

ABSTRAK

MAJERI, 2026. Efisiensi Pelabelan Data Pada Klasifikasi *Tuberculosis* Berbasis Citra X-Ray Menggunakan *Simple Contrastive Learning*. (dibimbing oleh Fahrir Irhamna Rachman, S.Kom., M.T, dan Ir. Muhammad Faisal, S.SI., M.T., Ph.D., IPM).

Pengembangan sistem *Computer-Aided Diagnosis* (CAD) untuk deteksi *tuberculosis* (TBC) sering terkendala oleh ketergantungan pada data medis berlabel dalam skala besar. Penelitian ini bertujuan mengevaluasi efisiensi pelabelan data pada klasifikasi citra X-Ray paru menggunakan pendekatan konseptual *Simple Contrastive Learning*, yang direalisasikan secara teknis melalui *framework* SimCLR dengan *encoder* ResNet-50. Evaluasi dilakukan dengan membandingkan tiga strategi: *Baseline* (inisialisasi ImageNet), SimCLR-LP (*Linear Probing*), dan SimCLR-FT (*Fine-Tuning*) pada variasi fraksi data berlabel 10%, 25%, 50%, dan 100% dari 606 citra latih yang tersedia. Pada kondisi data sangat terbatas, fraksi 10% yang hanya terdiri dari 60 citra berlabel, SimCLR-LP berhasil mencapai akurasi 85,50% dan F1-Score 86,33%, mengungguli *Baseline* yang turun ke akurasi 62,60% akibat kehilangan kemampuan diskriminatif. SimCLR-LP hanya membutuhkan 10% data berlabel untuk melampaui akurasi 80%, sedangkan *Baseline* memerlukan hingga 50%. Sebaliknya, SimCLR-FT unggul pada data berlabel penuh dengan sensitivitas tertinggi 92,86% pada fraksi 100%, menjadikannya strategi terbaik untuk skenario *screening* klinis. Analisis Grad-CAM membuktikan bahwa model berbasis representasi *contrastive* secara konsisten lebih berorientasi pada fitur patologi diagnostik yang relevan dibandingkan model *Baseline* yang cenderung mengaktifkan area non-diagnostik. Kesimpulannya, pendekatan *Simple Contrastive Learning* yang diimplementasikan melalui SimCLR terbukti efektif mengatasi keterbatasan anotasi medis dan berpotensi diterapkan pada dataset klinis yang lebih beragam di masa mendatang.

Kata Kunci: *Tuberculosis*, Citra X-Ray, *Simple Contrastive Learning*, SimCLR, Efisiensi Pelabelan