

REDUKSI DATA BERLABEL PADA DETEKSI TUBERKULOSIS BERBASIS CITRA X-RAY MENGGUNAKAN *FRAMEWORK* SIMCLR

Majeri¹⁾, Fahrin Irhamna Rachman²⁾, Muhammad Faisal³⁾

1. Informatika, Universitas Muhammadiyah Makassar
email: 105841103622@student.unismuh.ac.id
2. Informatika, Universitas Muhammadiyah Makassar
email: fahrim141020@unismuh.ac.id
3. Informatika, Universitas Muhammadiyah Makassar
email: muhfaisal@unismuh.ac.id

Abstract

The development of deep learning-based Computer-Aided Diagnosis (CAD) for tuberculosis (TB) detection faces a fundamental challenge: high reliance on massive annotated data, which requires scarce radiological expertise, considerable time, and high costs. This study proposes a Self-Supervised Learning (SSL) approach via the SimCLR framework as a strategy to reduce labeled data requirements in X-Ray-based TB classification. The model utilizes a ResNet-50 encoder trained contrastively on unlabeled data using the NT-Xent Loss, followed by downstream adaptation via linear probing (SimCLR-LP) and fine-tuning (SimCLR-FT). Utilizing datasets from UPF BBKPM Makassar, evaluations were conducted across four labeled data fractions (10%, 25%, 50%, 100%). Results demonstrated that at the 10% fraction, SimCLR-LP achieved 85.50% accuracy and an AUC of 0.9091, significantly outperforming the Baseline model (62.60% accuracy) which suffered from degenerate prediction. The SimCLR-LP variant achieved $\geq 80\%$ accuracy using only 60 labeled images, whereas the Baseline required 303 images to reach a comparable threshold, demonstrating a fivefold labeling efficiency. Grad-CAM analysis confirmed that SimCLR-FT yielded localized activations in the perihilar and lower lung lobes, unlike the Baseline's scattered activations without anatomical focus.

Kata Kunci : Data Labeling Efficiency, Deep Learning, Self-Supervised Learning, SimCLR, Tuberculosis

A. PENDAHULUAN

Seiring perkembangan teknologi, algoritma *Computer-Aided Diagnosis* (CAD) berbasis *deep learning* menunjukkan potensi besar dalam meningkatkan kecepatan dan akurasi diagnostik, khususnya pada penanganan penyakit menular krusial seperti *tuberculosis* (TBC) yang masih menjadi penyebab utama kematian infeksius global

[1]. Namun, pengembangannya menghadapi kendala utama berupa ketergantungan pada dataset berskala besar yang memerlukan anotasi oleh ahli radiologi, di mana proses ini langka, mahal, dan memakan waktu [2], [3].

Paradigma konvensional umumnya mengatasi kelangkaan data ini melalui teknik *transfer learning* dari dataset alami seperti *ImageNet* [4]. Sayangnya,

pendekatan *transfer learning* konvensional ini memiliki keterbatasan akibat ketidaksesuaian domain (*domain mismatch*). Fitur visual yang dipelajari dari citra alami tidak selalu selaras dengan pola patologis halus pada citra medis, sehingga manfaat transfer fitur ini dapat menurun secara signifikan pada dataset medis yang sangat spesifik, meskipun metode ini diandalkan saat data terbatas[5].

Untuk menanggulangi inefisiensi pelabelan tersebut, *Self-Supervised Learning* (SSL) muncul sebagai solusi alternatif yang menjanjikan dengan memungkinkan model mempelajari representasi fitur kuat dari data tidak berlabel sebelum dilatih pada data berlabel [6]. Salah satu metode SSL yang menonjol adalah SimCLR, yang beroperasi dengan memaksimalkan kesamaan representasi antar-augmentasi dari citra yang sama dalam ruang laten. Pendekatan kontrastif ini memungkinkan model mengekstraksi fitur visual yang invarian dan diskriminatif tanpa supervisi manusia di tahap awal, yang secara efektif mereduksi ketergantungan pada anotasi massal [7].

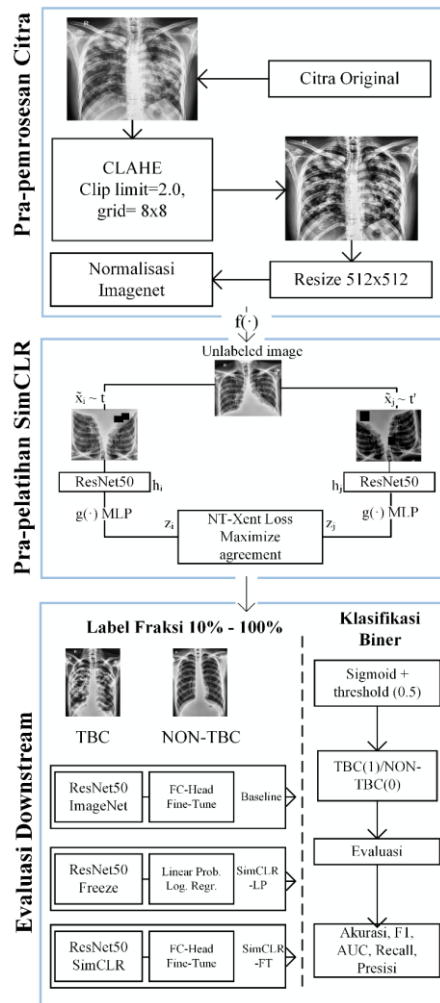
Efektivitas SimCLR telah dibuktikan mampu mempertahankan performa klasifikasi COVID-19 yang tinggi meski data berlabel rontgen dada direduksi hingga 40% [8]. Kajian terdahulu juga menegaskan bahwa model SSL yang dilatih langsung pada domain medis menghasilkan representasi yang lebih relevan dibanding inisialisasi citra alami [9], [10]. Hal ini disebabkan oleh perbedaan mendasar karakteristik visual antara citra alami dan citra radiografi X-Ray paru yang didominasi variasi tekstur halus serta pola densitas jaringan spesifik [11], [12]. Meskipun demikian, eksplorasi efisiensi pelabelan secara spesifik pada kasus klasifikasi TBC untuk memvalidasi apakah arsitektur *contrastive learning* mampu menyaingi model *supervised* penuh dengan fraksi data yang jauh lebih

sedikit masih menjadi celah penelitian yang perlu dikaji secara mendalam.

Penelitian ini bertujuan untuk menganalisis dan menguji efisiensi pelabelan kerangka kerja SimCLR (dibandingkan antara skenario Baseline, SimCLR-LP, dan SimCLR-FT) pada berbagai tingkat fraksi citra X-Ray paru (10%, 25%, 50%, dan 100%). Melalui pendekatan ini, penelitian diharapkan dapat membuktikan secara empiris sejauh mana teknik *Self-Supervised Learning* mampu menekan ketergantungan anotasi manual tanpa mengorbankan akurasi diagnostik serta interpretabilitas klinis pada sistem deteksi TBC.

B. METODE PENELITIAN

Penelitian eksperimental kuantitatif ini menguji efisiensi pelabelan klasifikasi tuberkulosis (TBC). Alur sistem (Gambar 1) meliputi tiga tahap: pra-pemrosesan citra, pra-pelatihan *encoder* SimCLR pada data tanpa label, dan evaluasi *downstream*.



Gambar 1. Diagram Alir Sistem Pra-Pelatihan SimCLR dan Evaluasi *Downstream* Klasifikasi TBC

Gambar 1 memperlihatkan bahwa alur sistem secara berurutan terbagi menjadi tiga blok utama dari atas ke bawah. Blok pertama (atas) merupakan tahap pra-pemrosesan yang mencakup pembagian set data, penerapan CLAHE, penyesuaian ukuran citra, dan normalisasi. Blok kedua (tengah) merepresentasikan fase Pra-pelatihan SimCLR yang beroperasi pada data tanpa label menggunakan mekanisme augmentasi kontrasif untuk menghasilkan dua pandangan citra, ekstraksi fitur melalui *encoder* ResNet-50, pemetaan pada *projection head* (MLP), dan optimalisasi dengan *NT-Xent Loss*. Selanjutnya, blok ketiga (bawah) adalah

fase Evaluasi *Downstream* di mana representasi diadaptasi ke tugas klasifikasi biner TBC/NON-TBC pada berbagai fraksi data latih menggunakan tiga strategi pemodelan yang dibandingkan secara paralel.

Dataset dan Pra-pemrosesan

Data X-Ray paru anonim dari arsip UPF BBKPM Makassar berjumlah 867 citra (461 TBC, 406 NON-TBC). Data dipartisi dengan rasio 70:15:15 menjadi himpunan pelatihan (606), validasi (130), dan pengujian (131). Fase pra-pelatihan SimCLR memanfaatkan 736 citra (pelatihan dan validasi) sebagai data kontrasif tak berlabel, sedangkan himpunan pengujian diisolasi ketat guna mencegah kebocoran data.

Pra-pemrosesan menerapkan *Contrast Limited Adaptive Histogram Equalization* (CLAHE) dengan *clip limit* 2,0 dan ukuran *grid* 8×8 piksel. Citra distandardisasi ke resolusi 512×512 piksel dan dinormalisasi mengikuti distribusi ImageNet. Terdapat dua *pipeline* augmentasi yaitu augmentasi kontrasif agresif untuk SimCLR (*random resized crop*, horizontal/vertikal *flip*, rotasi $\pm 15^\circ$, dan *coarse dropout*), serta augmentasi moderat dengan rotasi $\pm 10^\circ$ untuk fase *supervised fine-tuning*.

Sistem menggunakan *backbone encoder* ResNet-50. Pada pra-pelatihan SimCLR, *encoder* memanfaatkan *projection head* untuk memetakan fitur 2.048 dimensi menjadi ruang laten 128 dimensi via lapisan tersembunyi 512 neuron. Pada fase klasifikasi *downstream*, komponen ini diganti dengan *classification head* (2.048→256 neuron, *Dropout*, lalu *Sigmoid*). Tiga strategi model yang dievaluasi adalah:

1. **Baseline**, ResNet-50 dengan inialisasi bobot ImageNet, dilatih secara *full-supervised*.
2. **SimCLR-LP (Linear Probing)**, *Encoder* pra-pelatihan dibekukan,

dilanjutkan dengan pelatihan *Logistic Regression*.

3. **SimCLR-FT (Fine-Tuning)**, Seluruh parameter *encoder* dan *classification head* diperbarui melalui strategi pelatihan dua fase.

NT-Xent Loss

Dalam pra-pelatihan SSL, model dilatih untuk mengenali pasangan citra positif (augmentasi dari citra yang sama) di dalam sebuah *mini-batch* yang telah diencode. Optimalisasi fitur dilakukan dengan meminimalkan fungsi kerugian kontrastif (*Normalized Temperature-scaled Cross Entropy / NT-Xent*), di mana representasi citra yang sama ditarik mendekat sementara citra yang berbeda didorong menjauh [11]. Rumus fungsi kerugian tersebut secara spesifik didefinisikan sebagai berikut:

$$l_{i,j} = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} 1_{[k \neq i]} \exp(\text{sim}(z_i, z_k)/\tau)}$$

- $l_{i,j}$: Nilai *loss* kontrastif pasangan positif.
- $\text{sim}(\cdot)$: Kesamaan kosinus (*cosine similarity*).
- z_i, z_j : Representasi dari citra yang sama (*positive pair*).
- z_k : Representasi dari citra berbeda (*negative samples*).
- τ : Parameter *temperature*.
- N : Jumlah citra dalam satu *batch*.
- $1_{[k \neq i]}$: Fungsi indikator (bernilai 1 jika $1_{[k \neq i]}$).

Sesuai dengan fungsi tersebut, *NT-Xent* dioptimasi pada pra-pelatihan dengan suhu $\tau = 0,07$. Seluruh fase menggunakan *optimizer AdamW* (*learning rate* 5×10^{-4} , *weight decay* 10^{-4}). Pra-pelatihan dieksekusi dengan *batch size* 32 dan *seed* 42, serta dihentikan (*early*

stopping) jika kemiripan kosinus (Sim+) melampaui 0,88 guna mencegah *representational collapse*.

Eksperimen Efisiensi Pelabelan

Guna mengukur efisiensi pelabelan dan ketahanan model pada kondisi kelangkaan data medis, himpunan data pelatihan (606 citra) direduksi secara sistematis ke dalam empat skenario fraksi, yakni 10%, 25%, 50%, dan 100%. Reduksi data dilakukan secara proporsional (*stratified*) untuk memastikan keseimbangan rasio antara kelas TBC dan NON-TBC tetap terjaga pada setiap skenario. Himpunan data pengujian diisolasi dan dijaga konstan sebanyak 131 citra pada seluruh tahap evaluasi guna menjamin objektivitas perbandingan kinerja model. Rincian distribusi data berlabel disajikan pada Tabel 1.

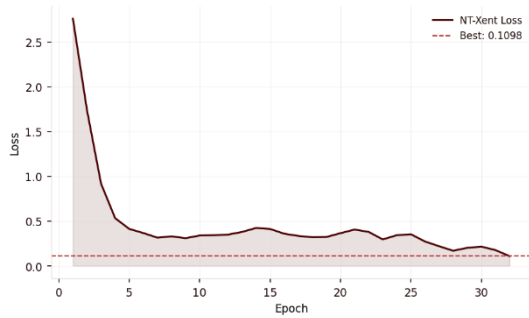
Tabel 1. Distribusi jumlah data berlabel pada skenario evaluasi *downstream*

Fraksi	Data latih			Total Uji
	Total	TBC	NON-TBC	
10%	60	32	28	131
25%	151	80	71	131
50%	303	161	142	131
100%	606	322	284	131

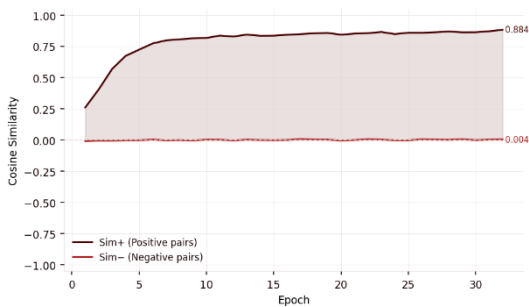
C. HASIL DAN PEMBAHASAN

Kinerja Pra-Pelatihan SimCLR

Fase pra-pelatihan SimCLR dievaluasi berdasarkan pergerakan fungsi kerugian dan kemiripan representasi. Gambar 2 menampilkan kurva penurunan *NT-Xent Loss* dan Gambar 3 menyajikan dinamika *Cosine Similarity* selama proses pelatihan.



Gambar 2. Kurva penurunan *NT-Xent Loss* selama pra-pelatihan SimCLR



Gambar 3. Kurva *Cosine Similarity* untuk pasangan positif (Sim+) dan negatif (Sim-)

Berdasarkan Gambar 2, nilai *NT-Xent Loss* awal tercatat sebesar 2.7592 dan mengalami konvergensi yang responsif hingga mencapai *best loss* sebesar 0.1098. Sejalan dengan hal tersebut, Gambar 3 menunjukkan bahwa *cosine similarity* untuk pasangan positif (Sim+) meningkat agresif dari 0.261 pada *epoch* pertama menjadi 0.8835 pada *epoch* ke-32. Pencapaian ambang batas Sim+ > 0.88 ini memicu mekanisme *early stopping* guna mencegah terjadinya *representational collapse* pada arsitektur. Di sisi lain, *cosine similarity* untuk pasangan negatif (Sim-) tetap stabil mendekati nol (0.0041). Penurunan kerugian yang konsisten serta pemisahan tegas antara Sim+ dan Sim- mengonfirmasi bahwa *encoder* telah berhasil membentuk ruang laten diskriminatif yang invarian terhadap augmentasi visual, sekaligus menjauhkan representasi antar instansi yang berbeda.

Komparasi Kinerja Klasifikasi

Setelah kualitas *encoder* tervalidasi, kinerja tiga strategi model (Baseline, SimCLR-LP, dan SimCLR-FT) dievaluasi pada empat skenario fraksi data berlabel (10%, 25%, 50%, dan 100%). Ringkasan metrik evaluasi disajikan pada Tabel 1.

Tabel 2. Komparasi Metrik Evaluasi Kinerja Klasifikasi

Fraksi	Model	Akurasi	Presisi	Recall	F1 Score	AUC
10%	SimCLR-LP	0.8550	0.8696	0.8571	0.8633	0.9091
10%	Baseline	0.6260	0.5897	0.9857	0.7380	0.8591
10%	SimCLR-FT	0.7939	0.7654	0.8571	0.8212	0.9611
25%	SimCLR-LP	0.8397	0.8890	0.8000	0.8421	0.9091
25%	Baseline	0.6489	0.6034	1.0000	0.7527	0.9029
25%	SimCLR-FT	0.8397	0.8356	0.8714	0.8531	0.9244
50%	SimCLR-FT	0.8855	0.8730	0.9000	0.8936	0.9412
50%	SimCLR-LP	0.8550	0.9230	0.8286	0.8933	0.9319

Fraksi	Model	Akurasi	Presisi	Recall	F1-Score	AUC
50%	Baseline	0.8473	0.8205	0.9143	0.8649	0.416
10%	SimCLR-LP	0.9008	0.9130	0.9000	0.9065	0.403
10%	SimCLR-FT	0.8855	0.8667	0.9286	0.8966	0.363
10%	Baseline	0.8702	0.9344	0.8143	0.8702	0.548

Analisis Tabel 2 mengungkap disparitas kinerja yang signifikan pada kondisi *few-shot learning* (fraksi 10% dan 25%). Dengan hanya ketersediaan 60 citra latih (10%), model Baseline mencatatkan *Recall* ekstrem (0.9857) namun dengan Presisi yang sangat rendah (0.5897) dan Akurasi hanya 62.60%. Hal ini mengindikasikan bahwa inisialisasi ImageNet konvensional kehilangan kemampuan diskriminatifnya dan secara artifisial memprediksi sebagian besar sampel sebagai kelas positif (TBC) untuk menekan *error* latih. Pada fraksi 25%, fenomena bias kelas mayoritas (*degenerate prediction*) ini semakin memburuk dengan pencapaian *Recall* absolut (1.0000).

Sebaliknya, pendekatan SimCLR-LP menunjukkan ketahanan luar biasa pada fraksi 10%, menjaga keseimbangan optimal antara Presisi (0.8696) dan *Recall* (0.8571), serta menghasilkan F1-Score tertinggi (0.8633) di antara ketiga strategi pada skenario tersebut. Menariknya, pada kondisi sangat terbatas ini, *Linear Probing* mengungguli *Fine-Tuning* (Akurasi

SimCLR-FT hanya 0.7939). Hal ini menegaskan bahwa pembaruan bobot *encoder* secara penuh pada data latih yang minim rentan memicu *catastrophic forgetting* terhadap representasi invarian yang telah dipelajari selama SSL. Saat data latih ditingkatkan (50% dan 100%), kinerja ketiga model mulai konvergen, dengan SimCLR-FT mengambil alih keunggulan seiring dengan kapasitas adaptasi domain yang lebih leluasa.

Analisis Efisiensi Pelabelan

Untuk mengukur sejauh mana pra-pelatihan kontrastif mereduksi kebutuhan anotasi manual. Tabel 3 membandingkan fraksi data minimum yang dibutuhkan setiap model untuk mencapai ambang batas (*threshold*) akurasi tertentu. Tabel 4 menyoroti selisih performa absolut (ΔAUC) model SSL terhadap Baseline.

Tabel 3. Kebutuhan Data Berlabel Minimum per Target Akurasi

Target Kinerja	f_L P	f_F T	f_B L	Keterangan
Akurasi ≥ 0.80	10%	25%	50%	Kedua SSL jauh lebih hemat
Akurasi ≥ 0.85	10%	50%	-	Baseline tidak tercapai
Akurasi ≥ 0.90	100%	-	-	Semua model terbatas

Tabel 4. Selisih AUC (ΔAUC) SSL terhadap Baseline per Skenario Fraksi

Fraksi	L P	F T	B L	LP - BL	FT - BL	Kesimpulan
10%	0.9091	0.8961	0.8591	+0.0500	+0.0370	SSL unggul

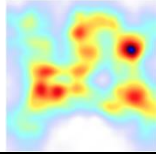
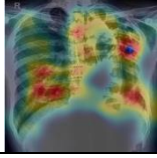
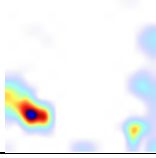
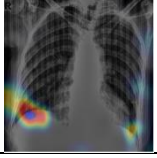
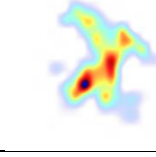
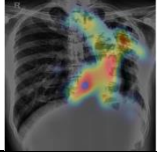
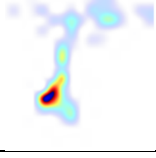
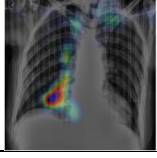
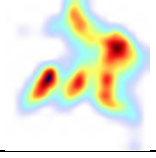
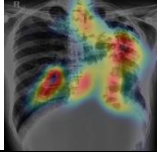
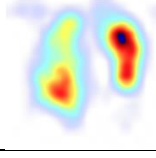
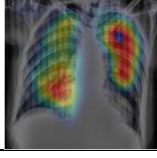
						signif ikan
25	0.9	0.9	0.9	+0.	+0.	SSL
%	09	24	02	006	021	masi
	1	4	9	2	5	h
						ungg
						ul
50	0.9	0.9	0.9	-	-	Ham
%	31	41	41	0.0	0.0	pir
	9	2	6	097	004	setara
10	0.9	0.9	0.9	-	-	Basel
0	40	36	54	0.0	0.0	ine
%	3	3	8	145	185	AUC
						tertin
						ggi

Analisis efisiensi pada Tabel 3 secara empiris membuktikan bahwa arsitektur berbasis SimCLR menghemat penggunaan data berlabel hingga lima kali lipat. Varian SimCLR hanya membutuhkan 60 anotasi manual (fraksi 10%) untuk melampaui akurasi 80%, sedangkan Baseline memerlukan setidaknya 303 anotasi (fraksi 50%). Tabel 3 memvalidasi temuan ini: keunggulan margin performa terbesar $\Delta AUC = +0.0500$) diraih oleh SimCLR-LP justru pada skenario data paling langka. Keunggulan tersebut menyusut seiring bertambahnya pasokan label, menegaskan bahwa paradigma SSL memiliki peran terapeutik yang paling krusial pada fasilitas kesehatan yang kekurangan sumber daya radiologis ahli.

Kualitas Representasi Spasial (Analisis Grad-CAM)

Guna memahami interpretasi keputusan model secara klinis, algoritma *Gradient-weighted Class Activation Mapping* (Grad-CAM) diekstraksi pada lapisan konvolusi terminal ResNet-50. Tabel 5 menyandingkan pola atensi spasial antara model pada kondisi kelangkaan data ekstrem (fraksi 10%).

Tabel 5. Komparasi Peta Aktivasi Spasial Grad-CAM (Fraksi 10%)

Model	Heatmap	Overlay
Baseline (TBC)		
		
Baseline (NONTBC)		
		
SimCLR-LP (TBC)		
		
SimCLR-LP (NON-TBC)		
		
SimCLR-FT (TBC)		
		
SimCLR-FT (NON-TBCTBC)		
		

Pada skenario data 10%, model Baseline mengeksibisi peta aktivasi yang terdistraksi dan memancar secara acak di luar area paru (seperti *costophrenic angle* dan jaringan lunak perifer). Ini mengonfirmasi bahwa penyesuaian bobot ImageNet ke domain radiografi gagal menangkap region patologi tanpa pasokan label yang masif.

Varian SimCLR-LP memperlihatkan lokalisasi atensi yang jauh lebih terfokus, membuktikan bahwa meskipun *encoder*

dibekukan sepenuhnya, ruang fitur yang dibentuk oleh pra-pelatihan kontrasif secara intrinsik telah sensitif terhadap opasitas anatomis paru. Adapun strategi SimCLR-FT mendemonstrasikan orientasi diagnosis yang paling persis, pada kelas positif TBC, model menavigasi aktivasi gradien maksimum pada zona perihilar dan lobus apikal bilateral, area yang secara histologis merupakan predileksi utama infiltrat *Mycobacterium tuberculosis*. Pada citra NON-TBC negatif, model menunjukkan atensi yang redup tanpa pola spesifik, mengesampingkan kekhawatiran terkait *false pattern recognition* [14].

D. KESIMPULAN DAN SARAN

Studi ini mendemonstrasikan bahwa *framework* SimCLR dengan *backbone* ResNet-50 secara efektif mereduksi ketergantungan pada anotasi berlabel dalam klasifikasi tuberkulosis berbasis citra X-Ray. Pada fraksi data berlabel minimum (10%, 60 citra), SimCLR-LP mencapai akurasi 85,50% dan F1-Score 86,33%, melampaui model Baseline (62,60%) yang mengalami *degenerate prediction* akibat inisialisasi ImageNet pada kondisi kelangkaan data. Temuan ini mengindikasikan bahwa pra-pelatihan kontrasif menghasilkan representasi fitur yang lebih *transferable* ke domain radiografi medis dibandingkan *transfer learning* konvensional, dengan efisiensi pelabelan lima kali lebih tinggi untuk mencapai ambang akurasi 80%. Validasi Grad-CAM lebih lanjut mengonfirmasi bahwa SimCLR-FT menghasilkan aktivasi yang terlokalisasi secara anatomis pada zona perihilar dan lobus bawah paru, berbeda dengan pola aktivasi Baseline yang tersebar tanpa korespondensi patologis.

Untuk penelitian lanjutan, disarankan perluasan korpus pra-pelatihan dari sumber *multi-center* guna meningkatkan

generalisasi lintas domain, eksplorasi strategi *semi-supervised learning* seperti *pseudo-labeling* dan *mean teacher* untuk optimasi utilisasi data tak berlabel, serta validasi klinis prospektif yang melibatkan spesialis radiologi sebagai prasyarat implementasi pada fasilitas kesehatan.

E. REFERENSI

- [1] S. Bagcchi, "WHO's Global Tuberculosis Report 2022," *Lancet Microbe*, vol. 4, no. 1, p. e20, Jan. 2023, doi: 10.1016/s2666-5247(22)00359-7.
- [2] L. Alzubaidi *et al.*, "Novel transfer learning approach for medical imaging with limited labeled data," *Cancers (Basel)*, vol. 13, no. 7, Apr. 2021, doi: 10.3390/cancers13071590.
- [3] I. Martinović *et al.*, "X-ray modalities in the era of artificial intelligence: overview of self-supervised learning approach," 2025, *Canadian Science Publishing*. doi: 10.1139/facets-2024-0229.
- [4] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, "Transfer learning for medical image classification: a literature review," Dec. 01, 2022, *BioMed Central Ltd*. doi: 10.1186/s12880-022-00793-7.
- [5] C. Matsoukas, J. F. Haslum, M. Sorkhei, M. Söderberg, and K. Smith, "What Makes Transfer Learning Work For Medical Images: Feature Reuse & Other Factors," Jun. 2022, [Online]. Available: <http://arxiv.org/abs/2203.01825>
- [6] J. Dominic *et al.*, "Improving Data-Efficiency and Robustness of Medical Imaging Segmentation Using Inpainting-Based Self-Supervised

- Learning,” *Bioengineering*, vol. 10, no. 2, Feb. 2023, doi: 10.3390/bioengineering10020207.
- [7] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A Simple Framework for Contrastive Learning of Visual Representations,” Jul. 2020, [Online]. Available: <http://arxiv.org/abs/2002.05709>
- [8] K. Imagawa and K. Shiimoto, “Evaluation of Effectiveness of Self-Supervised Learning in Chest X-Ray Imaging to Reduce Annotated Images,” *Journal of Imaging Informatics in Medicine*, vol. 37, no. 4, pp. 1618–1624, Aug. 2024, doi: 10.1007/s10278-024-00975-5.
- [9] S. Cheng, Z. A. Starosolski, and D. Subramanian, “Self-Supervised Learning for Building Robust Pediatric Chest X-ray Classification Models,” Aug. 2024, [Online]. Available: <http://arxiv.org/abs/2409.00231>
- [10] M. R. Hosseinzadeh Taher, F. Haghighi, R. Feng, M. B. Gotway, and J. Liang, “A Systematic Benchmarking Analysis of Transfer Learning for Medical Image Analysis,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Science and Business Media Deutschland GmbH, 2021, pp. 3–13. doi: 10.1007/978-3-030-87722-4_1.
- [11] S. Azizi *et al.*, “Big Self-Supervised Models Advance Medical Image Classification,” *Diagnostics*, vol. 15, no. 16, Apr. 2021, [Online]. Available: <http://arxiv.org/abs/2101.05224>
- [12] S. C. Huang, A. Pareek, M. Jensen, M. P. Lungren, S. Yeung, and A. S. Chaudhari, “Self-supervised learning for medical image classification: a systematic review and implementation guidelines,” Dec. 01, 2023, *Nature Research*. doi: 10.1038/s41746-023-00811-0.
- [13] L. Ericsson, H. Gouk, C. C. Loy, and T. M. Hospedales, “Self-Supervised Representation Learning: Introduction, Advances and Challenges,” Oct. 2021, doi: 10.1109/MSP.2021.3134634.
- [14] I. Matas *et al.*, “Mitigating Overfitting in Medical Imaging: Self-Supervised Pretraining vs. ImageNet Transfer Learning for Dermatological Diagnosis,” May 2025, [Online]. Available: <http://arxiv.org/abs/2505.16773>