Implementasi Attention Augmented Convolutional Networks untuk Deteksi Penyakit Covid-19 dari Gambar X-Ray Dada

Farid Alvin Fadillah
Fakultas Informatika, Universitas Telkom, Bandung
faridalvin@student.telkomuniversity.ac.id

Tjokorda Agung Budi Wirayuda
Fakultas Informatika, Universitas Telkom, Bandung
cokagung@telkomuniversity.ac.id

Abstrak

Dalam beberapa tahun terakhir, penyakit COVID-19, yang disebabkan oleh virus corona, telah menjadi masalah kesehatan yang menantang bagi masyarakat. Metode deteksi COVID-19 dengan bantuan komputer saat ini menghadapi kesulitan untuk membedakan antara COVID-19 dan pneumonia karena keduanya memiliki gejala yang sama. Karena itu, ada kebutuhan untuk mengotomatisasi dan membantu proses pendeteksian. Kemajuan terbaru dalam deep learning telah membuka jalan baru untuk memanfaatkan data pencitraan medis untuk deteksi penyakit secara otomatis. Penelitian menghasilkan peforma Attention Augmented Convolutional Networks (AACNs) yang sedikit lebih bagus (~1%), namun dinilai kurang optimal dibandingkan dengan ResNet50 sebagai baseline. Dikarenakan adanya artefak dan beragam variasi pada data yang dapat mempengaruhi mekanisme attention.

Kata kunci: covid-19, image classification, convolutional networks.

Abstract

In recent years, COVID-19, caused by the coronavirus, has become a challenging health issue for society. Current computer-assisted COVID-19 detection methods face difficulties in distinguishing between COVID-19 and pneumonia because they share similar symptoms. Therefore, there is a need to automate and assist the detection process. Recent advances in deep learning have paved the way for utilizing medical imaging data for automatic disease detection. The research produced Attention Augmented Convolutional Networks (AACNs) with slightly better performance (~1%), but they were considered less than optimal compared to ResNet50 as a baseline. This is due to artifacts and variations in the data that can affect the attention mechanism.

Keywords: covid-19, image classification, convolutional networks

1. Pendahuluan

Latar Belakang

Covid-19, atau Penyakit Coronavirus 2019, merupakan penyakit yang mudah menular karena virus SARS-CoV-2. Virus ini pertama kali ditemukan di Wuhan, China, pada akhir tahun 2019. Penyakit ini cepat menyebar ke berbagai negara, menimbulkan pandemi di seluruh dunia yang memberikan dampak besar pada kesehatan masyarakat, perekonomian, dan rutinitas sehari-hari [4]. Kematian dari pneumonia karena virus SARS-CoV-2 menjadi semakin meningkat [5].

Salah satu langkah krusial dalam menanggulangi penyebaran penyakit ini adalah *screening* pasien yang efektif dan cepat. Saat ini, metode standar emas untuk mendeteksi COVID-19 adalah tes *Reverse Transcription Polymerase Chain Reaction* (RT-PCR). Walaupun begitu, metode RT-PCR ini memiliki keterbatasan signifikan, yaitu proses manual yang rumit, memakan waktu lama, serta sensitivitas yang bervariasi tergantung pada waktu timbulnya gejala dan metode pengambilan sampel [3].

Salah satu metode pencitraan yang sangat penting dan juga paling umum digunakan untuk mendiagnosis pneumonia di seluruh dunia, terutama selama pandemi COVID-19, adalah X-ray dada, yang sering disebut sebagai rontgen dada [6]. Namun, membuat diagnosis yang tepat dan akurat dari gambar rontgen membutuhkan pengetahuan ahli dan pengalaman yang khusus [6]. Mendiagnosis menggunakan X-ray dada Jauh lebih sulit daripada modalitas pencitraan lainnya seperti CT atau MRI [5].

Kemajuan teknologi deep learning, terutama *Convolutional Neural Networks* (CNN), telah memberikan kesempatan baru dalam pengembangan sistem diagnosis pneumonia berbasis komputer. Penggunaan CNN memungkinkan analisis gambar medis secara lebih cepat, akurat, dan konsisten, yang dapat membantu para ahli dalam menetapkan diagnosis penyakit dari gambar X-ray dada [2][13].

Untuk mengatasi tantangan tersebut, teknologi Deep Learning, khususnya CNN, telah menjadi paradigma utama dalam analisis citra medis otomatis,. Salah satu arsitektur CNN yang dirancang khusus untuk deteksi COVID-19 adalah COVID-Net. Dikembangkan oleh Wang et al., 2020 [4], COVID-Net memanfaatkan desain *projection-expansion-projection-extension* (PEPX) yang ringan untuk mencapai akurasi tinggi pada dataset COVIDx. COVID-Net telah menjadi baseline penting karena sifatnya yang *open-source* dan kemampuannya dalam memberikan performa diagnostik yang menjanjikan.

Meskipun CNN seperti COVID-Net telah menunjukkan hasil yang baik, arsitektur berbasis konvolusi memiliki kelemahan fundamental. Operasi konvolusi standar bekerja pada lingkungan lokal (*local neighborhood*), yang dibatasi oleh ukuran *receptive field* kernelnya. Hal ini menyebabkan CNN kesulitan dalam menangkap informasi global atau dependensi jarak jauh (*long-range interactions*) yang terdapat dalam sebuah citra [1]. Padahal, dalam citra medis seperti X-Ray dada, hubungan antar fitur visual yang berjauhan mungkin mengandung informasi penting untuk membedakan patologi penyakit yang kompleks.

Di sisi lain, mekanisme *self-attention*, yang awalnya populer dalam pemrosesan bahasa alami (NLP) dengan arsitektur Transformer [7], telah terbukti mampu menangkap dependensi global tanpa mempedulikan jarak antar elemen dalam input. Studi oleh Rao et al. [12], menunjukkan bahwa CNN standar seringkali menganalisis fitur secara general dan kurang fokus. Penambahan mekanisme self-attention terbukti membantu model medis untuk fokus pada regio klinis yang relevan, meningkatkan akurasi dan interpretabilitas.

Bello et al., 2019 [1] memperkenalkan *Attention Augmented Convolutional Networks* (AACN), sebuah metode yang menggabungkan kekuatan ekstraksi fitur lokal dari konvolusi dengan kemampuan pemahaman konteks global dari self-

attention. AACN bekerja dengan memadukan (*concatenating*) *feature maps* hasil konvolusi dengan feature maps yang dihasilkan melalui mekanisme relative self-attention 2D [1]. Pendekatan hibrida ini terbukti meningkatkan akurasi klasifikasi pada dataset standar seperti ImageNet dan COCO dengan penambahan parameter yang minimal dibandingkan model CNN murni seperti ResNet.

Dalam penelitian ini, akan dilakukan uji coba arsitektur AACN yang memanfaatkan self-attention pada bagian konvolusional dari jaringan. Pendekatan ini diharapkan dapat memberikan informasi lebih lanjut terhadap efisiensi model dalam deteksi penyakit Covid-19, sekaligus membandingkan hasilnya dengan model lainnya.

2. Studi Terkait

López-Cabrera et al., 2021 [3] melakukan tinjauan komprehensif mengenai keterbatasan penggunaan kecerdasan buatan dalam klasifikasi COVID-19 dari citra chest X-ray (CXR). Studi ini menyoroti bahwa meskipun banyak model melaporkan tingkat akurasi yang sangat tinggi, sebagian besar model tersebut mengalami *overfitting* karena bias pada dataset, di mana gambar positif dan negatif sering kali berasal dari sumber yang berbeda dengan karakteristik akuisisi yang berlainan. Kelebihan dari tinjauan ini adalah analisis kritisnya terhadap protokol evaluasi yang ada, yang menunjukkan bahwa model sering kali mempelajari artefak sumber (seperti label teks atau parameter mesin) daripada patologi penyakit itu sendiri, sehingga generalisasi model pada pengaturan klinis nyata menjadi rendah.

He et al., 2016 [11] memperkenalkan kerangka kerja *Deep Residual Learning* untuk mengatasi masalah penurunan akurasi pelatihan pada jaringan saraf yang sangat mendalam. Metode ini bekerja dengan mereformulasi lapisan untuk mempelajari fungsi residual $F(x)=H(x)-x$ melalui penggunaan shortcut connections (pemetaan identitas) yang memungkinkan aliran gradien lebih lancar selama propagasi balik. Secara khusus, ResNet50 menggunakan desain blok bottleneck (terdiri dari tiga lapisan konvolusi: 1×1, 3×3, dan 1×1) untuk menjaga efisiensi komputasi. Hasil eksperimen menunjukkan bahwa arsitektur residual memenangkan kompetisi ILSVRC 2015 dengan tingkat kesalahan top-5 sebesar 3,57% pada dataset ImageNet, membuktikan bahwa jaringan yang lebih dalam dapat dilatih secara efektif dengan kompleksitas yang lebih rendah dibandingkan model VGG.

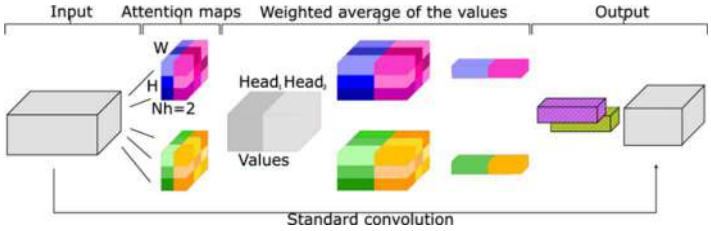
Bello et al., 2019 [1] mengusulkan arsitektur *Attention Augmented Convolutional Networks* (AACN) untuk menangkap konteks global yang sering terlewatkan oleh operasi konvolusi standar. Metode ini bekerja dengan menggabungkan convolutional feature maps (untuk lokalitas) dengan self-attentional feature maps (untuk dependensi jarak jauh) melalui konkatenasi, serta menggunakan mekanisme 2D relative self-attention untuk mempertahankan ekuivariansi translasi. Hasil eksperimen menunjukkan bahwa AACN mencapai peningkatan akurasi top-1 sebesar 1,3% pada dataset ImageNet dibandingkan baseline ResNet50 dan peningkatan 1,4 mAP pada deteksi objek COCO, serta mengungguli metode atensi lain seperti Squeeze-and-Excitation (SE) dengan jumlah parameter yang sebanding.

Wang et al., 2020 [4] mengembangkan COVID-Net, sebuah arsitektur CNN yang dirancang dengan khusus untuk mengenali dan mendeteksi penyakit COVID-19 dari citra X-ray pada dada (CXR) dengan menggunakan strategi desain generatif mesin-manusia. Model ini mencapai akurasi pengujian sebesar 93,3% pada dataset COVIDx, dengan sensitivitas 91,0% untuk kasus COVID-19. Keunggulan penelitian ini adalah transparansi model melalui metode *explainability* (GSInquire) yang memvalidasi bahwa keputusan model didasarkan pada indikator visual yang relevan di paru-paru dan penyediaan model open source serta dataset benchmark yang besar untuk komunitas riset, meskipun keterbatasan utamanya adalah jumlah data positif COVID-19 yang masih terbatas pada saat dilakukan penelitan oleh Wang et al. Pada penelitian ini, akan digunakan sebagian dataset COVIDx untuk pelatihan.

Pada dataset medis, seperti dataset COVIDx, diharapkan dapat meningkatkan akurasi dibandingkan model CNN standar (seperti ResNet atau COVID-Net). Hal ini dikarenakan penggabungan fitur konvolusional dan fitur atensi memungkinkan model untuk mempertahankan ekstraksi fitur lokal yang detail, serta menangkap konteks global seluruh area dada untuk membedakan pola penyebaran penyakit yang mungkin terlihat serupa secara lokal namun berbeda secara global.

Attention Augmented Convolutional Networks

Disingkat sebagi AACN, *Attention Augmented Convolutional Networks* adalah sebuah arsitektur *deep learning* yang diperkenalkan oleh Bello et al., [1] dari Google Brain pada tahun 2019. Arsitektur ini dirancang untuk mengatasi kelemahan fundamental pada *Convolutional Neural Networks* (CNN) standar.



GAMBAR 1
Gambar Attention-augmented convolution

Pada dasarnya, Attention Augmented Convolution bekerja dengan memproses input tensor melalui dua jalur paralel: jalur konvolusi standar dan jalur multi-head attention, kemudian menggabungkan hasilnya. Jika diberikan input tensor X, sebuah lapisan Attention Augmented Convolution (AAConv) dapat diformulasikan sebagai penggabungan dari peta fitur (feature maps) konvolusional dan peta fitur attentional sebagai berikut:

$$AAConv(X) = Concat[Conv(X), MHA(X)]$$

Dimana Conv(X) adalah hasil dari operasi konvolusi standar yang mendapatkan lokalitas, dan MHA(X) yang merupakan hasil dari mekanisme Multi-Head Attention yang menangkap dependensi global

- **Operasi Konvolusi (Local Feature Extraction)** Bagian ini bekerja seperti CNN standar, di mana filter (*kernel*) digeser ke seluruh citra input untuk menghasilkan feature map. Tujuannya adalah untuk menangkap fitur visual lokal seperti tepi, tekstur, dan bentuk spesifik pada area paru-paru.
- **Mekanisme Self-Attention (Global Feature Extraction)** Mekanisme *self-attention* yang digunakan dalam AACN diadaptasi dari arsitektur Transformer yang awalnya dikembangkan untuk pemrosesan bahasa alami oleh Vaswani et al., 2017 [7]. Berbeda dengan konvolusi, self-attention menghitung nilai rata-rata yang tertimbang (weighted average) dari nilai-nilai yang dihitung dari hidden units. Bobot dalam operasi ini dihasilkan secara dinamis melalui fungsi similaritas antar unit, sehingga interaksi antar sinyal input bergantung pada sinyal itu sendiri, bukan pada lokasi relatifnya yang telah ditentukan sebelumnya. Dalam AACN, mekanisme ini diimplementasikan sebagai berikut:
  - Flattening: Diberikan input tensor dengan dimensi (H, W, F_{in}), dimana merujuk pada *height* (tinggi), *width* (lebar), dan jumlah *input filters* pada *activation map*. Tensor tersebut

diratakan (flattened) menjadi matriks $X \in R^{HW \times F_{\text{in}}}$. Hal Ini memungkinkan algoritma memperlakukan piksel citra sebagai urutan fitur, mirip dengan token dalam teks.

○ ulti-head Attention: Mekanisme Multi-Head Attention merupakan komponen fundamental yang diadopsi dari arsitektur Transformer yang diperkenalkan oleh Vaswani et al., 2017 [7]. AACN menggunakan multi-head attention untuk memungkinkan model memperhatikan berbagai sub-ruang representasi secara bersamaan. Alih-alih hanya melakukan satu fungsi atensi tunggal dengan dimensi penuh, Multi-Head Attention memproyeksikan input X menjadi matriks queries (Q), keys (K), dan values (V) secara linear sebanyak h kali (di mana h adalah jumlah head) ke dalam dimensi yang berbeda

○ alkulasi Output: Output dari satu head (O_h) dihitung menggunakan persamaan berikut:

$$O_h = \text{Softmax}\left(\frac{(XW_q)(XW_k)^T}{\sqrt{d_k^h}}\right)(XW_v)$$

dimana W_q W_k dan W_v merupakan transformasi linear yang telah dipelajari untuk memetakan tensor masukan X, menjadi kueri $Q = XW_q$, *keys* $Q = XW_k$, dan *values* $Q = XW_v$. d_k^h adalah kedalaman key per head untuk normalisasi. Output dari semua head kemudian digabungkan (concatenated) dan diproyeksikan kembali untuk membentuk hasil akhir mekanisme atensi dengan rumus.

$$\text{MHA}(X) = \text{Concat}\left[O_1, \ldots, O_{N_h}\right]W^O$$

Jumlah filter yang dialokasikan untuk atensi ditentukan oleh rasio $v = \frac{d_v}{F_{out}}$ (rasio saluran atensi terhadap total filter) dan $k = \frac{d_k}{F_{out}}$ (rasio kedalaman key). Hal ini memungkinkan fleksibilitas dalam menyeimbangkan kompleksitas komputasi dan representasi fitur. Menurut Bello et al. multi-head attention menimbulkan kompleksitas sebesar $O((HW)^2 d_k)$ dan *memory cost* sebesar $O((HW)^2 N_h)$ karena diperlukan untuk menyimpan *attention maps* untuk setiap *head*.

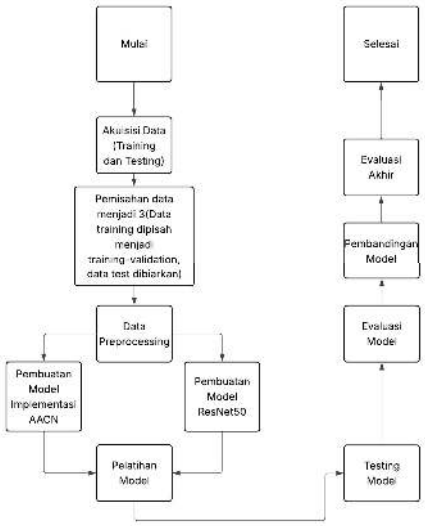
- **Two-Dimensional Relative Self-Attention** Tantangan utama dalam menerapkan self-attention pada citra adalah bahwa mekanisme standar bersifat permutation equivariant (tidak mengenali urutan atau posisi spasial piksel), yang membuatnya tidak efektif untuk data terstruktur seperti citra medis. Untuk mengatasi ini, Bello et al., [1] mengembangkan mekanisme Two-Dimensional Relative Self-Attention, berdasarkan konsep Shaw et al., [10] yang awalnya untuk data sekuensial 1D (teks) menjadi data spasial 2D (citra). Mekanisme ini menambahkan informasi posisi relatif (jarak antar piksel) ke dalam perhitungan atensi, bukan posisi absolut [1][10]. Implementasinya dilakukan dengan menambahkan relative height information dan relative width information secara independen. Persamaan logit atensi yang diperbarui menjadi:

$$l_{i,j} = \frac{q_i^T}{\sqrt{d_k^h}}\left(k_j + r^W_{j_x - i_x} + r^H_{j_y - i_y}\right)$$

Di mana r^W dan r^h adalah *learned embeddings* untuk jarak relatif lebar dan tinggi. Dengan cara ini, model dapat memahami struktur spasial citra X-Ray, memastikan bahwa hubungan antar fitur (seperti posisi paru-paru relatif terhadap tulang rusuk) tetap terjaga.

3. Sistem yang Dibangun

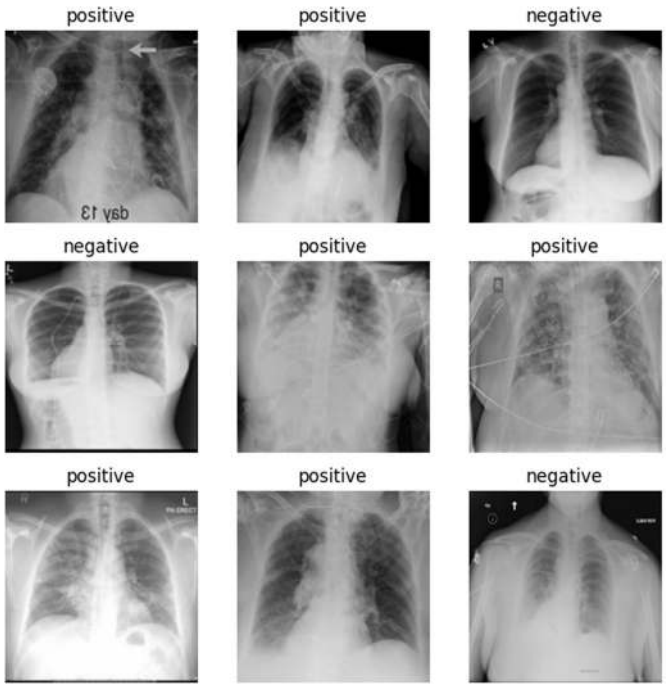
M



GAMBAR 2
Flowchart penelitian

Di penelitian ini, akan diambil dataset dari penelitian COVID-Net yang berjumlah 16.351 gambar. Dataset terdiri dari train dan test, dengan jumlah data test ialah 400, dengan 200 untuk positif Covid-19 dan 200 untuk negatif. Sedangkan data train, terdapat 13.793 data negatif Covid-19 dan 2.158 data positif Covid-19.

Dikarenakan ketidakseimbangan data yang didapat, dimana perbandingan gambar pelatihan yang berlabel negatif dan positif Covid-19 sangatlah jauh, maka data negatif yang diambil juga 2.158 supaya seimbang. Hasil jumlah dataset akhir ialah 4.316 train dan 400 test, dengan jumlah positif dan negatifnya seimbang. Lalu, pada data training akan dipisah lagi menjadi data training dan validation dengan rasio 80 : 20 untuk pelatihan, sedangkan data test tidak diubah.



GAMBAR 3
Gambar dataset X-ray dada

Seluruh gambar yang dimuat dikenakan pra-pemrosesan dasar

yaitu pengubahan ukuran menjadi 224x224, supaya sesuai dengan input ResNet50.

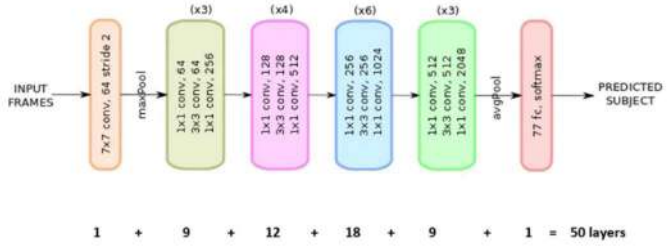
Pemilihan ResNet50 sebagai baseline dikarenakan sudah ada penelitian sebelumnya yang juga mengimplementasikan ResNet50 [1][3]. Terlebih lagi, Berdasarkan review oleh Sanapala et al., [14] ResNet-50 adalah arsitektur yang terbukti andal untuk diagnosis medis, termasuk pada citra X-Ray paru-paru.

3.1 Attention Augmented Convolutional Network

Implementasi dari layer AACN ini akan menggantikan *convolution 3×3* di dalam *bottleneck block* pada *layer*. Dikarenakan ResNet50 memiliki 4 *layer* atau tahapan konvolusi, maka ada 4 tempat kemungkinan pengimplementasian dari AACN pada arsitektur ResNet50.

TABEL 1
Tabel perbedaan ResNet50 dengan implementasi AACN terhadap ResNet50

No	ResNet-50 (Arsitektur Standar) [2]	ResNet dengan AACN [1]
Tujuan Utama	Arsitektur jaringan saraf konvolusional dalam untuk mengatasi masalah degradasi pada jaringan saraf yang sangat dalam menggunakan koneksi skip (residual).	Meningkatkan kemampuan ResNet untuk menangkap fitur yang kompleks dan asimetris.
Mekanisme Konvolusi	Konvolusi simetris standar, kernel 3x3.	Menggunakan kernel konvolusi yang asimetris (misalnya, 1xN dan Nx1) untuk menangkap dependensi spasial secara adaptif.
Kompleksitas	Rendah (Lebih cepat).	Tinggi.
Fokus	Mencari fitur secara seragam.	Bobot fitur berdasarkan relevansi spasial pada gambar.



GAMBAR 4
Gambar arsitektur ResNet50

Dengan adanya 4 *slot* atau *layer* untuk mengimplementasikan AACN, maka terdapat $2^4 = 16$ total kombinasi, dimana setiap *slot* bisa menggunakan AACN ataupun tidak (menggunakan *convolution* 3×3 seperti biasa pada ResNet50). Dalam implementasinya, akan digunakan pengkodean F (*False*) dan T (*True*) pada setiap kombinasi arsitektur yang ada, dengan contoh seperti berikut:

- TTFF: *True, True, False, False.* Menggunakan AACN pada *layer* 1 dan 2, dan *convolution* 3×3 biasa pada *layer* 3 dan 4.
- TFFF: *True, False, False, False.* Menggunakan AACN pada *layer* 1 saja, dan sisanya *convolution* 3×3 biasa.
- TFFT: *True, False, False, True.* Menggunakan AACN pada *layer* 1 dan 4, *convolution* 3×3 biasa pada *layer* 2 dan 3.
- FFFF: *False, False, False, False.* Karena tidak menggunakan AACN, maka arsitektur ini adalah ResNet50 biasa.

Karena *memory cost* $O((HW)^2 N_h)$ dapat menjadi terlalu besar untuk dimensi spasial yang besar, penelitian ini akan menggunakan konvolusi AACN perhatian mulai dari lapisan terakhir (dengan dimensi spasial terkecil) hingga mencapai batasan memori. Dikarenakan hal itu juga, adanya kemungkinan keterbatasan *memory* untuk beberapa kombinasi arsitektur, sehingga tidak dimungkinkan untuk dilatih dengan *hardware* yang terbatas.

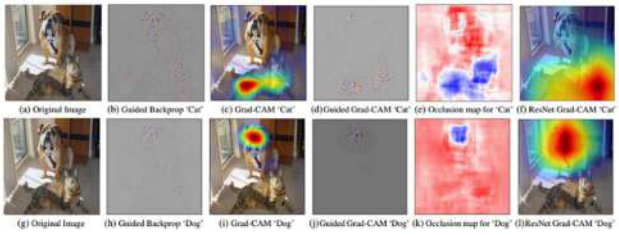
Implementasi AACN juga menggunakan nilai $k = 0.2\ v = 2$, mengikuti setelan yang dilakukan Bello et al.[1] pada eksperimennya dengan ResNet50.

3.2 Grad-CAM

Gradient-weighted Class Activation Mapping, atau yang biasa dikenal sebagai Grad-CAM, adalah teknik visualisasi yang dikembangkan oleh Selvaraju et al. [8] untuk memproduksi "penjelasan visual" bagi keputusan yang dibuat oleh berbagai arsitektur CNN. Teknik ini menghasilkan peta lokalisasi kasar (*coarse localization map*) yang menyoroti area-area penting pada citra input yang digunakan model untuk memprediksi konsep atau kelas tertentu.

Grad-CAM dapat diterapkan pada berbagai keluarga model CNN tanpa memerlukan perubahan arsitektur atau pelatihan ulang (*re-training*), termasuk CNN dengan *fully-connected layers* (seperti VGG), model untuk *structured outputs* (seperti captioning), dan model dengan *input* multi-modal. Hal ini menjadikan Grad-CAM cocok untuk digunakan dalam mempelajari peforma dari model pada penelitian ini, serta memverifikasi apakah apakah model benar-benar "melihat" pada area paru-paru yang terinfeksi, atau justru melakukan prediksi berdasarkan artefak yang tidak relevan (seperti teks pada X-Ray atau tulang rusuk).

Prinsip dasar Grad-CAM adalah memanfaatkan informasi gradien yang mengalir ke lapisan konvolusi terakhir (*final convolutional layer*) dari CNN. Lapisan konvolusi terakhir dipilih karena lapisan ini dianggap memiliki kompromi terbaik antara semantik tingkat tinggi (*high-level semantics*) dan informasi spasial yang rinci. Hasil akhir dari Grad-CAM adalah heatmap yang memperlihatkan bagian mana yang diperhatikan oleh model.



GAMBAR 5
Contoh penjelasan visual dari arsitektur Grad-CAM yang digunakan pada penelitian Selvaraju et al. [8]

3.3 Validasi Data

Beberapa metrik evaluasi yang dipakai ialah:

- Confusion Matrix

Confusion matrix merangkum hasil prediksi model dengan membedakan klasifikasi yang benar dan salah. True positive (TP) merujuk pada kasus Covid-19 yang diidentifikasi dengan benar oleh model, sedangkan true negative (TN) adalah kasus non Covid-19 yang dikenali dengan benar. False positive (FP) terjadi

jika model salah mengklasifikasikan kasus sehat atau non Covid-19 sebagai Covid-19, dan false negative (FN) terjadi ketika model salah mendeteksi kasus sehat ata non Covid-19 sebagai Covid-19. Metrik-metrik ini memberikan evaluasi yang jelas tentang akurasi model dan distribusi kesalahan.

- Accuracy

Menilai tingkat efektivitas model secara menyeluruh dengan membandingkan jumlah prediksi yang tepat terhadap total seluruh data.

$$A = \frac{TP + TN}{TP + TN + FP + FN}$$

- Precision

Fokus pada akurasi hasil. Precision menilai seberapa besar model berhasil memberikan label positif tanpa membuat terlalu banyak kesalahan "salah tebak" (*false positive*).

$$P = \frac{TP}{TP + FP}$$

- Recall

Mengukur kemampuan model untuk mendeteksi atau mengenal kembali semua kasus yang memang benar-benar positif dari data yang tersedia.

$$R = \frac{TP}{TP + FN}$$

- F1-Score

Merupakan metrik gabungan yang mengimbangkan nilai recall dan precision. F1-Score memberikan gambaran kinerja model dengan lebih objektif terutama ketika ada ketidakseimbangan pada jumlah data yang ada.

$$F_1 = 2 \cdot \frac{P \cdot R}{P + R}$$

4. Evaluasi

4.1 Hasil Pengujian

Setelah melakukan pengujian, penggunaan AACN pada *layer* pertama menggunakan *memory* yang sangat besar, sehingga tidak bisa melanjutkan semua kombinasi arsitektur dengan AACN pada *layer* pertama, dan menyisakan 8 kombinasi yang tersisa.

Pangkat dua pada $(HW)^2$ $dalam\ Memory\ cost\ O((HW)^2 N_h)$ menunjukkan bahwa kebutuhan memori meningkat secara kuadratik seiring dengan bertambahnya resolusi citra. Pada *layer* pertama, dimensi spasial (H) dan (W) masih sangat besar karena citra input belum mengalami proses downsampling yang signifikan. Sebaliknya, pada layer-layer yang lebih dalam, dimensi spasial feature map telah direduksi secara signifikan melalui serangkaian operasi pooling atau konvolusi.

TABEL 2
Hasil Testing Implementasi AACN

	Pred / True	Confusion Matrix		Precision	Recall	F1-Score	Overall Accuracy
FFFF (ResNet 50)		Negative	Positive				
	Negative	192	8	94%	96%	95%	94.75%
	Positive	13	187	96%	94%	95%	

	Pred / True	Confusion Matrix		Precision	Recall	F1-Score	Overall Accuracy
FFFT		Negative	Positive				
	Negative	191	9	96%	95%	96%	95.75%
	Positive	8	192	96%	96%	96%	

	Pred / True	Confusion Matrix		Precision	Recall	F1-Score	Overall Accuracy
FFTF		Negative	Positive				
	Negative	191	9	92%	95%	94%	93.75%
	Positive	16	184	95%	92%	94%	

	Pred / True	Confusion Matrix		Precision	Recall	F1-Score	Overall Accuracy
FFTT		Negative	Positive				
	Negative	176	24	92%	88%	90%	90.25%
	Positive	15	185	89%	93%	90%	

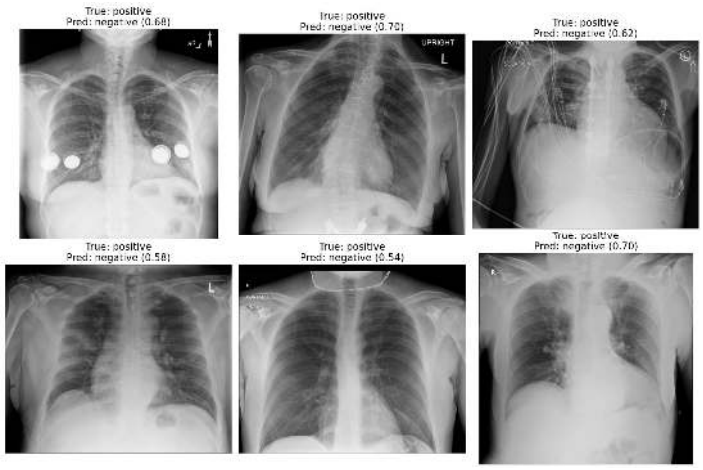
	Pred / True	Confusion Matrix		Precision	Recall	F1-Score	Overall Accuracy
FTFF		Negative	Positive				
	Negative	185	15	96%	93%	94%	94.50%
	Positive	7	193	93%	96%	95%	

	Pred / True	Confusion Matrix		Precision	Recall	F1-Score	Overall Accuracy
FTFT		Negative	Positive				
	Negative	179	21	93%	90%	91%	91.50%
	Positive	13	187	90%	94%	92%	

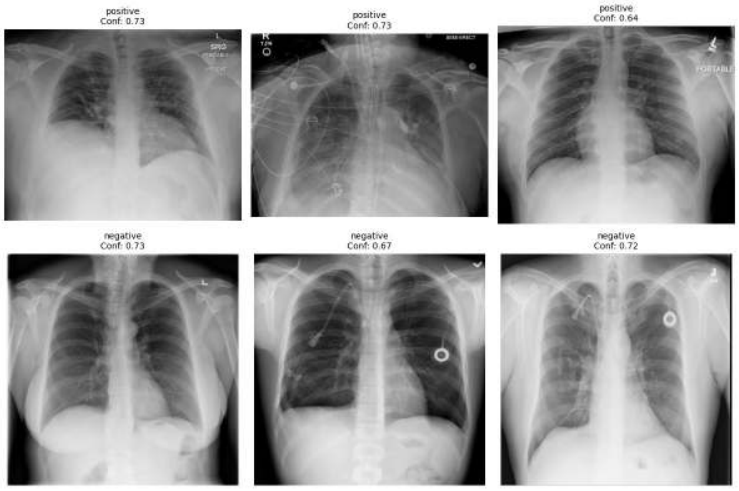
	Pred / True	Confusion Matrix		Precision	Recall	F1-Score	Overall Accuracy
FTTF		Negative	Positive				
	Negative	168	32	96%	84%	90%	90.25%
	Positive	7	193	86%	96%	91%	

	Pred / True	Confusion Matrix		Precision	Recall	F1-Score	Overall Accuracy
FTTT		Negative	Positive				
	Negative	178	22	88%	89%	89%	88.50%
	Positive	24	176	89%	88%	88%	

Walaupun FFFT memiliki akurasi yang 1% lebih baik dibandingkan dengan ResNet50 (FFFF), arsitektur yang lain memiliki peforma yang lebih buruk dalam mendeteksi Covid-19. FFTF dan FTFF merupakan arsitektur selanjutnya yang memiliki peforma terbagus, lalu dilanjutkan dengan arsitektur *layer* AACN yang lebih banyak.

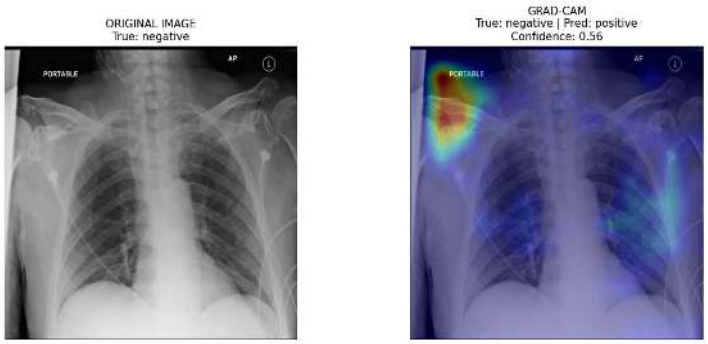


GAMBAR 6
Contoh gambar yang salah klasifikasinya pada testing dengan arsitektur FFFT

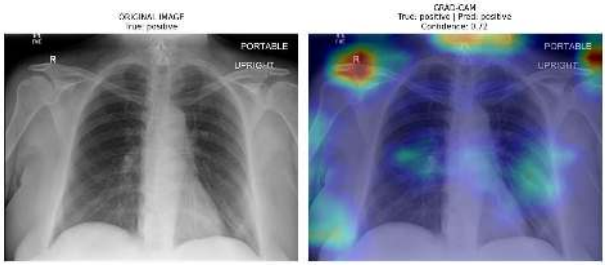


GAMBAR 7
Contoh gambar yang benar klasifikasinya pada testing dengan arsitektur FFFT

Secara sekilas, perbedaan cara model mengklasifikasi yang benar dengan yang salah kurang bisa dijelaskan. Kedua kelompok gambar diatas menunjukkan adanya teks, simbol, serta peralatan medis yang dapat mempengaruhi kinerja model.

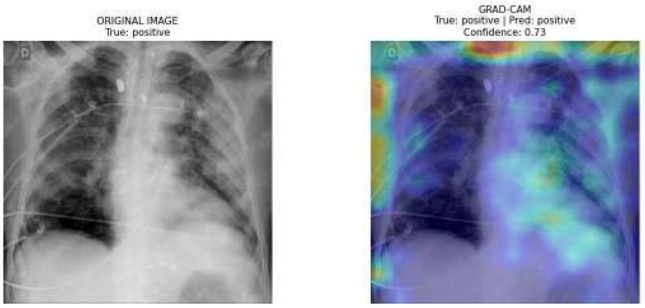


GAMBAR 8
Contoh gambar yang salah klasifikasinya pada testing dengan arsitektur FFFT dengan menggunakan Grad-CAM

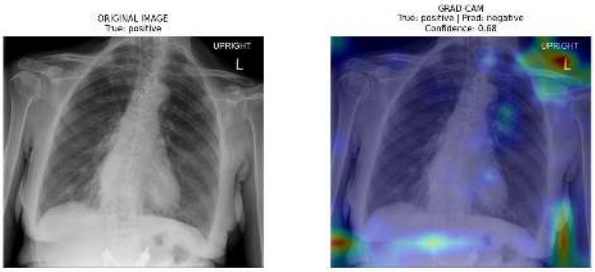


GAMBAR 9
Contoh gambar yang benar klasifikasinya pada testing dengan arsitektur FFFT dengan menggunakan Grad-CAM

Jika dilihat dengan Grad-CAM, *heatmap* yang muncul menunjukkan pendeteksian oleh model berfokus lebih ke simbol, bahu, lehet, serta ujung gambar, dibandingkan dengan dada atau paru-paru. Hal ini menandakan bahwa model banyak belajar serta fokus pada hal-hal yang berkaitan dengan penyakit Covid-19 dan juga area paru-paru.



GAMBAR 10
Contoh gambar yang benar klasifikasinya pada testing dengan arsitektur FFFF (ResNet50) dengan menggunakan Grad-CAM



GAMBAR 11
Contoh gambar yang salah klasifikasinya pada testing dengan arsitektur FFFF (ResNet50) dengan menggunakan Grad-CAM

Sedangkan pada ResNet50 (FFFF), tidak jauh berbeda dengan FFFT. Hasil Grad-CAM menunjukkan bahwa pembelajaran dan klasifikasi berfokus pada ujung gambar, dan simbol. Walaupun begitu, ResNet50 lebih fokus dalam mendeteksi pada bagian dada atau paru-paru, dibandingkan dengan FFFT, seperti yang ada pada gambar 10.

4.2 Analisis Hasil Pengujian

Setelah melakukan beberapa kali percobaan, sebagian besar hasil ResNet50 dengan implementasi AACN memiliki peforma yang sedikit lebih buruk daripada ResNet50 biasa. Walaupun arsitektur FFFT memberikan akurasi yang sedikit lebih bagus, penggunaan AACN ini masih tetap dinilai kurang mampu dalam mempelajari dataset yang ada.

Penelitian López-Cabrera et al., 2021 [2] memperingatkan bahwa model Deep Learning sering kali mempelajari bias dari sumber data, seperti label teks, sudut gambar, atau artefak mesin, daripada patologi penyakit itu sendiri. Akibatnya, AACN mungkin mendapatkan akurasi pelatihan yang tinggi, namun gagal saat diuji pada data bersih tanpa simbol (generalisasi buruk) karena ia mempelajari artefak, bukan fitur paru-paru.

Adapun hasil yang didapatkan dari Grad-CAM, menunjukkan

bahwa model berfokus terhadap hal-hal diluar dada atau paru-pau. Hal ini juga dijelaskan oleh López-Cabrera et al., 2021 [9], dimana ditemukan bukti bahwa banyak model CNN berkinerja tinggi sebenarnya "melihat" ke daerah di luar paru-paru (seperti bahu atau latar belakang) untuk membuat keputusan. Ini membuktikan bahwa model tersebut melakukan *cheating* atau *shortcut learning*.

Selain itu, AACN menggunakan 2D relative positional embeddings untuk mempertahankan ekuivariansi translasi dan memahami struktur spasial gambar. Namun, jika simbol atau artefak muncul di posisi yang acak akibat variasi dimensi asli dan proses resizing, mekanisme atensi mungkin kesulitan mempelajari embedding posisi relatif yang konsisten. Bello et al., 2019 [1] mencatat bahwa tanpa pengkodean posisi yang tepat, kemampuan atensi untuk memodelkan struktur visual akan menurun.

Dataset yang didapatkan dari penelitian Wang et al., 2020 [3] merupakan gabungan dari beberapa sumber dataset lainnya. Walaupun begitu, keberagaman dataset bisa memperburuk pelatihan dikarenakan variasi dari gambarnya yang memiliki dimensi, posisi dan juga simbol atau artefak yang berbeda antar sumber dataset. Terlebih lagi, Ada kemungkinan bahwa fitur pembeda utama untuk COVID-19 pada dataset bersifat sangat lokal (tekstur spesifik pada lobus paru) dan tidak memerlukan konteks global jarak jauh yang ditawarkan oleh atensi. Jika dependensi jarak jauh (long-range interactions) antar piksel tidak signifikan untuk klasifikasi kasus ini, penambahan mekanisme atensi hanya menambah kompleksitas komputasi tanpa memberikan keuntungan informasi yang relevan.

5. Kesimpulan

Walaupun AACN menghasilkan akurasi yang sedikit lebih bagus, implementasi AACN dinilai kurang optimal pada pelatihan ini. Keberagaman dataset serta adanya simbol dapat memperburuk pelatihan. Selain itu, dibutuhkan penelitian lebih lanjut karena penelitian ini menggunakan sebagian kecil dari dataset COVIDx dan juga kurangnya variasi pada pelatihan.

Untuk pengembangan selanjutnya, disarankan agar evaluasi terhadap mekanisme attention tidak hanya dilakukan secara kualitatif melalui visualisasi visual semata, melainkan juga melalui validasi kuantitatif yang objektif. Penelitian ini merekomendasikan penerapan strategi machine-centric sebagaimana diusulkan oleh Lin et al., [15] dalam paper mereka yang berjudul *'Do Explanations Reflect Decisions?*'.

Metode ini memperkenalkan pengukuran *Impact Score*, yaitu teknik audit model dengan cara menghapus atau menutupi (*masking*) area yang memiliki bobot atensi tinggi dan mengukur penurunannya terhadap probabilitas prediksi kelas [15]. Implementasi metode ini penting untuk membuktikan bahwa fitur spasial yang dipelajari oleh AACN benar-benar berkontribusi terhadap keputusan diagnostik COVID-19, serta untuk memastikan bahwa model tidak mengalami *shortcut learning* dengan berfokus pada artefak yang tidak relevan.

Daftar Pustaka

[1] Bello, Irwan, et al. "Attention augmented convolutional networks." *Proceedings of the IEEE/CVF international conference on computer vision*. 2019. He, K., Zhang, X., Ren, S., dan Sun, J. 2016. Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 770-778

[2] López-Cabrera, José Daniel, et al. "Current limitations to identify COVID-19 using artificial intelligence with chest X-ray imaging." *Health and Technology* 11.2 (2021): 411-424.

[3] Wang, Linda, Zhong Qiu Lin, and Alexander Wong. "COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images." *Scientific reports* 10.1 (2020): 19549.

[4] Ciotti, Marco, et al. "The COVID-19 pandemic." *Critical reviews in clinical laboratory sciences* 57.6 (2020): 365-388.

[5] Narin, Ali, Ceren Kaya, and Ziynet Pamuk. "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks." *Pattern Analysis and Applications* 24.3 (2021): 1207-1220.

[6] Jaiswal, Amit Kumar, et al. "Identifying pneumonia in chest X-rays: A deep learning approach." *Measurement* 145 (2019): 511-518.

[7] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).

[8] Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." Proceedings of the IEEE international conference on computer vision. 2017.

[9] López-Cabrera, José Daniel, et al. "Current limitations to identify covid-19 using artificial intelligence with chest x-ray imaging (part ii). The shortcut learning problem." *Health and technology* 11.6 (2021): 1331-1345.

[10] Shaw, Peter, Jakob Uszkoreit, and Ashish Vaswani. "Self-attention with relative position representations." *arXiv preprint arXiv:1803.02155* (2018).

[11] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[12] Rao, Adrit, et al. "Studying the effects of self-attention for medical image analysis." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

[13] Nguyen, Thanh Thi, et al. "Artificial intelligence in the battle against coronavirus (COVID-19): a survey and future research directions." arXiv preprint arXiv:2008.07343 (2020).

[14] Sanapala, Aparna, et al. "A Review On Image Based Diagnosis Using ResNet-50." International Journal of Research and Analytical Reviews (IJRAR), IJRAR.org. 2024.

[15] Lin, Zhong Qiu, et al. "Do explanations reflect decisions? A machine-centric strategy to quantify the performance of explainability algorithms." *arXiv preprint arXiv:1910.07387* (2019).