

MAKALAH

Deteksi COVID-19 pada Dataset X-Ray Dada dengan Metode Capsule Networks (CapsNets)



Disusun oleh Tim **Yakuy 2** :

Ardacandra Subiantoro (18/427572/PA/18532)

Arief Pujo Arianto (18/430253/PA/18766)

Chrystian (18/430257/PA/18770)

**PROGRAM STUDI S1 ILMU KOMPUTER
DEPARTEMEN ILMU KOMPUTER DAN ELEKTRONIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS GADJAH MADA
YOGYAKARTA**

2020

1 Latar Belakang

Jumlah kasus COVID-19 terus meningkat secara eksponensial. Jumlah kasus total COVID-19 di seluruh dunia sudah mencapai lebih dari 19 juta kasus, dengan lebih dari 700 ribu kematian akibat COVID-19. Kasus COVID-19 di Indonesia sudah mencapai lebih dari 120 ribu dengan lebih dari 5 ribu kematian. Dampak negatif dari pandemi COVID-19 ini sangat terasa di Indonesia. Direktur Jenderal Pajak Kementerian Keuangan (Kemenkeu) Suryo Utomo membagi dampak pandemi COVID-19 menjadi tiga garis besar [6]. Dampak pertama adalah membuat konsumsi rumah tangga atau daya beli yang merupakan penopang 60 persen terhadap ekonomi jatuh cukup dalam. Hal ini dibuktikan dengan data dari BPS yang mencatatkan bahwa konsumsi rumah tangga turun dari 5,02 persen pada kuartal I 2019 ke 2,84 persen pada kuartal I tahun ini. Dampak kedua yaitu pandemi menimbulkan adanya ketidakpastian yang berkepanjangan sehingga investasi ikut melemah dan berimplikasi pada terhentinya usaha. Dampak ketiga adalah seluruh dunia mengalami pelemahan ekonomi sehingga menyebabkan harga komoditas turun dan ekspor Indonesia ke beberapa negara juga terhenti.

Jumlah alat testing yang terbatas membuat tidak mungkin setiap pasien dengan penyakit pernafasan untuk dites dengan teknik konvensional (RT-PCR) untuk mendeteksi COVID-19. Jumlah ruang di rumah sakit, jumlah ventilator, dan jumlah alat protektif (PPE) bagi tenaga medis sangat terbatas, sehingga perlu digunakan dengan seefisien mungkin. Untuk meningkatkan efisiensi penggunaan alat medis, sangat penting untuk dapat membedakan pasien dengan *severe acute respiratory illness* (SARI) yang mungkin terinfeksi COVID-19. Oleh karena itu, kami mencoba menggunakan gambar X-Ray dada untuk mendeteksi infeksi COVID-19. Penggunaan gambar X-Ray memiliki beberapa keuntungan dibandingkan tes konvensional [5] :

1. Pengambilan gambar X-Ray lebih banyak digunakan dan lebih murah dibandingkan tes konvensional.
2. Gambar X-Ray dapat dengan mudah ditransfer tanpa perlu ada alat transportasi dari tempat pengambilan gambar ke tempat analisis gambar, sehingga proses diagnostik dapat dilaksanakan dengan cepat.
3. Berbeda dengan CT Scan, terdapat alat-alat pengambil gambar X-Ray portable sehingga tes dapat dilakukan di ruangan-ruangan isolasi. Hal ini menyebabkan kurangnya penggunaan alat-alat protektif dan mengurangi risiko infeksi tersebar ke pasien-pasien lain di rumah sakit.

Dataset yang akan kami gunakan berasal dari dua sumber. Sumber pertama adalah Kaggle Chest X-Ray Images (Pneumonia) dataset (2018) milik Paul Mooney [3] yang berisi 5,863 gambar X-Ray dada yang dipilih dari pasien-pasien anak berumur satu sampai lima tahun dari Guangzhou Women and Children's Medical Center, Guangzhou. Sumber kedua adalah dari COVID-19 Chest X-Ray dataset (2020) milik Bachir [4] yang berisi 357 gambar X-Ray dada yang terinfeksi COVID-19 dan juga metadata yang mendeskripsikan informasi-informasi seperti jenis kelamin, umur, lokasi, dll dari pasien.

2 Tujuan dan Manfaat

Tujuan utama kami adalah untuk mengusulkan model berbasis *neural network* yang dapat secara akurat mendeteksi infeksi COVID-19 dari gambar X-Ray dada pasien. Model ini diharapkan dapat bertindak sebagai alat otomatis yang dapat memandu petugas kesehatan dalam mendiagnosis gambar X-Ray agar mengambil kesimpulan yang akurat. Manfaat yang diharapkan adalah peningkatan akurasi dan aksesibilitas tes deteksi infeksi COVID-19.

Perlu kami tekankan bahwa kami tidak mengusulkan penggunaan model ini untuk menggantikan metode tes diagnostik konvensional. Kami berharap bahwa model ini dapat bersifat saling melengkapi dengan tes konvensional, dimana hasil prediksi dari model dapat memastikan hasil dari tes konvensional atau menjadi metode alternatif bila tes konvensional tidak memungkinkan.

Tujuan kedua dari penambangan data yang kami lakukan adalah untuk mengeksplorasi metadata dari dataset yang tersedia melalui *Exploratory Data Analysis*. Kami mencoba untuk menelusuri apa pengaruh unsur-unsur seperti jenis kelamin dan umur terhadap infeksi COVID-19. Kami berharap informasi yang dapat kami ekstraksi dapat bermanfaat dalam upaya pendeteksi dan pengendalian kasus COVID-19.

3 Batasan yang Digunakan

Batasan masalah yang akan kami gunakan adalah sebagai berikut :

- Makalah ini akan membahas eksplorasi metadata dan penyusunan model untuk mendeteksi infeksi COVID-19 dari gambar X-Ray dada pasien.
- Dataset yang akan digunakan dibatasi pada Kaggle Chest X-Ray Images (Pneumonia) dataset (2018) milik Paul Mooney [3] dan COVID-19 Chest X-Ray dataset (2020) milik Bachir [4].

- Jenis penyakit yang akan diklasifikasi dibatasi pada COVID-19 dan Pneumonia saja.
- Metode penambangan data yang akan kami gunakan adalah Capsule Networks (CapsNets).

4 Metode Penambangan Data

4.1 Introduksi CapsNet

CapsNet (Capsule Neural Network) adalah sistem pembelajaran mesin bertipe *Artificial Neural Network* yang dapat memodelkan hubungan-hubungan bersifat hierarkis. Pendekatan ini dikembangkan dengan mengikuti organisasi syaraf biologis. Perbedaan CapsNet dengan *Convolutional Neural Network* (CNN) biasa adalah penambahan struktur *capsule*, dimana *capsule* dengan tingkat lebih tinggi akan menggunakan kembali keluaran dari beberapa *capsule* tingkat rendah untuk menghasilkan representasi informasi yang lebih stabil. Keluaran dari CapsNet adalah vektor berisi probabilitas dan *pose* (kombinasi dari posisi dan orientasi) dari sebuah observasi.

Salah satu keuntungan utama CapsNet adalah CapsNet dapat menjadi solusi “*Picasso Problem*” di bidang pengenalan gambar. Contoh “*Picasso Problem*” adalah ketika sebuah gambar wajah seseorang memiliki semua fitur-fitur seperti hidung dan mulut, namun posisi hidung ditukar dengan posisi mulut. *Convolutional Neural Network* biasa akan kesulitan mendeteksi gambar tersebut sebagai sebuah wajah. CapsNet dapat mengatasi masalah ini dengan mengeksplorasi fakta bahwa walaupun perubahan sudut pandang memiliki efek non-linear di tingkat pixel, efeknya linear di tingkat objek.

4.2 Latar Belakang

Awal mula perkembangan CapsNet terjadi pada tahun 2000, dimana Geoffrey Hinton mendeskripsikan sistem untuk merepresentasikan gambar dengan menggunakan gabungan teknik *segmentation* dan *parse trees*. Sistem ini terbukti berguna dalam mengklasifikasikan digit-digit tulisan tangan di dataset MNIST.

Pada tahun 2017, Hinton dan timnya memperkenalkan mekanisme *dynamic routing* untuk *capsule networks*. Pendekatan ini berhasil meningkatkan performa model dalam mengolah data MNIST dan mengurangi ukuran data latihan. Hasilnya diklaim mengalahkan performa CNN terutama pada digit-digit yang *overlap*.

4.3 Transformasi

Dalam bidang citra komputer, terdapat tiga jenis sifat yang dapat dimiliki sebuah objek :

1. *Invariant* : sifat objek yang tidak berubah sebagai hasil dari suatu bentuk transformasi. Sebagai contoh, area dari lingkaran tidak berubah ketika lingkaran digeser ke kiri atau kanan.
2. *Equivariant* : sifat objek yang perubahannya dapat diprediksi bila dilakukan sebuah transformasi. Contohnya adalah titik pusat lingkaran berubah sesuai dengan arah gerak lingkaran tersebut.
3. *Non-equivariant* : sifat objek yang perubahannya tidak dapat diprediksi bila dilakukan sebuah transformasi. Contohnya adalah bila sebuah lingkaran ditransformasi menjadi sebuah oval, rumus untuk menghitung keliling objek tersebut sudah bukan $2 \pi r$.

Kelas sebuah objek dalam citra komputer diharapkan bersifat *invariant* ketika diterapkan banyak transformasi. Sebagai contoh, sebuah mobil harusnya tetap diklasifikasikan sebagai mobil walaupun gambarnya dibalik atau dkecilkan. Namun pada kenyataannya, sebagian besar karakteristik gambar bersifat *equivariant*.

Karakteristik-karakteristik *equivariant* seperti volume dan letak bagian-bagian objek disimpan dalam sebuah *pose*. *Pose* adalah data yang mendeskripsikan translasi, rotasi, skala, dan refleksi dari objek. Translasi adalah perubahan lokasi di satu atau lebih dimensi, rotasi adalah perubahan orientasi, skala adalah perubahan ukuran, dan refleksi adalah gambar yang dicerminkan.

CapsNet mempelajari *linear manifold* global antara objek dan pose objek tersebut dan merepresentasikan informasi dalam bentuk matriks. Dengan menggunakan matriks tersebut, CapsNet dapat mengidentifikasi objek walaupun objek tersebut sudah mengalami sejumlah transformasi. Informasi spasial dari objek dipisah dari informasi untuk mengklasifikasi objek itu sendiri, sehingga objek dapat diklasifikasi secara independen dari *pose*-nya.

4.4 Pooling

Dalam CNN konvensional, digunakan teknik *pooling layer* yang berguna untuk mengurangi jumlah detail informasi yang diproses di lapisan yang lebih tinggi. *Pooling* mengizinkan sedikit *invariance translasional* (objek berada di lokasi yang berbeda) dan meningkatkan jumlah tipe fitur yang dapat direpresentasikan. CapsNet menolak penggunaan teknik *pooling*, dengan alasan :

1. *Pooling* melanggar persepsi bentuk biologis karena tidak ada bingkai koordinat intrinsik;
2. *Pooling* membuang informasi posisional, sehingga menghasilkan *invariance* dan bukan *equivariance*;
3. *Pooling* mengabaikan *linear manifold* yang mendasari banyak variasi dalam gambar-gambar;
4. *Pooling* membuat rute informasi antar lapisan secara statis, tidak mengkomunikasikan info potensial ke fitur-fitur secara dinamis;
5. *Pooling* merusak detektor fitur yang dengan dengan *pooling layer*, karena ada sejumlah informasi yang dihapus.

4.5 Capsules

Sebuah *capsule* adalah kumpulan neuron yang memiliki vektor aktivitas yang merepresentasikan berbagai properti-properti dari suatu tipe entitas yang terdapat pada gambar, seperti posisi, ukuran, orientasi, dll [7]. Kumpulan neuron tersebut secara kolektif menghasilkan vektor aktivitas yang diderivasi oleh CapsNet dari data masukan. Kemungkinan adanya suatu entitas dalam gambar direpresentasikan oleh panjang vektor, sedangkan orientasi vektor mengkuantifikasikan properti dari *capsule*.

Di struktur *Artificial Network* tradisional, keluaran dari neuron-neuron adalah sebuah nilai skalar yang secara longgar merepresentasikan probabilitas dari sebuah observasi. CapsNet menggantikan detektor fitur yang menghasilkan nilai skalar dengan *capsule-capsule* yang menghasilkan vektor dan *max-pooling* dengan metode *routing-by-agreement*.

Capsules bersifat mandiri dari satu sama lain, sehingga probabilitas deteksi benar meningkat drastis bila beberapa *capsule* setuju akan suatu prediksi. Dua *capsule* yang mengolah entitas berdimensi enam hanya akan sepakat akan suatu nilai dengan margin 10 persen karena kebetulan hanya terjadi satu dalam satu juta kali percobaan. Bila dimensi entitas bertambah, maka kemungkinan kesepakatan karena kebetulan berkurang secara eksponensial.

Capsules di lapisan lebih tinggi mengambil keluaran dari *capsules* di lapisan lebih rendah, lalu menerima dari *capsules* yang keluarannya berkelompok. Sebuah kelompok akan menyebabkan *capsule* lebih tinggi untuk menghasilkan keluaran dengan probabilitas tinggi bahwa entitas terdapat dalam observasi tersebut. *Capsules* tingkat tinggi mengabaikan *outlier* karena mereka hanya berkonsentrasi pada keluaran yang berkelompok dengan keluaran-keluaran lain.

4.6 Routing-by-agreement

Keluaran dari suatu *capsule* (anak) akan dihubungkan ke *capsule* (orangtua) di lapisan berikutnya sesuai dengan kemampuan *capsule* anak untuk memprediksi keluaran dari *capsule* orangtua. Setelah melalui beberapa iterasi, keluaran setiap *capsule* orangtua dapat bersatu dengan prediksi beberapa *capsule* anak dan berpisah dengan beberapa *capsule* anak lain.

Untuk setiap *capsule* orangtua yang mungkin, setiap *capsule* anak menghitung vektor prediksi dengan mengalikan keluarannya dengan matriks beban yang dilatih dengan *backpropagation*. Lalu, keluaran dari *capsule* orangtua dihitung sebagai produk skalar dari sebuah prediksi dengan koefisien yang merepresentasikan probabilitas bahwa *capsule* anak adalah milik *capsule* orangtua tersebut. Sebuah *capsule* anak yang prediksinya relatif dekat dengan keluaran hasilnya akan meningkatkan koefisien antara *capsule* anak dan *capsule* orangtua, sedangkan prediksi yang relatif jauh akan mengurangi koefisien. Koefisien yang meningkat akan meningkatkan kontribusi dari *capsule* anak ke *capsule* orangtua, meningkatkan produk skalar dari prediksi *capsule* dengan keluaran *capsule* orangtua. Setelah beberapa iterasi, koefisien yang dihasilkan akan menghubungkan secara kuat *capsule* orangtua dengan *capsule-capsule* yang paling mungkin sebagai anak. Hal ini mengindikasikan bahwa keberadaan *capsule-capsule* anak mengimplikasikan keberadaan entitas yang direpresentasikan *capsule* orangtua. Semakin banyak *capsule* anak yang prediksinya mirip dengan keluaran *capsule* orangtua, semakin cepat koefisien bertambah dan mempercepat konvergensi. *Pose* dari *capsule* orangtua secara progresif menjadi lebih cocok dengan *pose* anak-anaknya.

Logit awal dari koefisien-koefisien adalah probabilitas log prior bahwa sebuah *capsule* anak adalah miliki sebuah *capsule* orangtua. Prior-prior dapat dilatih bersama dengan beban-beban pada matriks beban. Prior tergantung lokasi dan tipe dari *capsule* orangtua dan anak, namun tidak tergantung masukan yang sedang diolah. Di setiap iterasi, koefisien akan disesuaikan melalui metode *softmax* sehingga jumlah setiap koefisien adalah 1. *Softmax* akan semakin memperbesar nilai-nilai yang besar dan semakin mengurangi nilai-nilai kecil.

Mekanisme pembuatan rute dinamis ini menyediakan sarana yang dibutuhkan untuk memisahkan objek-objek yang *overlap*.

Algoritma ini dapat direpresentasikan secara matematis sebagai berikut :

$$\hat{u}_{j|i} = W_{ij}u_i$$

Pose vektor u_i dirotasikan dan ditranslasi oleh matriks W_{ij} menjadi vektor $\hat{u}_{j|i}$ yang memprediksikan keluaran dari *capsule* orangtua.

$$s_j = \sum c_{ij} \hat{u}_{j|i}$$

Capsule s_j di lapisan yang lebih tinggi akan diberikan jumlahan dari prediksi-prediksi semua *capsule-capsule* di lapisan yang lebih rendah, dimana setiap prediksi dikalikan dengan koefisien c_{ij} .

4.7 Training

CapsNet dilatih secara *supervised* dengan cara meminimalisir jarak euclidean antara gambar dan keluaran dari sebuah CNN yang merekonstruksikan gambar masukan dari keluaran *capsule-capsule* terminal. Jaringan dilatih secara diskriminatif dengan menggunakan *routing-by-agreement* iteratif.

Terdapat dua tipe *loss* yang digunakan untuk melatih model. *Loss* tipe pertama adalah *margin loss*. Panjang dari sebuah vektor instansiasi merepresentasikan probabilitas bahwa entitas dari *capsule* terdapat pada sebuah gambar. Untuk membolehkan keberadaan entitas-entitas dengan jumlah banyak, *margin loss* terpisah harus dihitung untuk setiap *capsule*. Menurunkan bobot *loss* dari entitas-entitas yang absen menghentikan pembelajaran mengecilkan panjang vektor aktivitas untuk semua entitas. *Loss* total adalah jumlah dari *loss* setiap entitas. Hinton menggunakan *loss function* sebagai berikut :

$$L_k = \underbrace{T_k \max(0, m^+ - \|\mathbf{v}_k\|)^2}_{\text{class present}} + \underbrace{\lambda (1 - T_k) \max(0, \|\mathbf{v}_k\| - m^-)^2}_{\text{class not present}}, \quad T_k = \begin{cases} 1, & \text{digit of class } k \text{ present} \\ 0, & \text{otherwise} \end{cases}$$

Hinton menggunakan nilai 0.9 untuk m^+ dan nilai 0.1 untuk m^- supaya panjang tidak terlalu maksimal atau jatuh. Penurunan bobot dari bobot inisial kelas-kelas yang abses diatur dengan λ , dimana Hinton menggunakan nilai 0.5.

Loss tipe kedua adalah *reconstruction loss*. *Reconstruction loss* mendorong entitas-entitas untuk menyandikan parameter-parameter instansiasi dari masukan mereka. Vektor aktivitas final lalu digunakan untuk merekonstruksi gambar masukan menggunakan CNN *decoder* yang terdiri dari tiga lapisan yang terhubung penuh. Rekonstruksi meminimalisir jumlah dari kuadrat perbedaan antara keluaran unit logistik dan intensitas pixel. Hinton menggunakan nilai 0.0005 sebagai koefisien untuk memperkecil *reconstruction loss* supaya tidak mendominasi *margin loss* ketika latihan.

4.8 Kelebihan CapsNet

CapsNet diklaim memiliki empat kelebihan konseptual dibandingkan *convolutional neural networks* :

1. *Viewpoint invariance* : penggunaan matriks pose membolehkan jaringan CapsNet untuk mengenali objek-objek tanpa mempedulikan perspektif dari mana mereka dipandang.
2. Parameter lebih sedikit : *capsule* mengelompokkan neuron, sehingga koneksi antar lapisan membutuhkan lebih sedikit parameter.
3. Generalisasi lebih baik untuk sudut pandang baru : CapsNet lebih baik mengenali objek dari beberapa rotasi berbeda karena matriks *pose* menangkap karakteristik-karakteristik ini sebagai transformasi linear.
4. Dapat menangani *white-box adversarial attacks* : *Fast Gradient Sign Method* (FGSM) biasanya digunakan untuk menyerang CNN biasa dengan mengevaluasi gradien dari tiap pixel terhadap *loss* dari jaringan, lalu mengubah setiap pixel dengan maksimum epsilon (*error term*) untuk memaksimalkan *loss*. Metode ini dapat mengurangi akurasi CNN secara drastis (hingga di bawah 20 persen), namun CapsNet dapat mempertahankan akurasi di atas 70 persen.

5 Desain dan Implementasi Penambangan Data

6 Analisis

7 Kesimpulan

8 Dokumentasi

9 Daftar Pustaka

1. Cohen, J.P., Morrison, P., Dao, L.: Covid-19 image data collection. arXiv 2003.11597 (2020), <https://github.com/ieee8023/covid-chestxray-dataset> .
2. Bachir: Detecting COVID-19 in X-ray images with TensorFlow (2020), <https://www.kaggle.com/bachrr/detecting-covid-19-in-x-ray-images-with-tensorflow> .
3. Mooney, P.: Kaggle chest x-ray images (pneumonia) dataset (2018), <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia> .
4. Bachir: COVID-19 chest x-ray dataset (2020), <https://www.kaggle.com/bachrr/covid-chest-xray> .
5. Mangal, A., Kalia, S., Rajgopal, H., Rangarajan, K., Namboodiri, V., Banerjee, S., Arora, C. : CovidAID: COVID-19 Detection Using Chest X-Ray. arXiv:2004.09803v1 [eess.IV] (2020)
6. Zuraya, N. : Tiga Dampak Besar Pandemi Covid-19 bagi Ekonomi RI (2020). <https://republika.co.id/berita/qdgt5p383/tiga-dampak-besar-pandemi-covid19-bagi-ekonomi-ri>.
7. Hinton, G., Sabour, S., Frosst, N. : Dynamic Routing Between Capsules. arXiv:1710.09829v2 [cs.CV] (2017)
8. Hinton, G., Krizhevsky, A., Wang, S. : Transforming Auto-encoders.