merupakan kondisi di mana sel abnormal pada tubuh seseorang terus-menerus tumbuh tanpa terkontrol. Sel tua yang seharusnya mati malah tumbuh dan memunculkan sel-sel abnormal lainnya. Dari semua jenis kanker yang meneror orang-orang, jenis kanker yang paling banyak menelan korban jiwa adalah kanker paru-paru.

Seperti namanya, kanker paru-paru adalah jenis kanker yang tumbuh dalam paru-paru manusia. Kanker ini bisa menjangkit siapa saja, tetapi perokok, baik yang aktif maupun pasif, merupakan golongan orang yang paling rentan terhadap penyakit ini.

Gejala-gejala dari penyakit ini meliputi batuk yang tidak kunjung berhenti, batuk berdarah, sesak nafas, sakit pada bagian dada, tenggorokan kering, penurunan berat badan yang tidak wajar, nyeri pada tulang, dan sakit kepala. Jika Anda mengalami gejala-gejala ini, Anda disarankan untuk konsultasi dengan dokter sesegera mungkin.

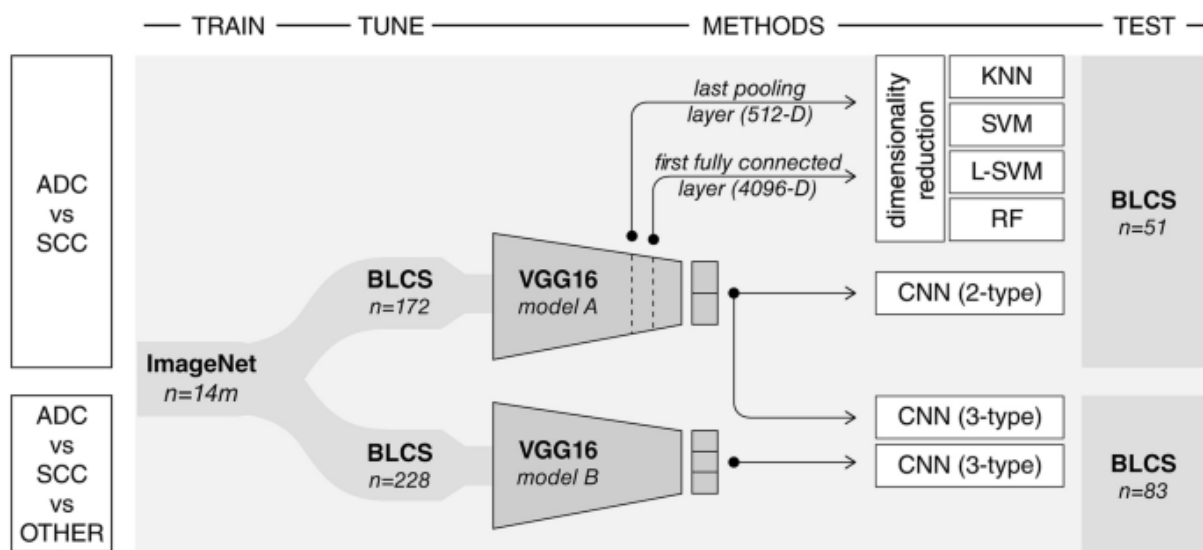
Sebenarnya kanker paru-paru sudah bisa ditangani berkat perkembangan teknologi jika penyakit ini dapat terdeteksi di masa awal pertumbuhannya. Pada umumnya, pasien yang memiliki gejala-gejala kanker paru-paru akan melalui beberapa tes dan prosedur yang meliputi *X-ray*, *CT scan*, *Magnetic Resonance Imaging* (MRI), *Positron Emission Tomography* (PET), *Sputum Cytology*, dan *Biopsy*. Yang menjadi masalah adalah semua tes dan prosedur ini masih bergantung pada kejelian dokter yang menganalisis citra yang dihasilkan.

Manusia adalah makhluk yang penuh dengan kekurangan. Ada faktor-faktor internal yang bisa mengakibatkan terjadinya error dalam menganalisis data dalam citra, khususnya jika citra yang dianalisis berukuran kecil atau tertutup obyek lainnya. Di sisi lain komputer akan menghasilkan prediksi dan analisis dengan presisi yang lebih tinggi dibandingkan manusia. Dengan alasan inilah para *data scientist* berlomba-lomba melatih komputer untuk mendeteksi kanker paru-paru dengan akurasi tinggi untuk membantu dokter dalam menangani kanker paru-paru sejak dini.

Radiomics muncul sebagai alat kuantifikasi fenotip tumor solid melalui ekstraksi fitur kuantitatif radiografi (Chaunzwa et al., 2021). Meskipun sebelumnya *radiomics* bergantung pada ekstraksi eksplisit dari fitur citra yang disesuaikan langsung oleh manusia, penelitian-penelitian terkini terkait topik ini sudah mulai berpindah menggunakan *deep learning*,

khususnya *Convolutional Neural Network* (CNN), di mana fitur representatif dipelajari langsung dari data citra. Penelitian terkini bahkan menghasilkan sensitifitas lebih dari 99% dalam *screening* benjolan paru-paru menggunakan *CT-scan* (Xu et al., 2019).

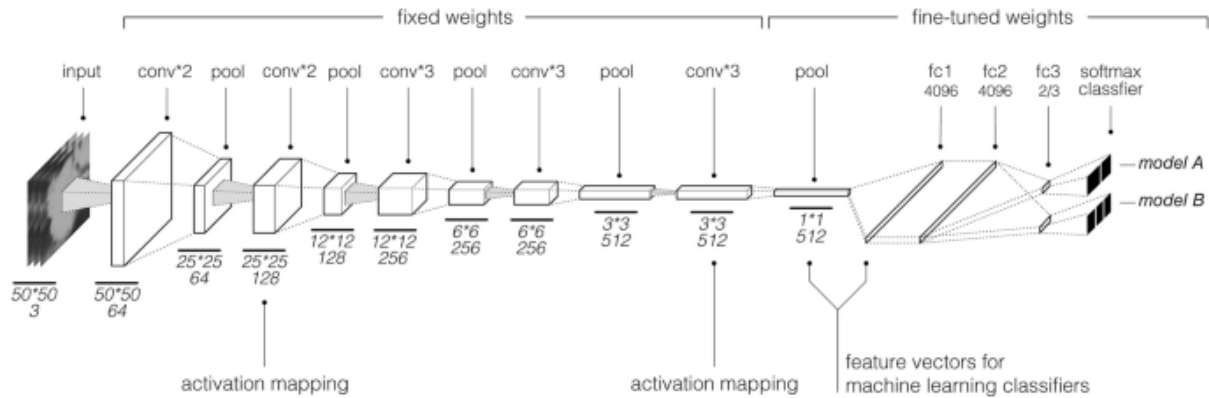
Berikut ini merupakan desain eksperimental dari *neural network* untuk mengklasifikasi histologi kanker paru-paru menggunakan citra *CT-scan*. *Network* ini tidak hanya mendeteksi adanya kanker paru-paru, tapi juga mengklasifikasikan kanker tersebut.



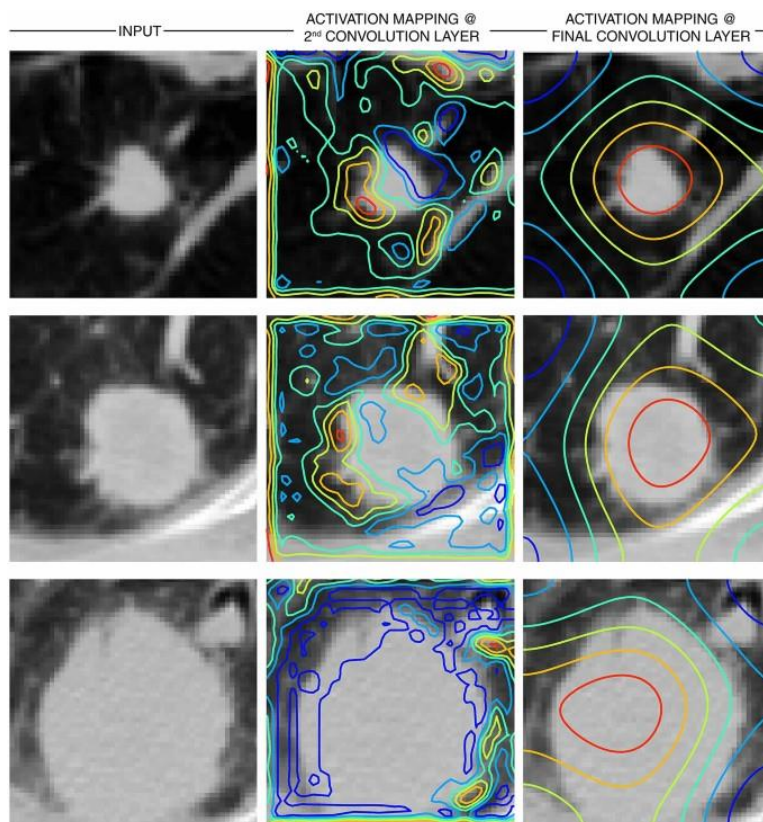
Dalam desain di atas, Sejenis CNN (VGG16) dilatih menggunakan dataset *ImageNet* berukuran besar yang memiliki lebih dari 14 juta citra. Untuk model A VGG16 disesuaikan menggunakan 172 sampel pasien dengan *adenocarcinoma* atau *squamous cell carcinoma* yang kemudian akan digunakan untuk memprediksi kasus kanker dengan tipe histologi ini menggunakan *test set* berupa 51 pasien dengan dua tipe histologi tersebut. Model ini digunakan sebagai ekstraktor fitur untuk menilai klasifikasi menggunakan *machine learning* berupa KNN, SVM, L-SVM, dan RF. Fitur kuantitatif radiografi didapatkan dari *last pooling* dan *first fully connected layers* dengan vektor masing-masing 512-D dan 4096-D. Model ini juga digunakan sebagai pengklasifikasi probabilistik dari histologi dan dites menggunakan *test set* berupa 83 kasus yang terdiri dari 3 tipe histologi *adenocarcinoma*, *squamous cell carcinoma*, dan “lainnya”. Untuk model B VGG16 disesuaikan menggunakan sampel heterogen dari 228 kasus yang terdiri dari 3 tipe histologi tersebut dan dites pada 83 sampel pasien. Data yang digunakan dibangun dari data yang terkumpul melalui *Boston Lung Cancer Survival* (BLCS) yang menghasilkan dataset berupa 311 pasien dengan *non-small cell lung cancer* (NSCLC) pada

tahap awal yang mendapatkan perawatan di *Massachusetts General Hospital* (MGH) dari 1999-2011.

Berikut ini adalah skematik untuk kedua model.



Berikut ini merupakan visualisasi dari cara kerja *network* pada model A.



Akhir kata, *data science* bisa diimplementasikan pada berbagai bidang, termasuk bidang kesehatan. Dengan adanya bantuan dari komputer yang dilatih berkat perkembangan *data science*, diharapkan di masa depan kanker paru-paru dapat terdeteksi sejak dini sehingga dapat ditangani dengan maksimal.

Daftar Pustaka

- Chaunzwa, T. L., Hosny, A., Xu, Y., Shafer, A., Diao, N., Lanuti, M., Christiani, D. C., Mak, R. H., & Aerts, H. J. W. L. (2021). Deep learning classification of lung cancer histology using CT images. *Scientific Reports*, 11(1), Article 1.
<https://doi.org/10.1038/s41598-021-84630-x>
- Xu, Y., Hosny, A., Zeleznik, R., Parmar, C., Coroller, T., Franco, I., Mak, R. H., & Aerts, H. J. W. L. (2019). Deep Learning Predicts Lung Cancer Treatment Response from Serial Medical Imaging. *Clinical Cancer Research*, 25(11), 3266–3275.
<https://doi.org/10.1158/1078-0432.CCR-18-2495>

Nama: Dwi Rangga Rhaditya Siwi Widodo

LinkedIn: <https://www.linkedin.com/in/dwi-rangga-rhaditya-siwi-widodo-5456681b7>

GitHub: <https://github.com/RanggaWidodo>

Portfolio Webpage: <https://ranggawidodo.github.io/devportfolio/>