ANALISIS PERBANDINGAN CNN DENGAN COST-SENSITIVE LEARNING UNTUK DETEKSI TUBERKULOSIS PARU PADA CITRA X-RAY THORAX

Yusraka Dimas Al Iman[1)], Septine Eka Putri [2)], dan Agus Supriyanto[3)]
[1)] Program Studi Teknik Biomedis, Sekolah Tinggi Kesehatan Semarang
Jl. Kolonel Warsito Sugiarto KM. 2,5 Sadeng, Gunungpati, Semarang, Jawa Tengah
[2)] Program Studi Keselamatan dan Kesehatan Kerja, Sekolah Tinggi Kesehatan Semarang
Jl. Kolonel Warsito Sugiarto KM. 2,5 Sadeng, Gunungpati, Semarang, Jawa Tengah
[3)] Teknik Elektronika, SMK Pembangunan Nasional Purwodadi
Jl. Gajah Mada No.5, Jetis Selatan, Purwodadi, Grobogan, Jawa Tengah
e-mail: yusraka5@gmail.com[1)], septineekap5@gmail.com[2)], dan agussupriyanto.elektro@gmail.com[3)]

ABSTRACT

Tuberculosis (TB) remains one of the heaviest infectious disease burdens in Indonesia, with an estimated 1.09 million cases and 125,000 deaths each year, while access to radiologists in primary health facilities is limited. This study compares two EfficientNetB0 models for classifying chest X-ray images as normal or TB: a baseline model trained without class weighting and a cost-sensitive model that applies class weights of 0.6 for the normal class and 3.0 for the TB class. The Tuberculosis (TB) Chest X-ray Database, containing 3,500 normal and 700 TB images, was divided into 3,000 training, 950 validation, and 250 test images (200 normal, 50 TB). On the test set at a 0.50 threshold, the baseline model achieved 97.20% accuracy, 98.00% sensitivity, 97.00% specificity, 89.09% precision, and a 93.33% F1-score, with 6 false positives and 1 false negative. The cost-sensitive model achieved 98.80% accuracy, 98.00% sensitivity, 99.00% specificity, 96.08% precision, and a 97.03% F1-score, with 2 false positives and 1 false negative. The cost-sensitive model therefore performed better on this test set, mainly by reducing false positives from 6 to 2, while both models missed the same number of TB cases (1). The small test set, the single training run, and the absence of external validation limit these findings, so they need confirmation through repeated training with different random seeds, statistical testing, and external validation.

***Keywords**: chest X-ray, convolutional neural network, cost-sensitive learning, EfficientNetB0, tuberculosis*

ABSTRAK

Tuberkulosis (TB) tergolong beban penyakit infeksi terberat di Indonesia dengan estimasi 1,09 juta kasus dan 125.000 kematian setiap tahun, sementara ketersediaan radiolog di fasilitas kesehatan primer masih terbatas. Penelitian ini membandingkan dua model EfficientNetB0 untuk mengklasifikasikan citra X-ray toraks sebagai normal atau TB, yaitu model baseline tanpa pembobotan kelas dan model cost-sensitive dengan bobot kelas 0,6 untuk kelas normal dan 3,0 untuk kelas TB. Tuberculosis (TB) Chest X-ray Database yang memuat 3.500 citra normal dan 700 citra TB dibagi menjadi 3.000 citra latih, 950 citra validasi, dan 250 citra uji (200 normal, 50 TB). Pada data uji dengan ambang 0,50, model baseline memperoleh akurasi 97,20%, sensitivitas 98,00%, spesifisitas 97,00%, presisi 89,09%, dan F1-score 93,33% dengan 6 false positive dan 1 false negative. Model cost-sensitive memperoleh akurasi 98,80%, sensitivitas 98,00%, spesifisitas 99,00%, presisi 96,08%, dan F1-score 97,03% dengan 2 false positive dan 1 false negative. Model cost-sensitive lebih unggul pada data uji ini, terutama karena menurunkan false positive dari 6 menjadi 2, sedangkan jumlah false negative sama pada kedua model (1). Jumlah data uji yang kecil, pelatihan yang hanya dilakukan sekali, dan belum adanya validasi eksternal membatasi temuan ini, sehingga perlu dikonfirmasi melalui pelatihan ulang dengan beberapa random seed, uji statistik, dan validasi eksternal.

***Kata Kunci**: citra X-ray toraks, convolutional neural network, cost-sensitive learning, EfficientNetB0, tuberculosis*

I. PENDAHULUAN

TB paru tergolong penyebab kematian akibat infeksi terbesar di dunia. Insidensi global pada 2023 mencapai sekitar 10,8 juta kasus, naik tipis dari 10,7 juta kasus pada 2022. Indonesia menempati peringkat kedua setelah India dengan kontribusi sekitar 10 persen kasus dunia [1], [2]. Estimasi nasional mencatat 1,09 juta penderita dan sekitar 125.000 kematian setiap tahun, atau setara dengan kurang lebih 14 kematian setiap jam [1], [3].

Penderita menular yang belum memperoleh pengobatan dapat menularkan penyakit kepada kontak erat, sehingga penundaan diagnosis memperpanjang rantai penularan di lingkungan rumah, tempat kerja, maupun layanan primer [1]. Di Indonesia, pelaporan kasus yang belum lengkap juga menjadi tantangan [4]. Keputusan klinis untuk skrining TB tetap banyak bersandar pada pembacaan foto toraks, sementara tenaga pembaca yang terlatih belum tentu tersedia di setiap fasilitas kesehatan primer.

WHO melalui pedoman tahun 2021 menganjurkan pemanfaatan *computer-aided detection* (CAD) untuk skrining TB pada kelompok usia 15 tahun ke atas [5]. Evaluasi beberapa algoritma kecerdasan buatan untuk triase TB pada wilayah dengan beban TB tinggi juga telah dilaporkan [6]. Hal ini mempertegas pentingnya pengembangan sistem deteksi otomatis yang andal sekaligus dapat dipertanggungjawabkan secara klinis.

Convolutional Neural Network (CNN) menjadi pilihan utama untuk persoalan ini karena mampu

mempelajari pola spasial secara langsung dari piksel tanpa rekayasa fitur manual. Sejumlah tinjauan dan penelitian terdahulu melaporkan capaian yang menjanjikan pada foto toraks TB [6], [7], [8], [9]. Penelitian-penelitian tersebut umumnya menekankan akurasi keseluruhan, padahal dalam konteks skrining penyakit menular, bobot kedua jenis kesalahan tidaklah setara. Negatif palsu berisiko memulangkan penderita dalam kondisi masih menular dan belum tertangani, sedangkan positif palsu masih dapat dikoreksi melalui pemeriksaan lanjutan [5].

Cost-sensitive learning menjawab ketimpangan bobot kesalahan tersebut dengan memberikan penalti lebih besar pada kesalahan terhadap kelas TB. Pendekatan ini masih jarang diperbandingkan secara berdampingan dengan model tanpa pembobotan pada arsitektur yang sama. Kesenjangan itulah yang menjadi fokus penelitian ini.

Penelitian ini menyumbangkan tiga kontribusi. Pertama, perbandingan antara EfficientNetB0 *baseline* dan *cost-sensitive* pada pembagian data serta protokol pelatihan yang sama. Kedua, evaluasi yang tidak terbatas pada akurasi, tetapi juga memperhitungkan sensitivitas, spesifisitas, presisi, F1-score, ROC-AUC, dan analisis ambang keputusan. Ketiga, implementasi antarmuka yang menampilkan probabilitas prediksi beserta peta *Grad-CAM* sebagai alat bantu visual. Peta *Grad-CAM* pada penelitian ini hanya ditelaah secara kualitatif dan tidak dievaluasi secara kuantitatif.

II. TINJAUAN PUSTAKA

A. Tuberkulosis Paru dan Tantangan Deteksi Radiologis

Tuberkulosis paru bersumber dari bakteri *Mycobacterium tuberculosis* yang menyebar melalui percikan dahak dari penderita aktif. Infeksi ini menyerang saluran pernapasan dan berpotensi menjalar ke organ lain apabila tidak ditangani sejak awal. Gejala klinis pada stadium dini kerap samar sehingga sukar dibedakan dari infeksi saluran napas lainnya. Batuk berkepanjangan, demam, keringat malam, dan penurunan berat badan merupakan gejala yang mendorong pemeriksaan lanjutan [1], [5].

Foto toraks menjadi salah satu modalitas utama dalam skrining dan penegakan diagnosis TB karena ketersediaannya yang luas dan biayanya yang relatif terjangkau [5]. Gambaran radiologis TB beragam, dapat berupa infiltrat, nodul, kavitas, atau fibrosis, bergantung pada stadium penyakit dan status imunitas penderita. Keragaman ini menjadikan pembacaan foto toraks TB menuntut keahlian dan pengalaman klinis. Kondisi tersebut membuat sistem CAD berbasis deep learning relevan untuk dikembangkan, terutama di fasilitas yang tidak memiliki radiolog tetap [5], [6].

B. CNN dan Transfer Learning untuk Deteksi TB

Convolutional Neural Network merupakan arsitektur jaringan saraf tiruan yang dirancang untuk mengolah data berbentuk grid, termasuk citra digital. Lapisan konvolusi mengekstraksi fitur spasial secara bertingkat, mulai dari tepi dan tekstur pada lapisan awal sampai pola semantik tingkat tinggi pada lapisan yang lebih dalam. Kemampuan ini membuat CNN banyak dipakai pada deteksi TB dari foto toraks, sebagaimana dirangkum oleh beberapa tinjauan sistematis [7], [8].

Transfer learning memanfaatkan bobot hasil pembelajaran dari data berskala besar seperti *ImageNet* untuk mempercepat pelatihan pada data medis yang berukuran lebih kecil. Showkatian dkk. [9] melaporkan bahwa deteksi TB pada foto toraks dengan pendekatan deep learning dapat mencapai kinerja tinggi. Mirugwe dkk. [10] membandingkan beberapa arsitektur CNN berbasis transfer learning untuk deteksi TB. Witarsyah dkk. [11] mengevaluasi VGG16, DenseNet121, MobileNetV2, dan CNN kustom. Huy dan Lin [12] mengusulkan model DenseNet yang ditingkatkan. Pal dkk. [13] dan Goswami dkk. [14] juga memanfaatkan model CNN pralatih pada citra X-ray toraks.

EfficientNetB0 berasal dari pendekatan compound scaling, yaitu penskalaan serentak terhadap kedalaman, lebar, dan resolusi jaringan dengan rasio tetap. Pendekatan model yang ringkas juga dilaporkan dalam deteksi TB, misalnya model hibrida ringan oleh Owda dkk. [15] dan CNN yang disesuaikan oleh Rahman dkk. [16].

C. Cost-Sensitive Learning untuk Data Tidak Seimbang

Ketidakseimbangan kelas merupakan tantangan yang umum pada data medis. Pada data TB, jumlah citra normal biasanya melampaui jumlah citra TB. Model yang dilatih tanpa penanganan khusus cenderung condong pada kelas mayoritas sehingga sensitivitas terhadap kelas TB dapat menurun.

Cost-sensitive learning menanggulangi persoalan tersebut melalui modifikasi fungsi rugi agar kesalahan pada kelas minoritas diperhitungkan lebih berat. Berbeda dengan oversampling seperti SMOTE yang menambah citra sintetis, pembobotan kelas tidak menambah atau mengubah citra latih sehingga distribusi data fisik tetap sama.

D. Interpretabilitas Model dengan Grad-CAM

Gradient-weighted Class Activation Mapping (Grad-CAM) adalah teknik visualisasi yang menghasilkan peta panas berdasarkan gradien dari keluaran model menuju lapisan konvolusi terakhir [6]. Peta tersebut menandai area citra yang paling berkontribusi terhadap keputusan klasifikasi. Pada ranah medis, Grad-CAM dipakai untuk memeriksa apakah model memperhatikan area anatomis yang relevan atau justru artefak yang tidak bermakna secara klinis. Visualisasi ini merupakan

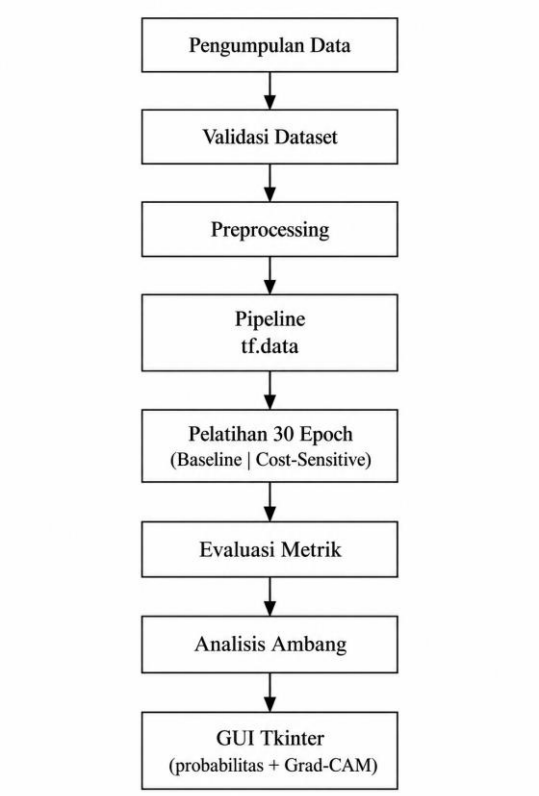
Jurnal Elektrika Vol. 18 No. 2 2026 P-ISSN 2085-0565 | E-ISSN 2580-8486
Received 23 Oct 2026, Revised 29 Oct 2026, Accepted 30 Oct 2026
DOI 10.26623/elektrika.v18i2.16277

alat bantu interpretasi dan tidak membuktikan bahwa model menggunakan ciri medis yang benar.

III. METODE PENELITIAN

A. Rancangan dan Alur Penelitian

Penelitian ini berbentuk eksperimen komputasional yang membandingkan dua skenario model CNN. Kedua model menggunakan arsitektur, pembagian data, dan protokol pelatihan yang sama sehingga perbedaan hasil terutama berkaitan dengan satu faktor perlakuan, yaitu pembobotan kelas. Gambar 1 menyajikan alur kerja penelitian mulai dari pengumpulan data sampai antarmuka pengguna.



Gambar 1. Diagram alir penelitian dari pengumpulan data sampai antarmuka pengguna.

B. Dataset

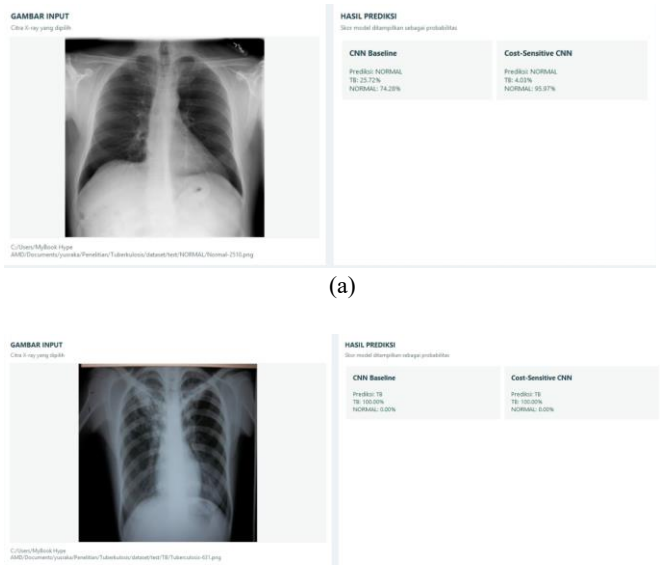
Data yang digunakan adalah Tuberculosis (TB) *Chest X-ray Database* hasil kompilasi Rahman dkk. dari Qatar University dan University of Dhaka. Data tersedia secara publik melalui platform Kaggle pada tautan: https://www.kaggle.com/datasets/tawsifurrahman/tuberculosis-tb-chest-xray-dataset

Kompilasi data ini merupakan hasil kolaborasi peneliti dari Qatar University, University of Dhaka, dan North South University bersama dokter dari Hamad Medical Corporation Qatar dan Bangladesh. Citra TB bersumber dari empat repositori, yaitu: (1) *Montgomery County X-ray Set* dan Shenzhen Hospital X-ray Set dari *National Library of Medicine* Amerika Serikat; (2) Belarus Set yang dihimpun untuk studi resistensi obat oleh *National Institute of Allergy and Infectious* Diseases; (3) sebagian citra dari NIAID TB Portal Program Dataset; serta (4) RSNA Chest X-ray Dataset.

Jumlah citra yang tersedia sebanyak 4.200, terdiri atas 3.500 citra Normal dan 700 citra TB. Rasio Normal terhadap TB sebesar 5:1 mencerminkan ketimpangan kelas sekaligus menjadi alasan penerapan *cost-sensitive learning*. Pembagian data dilakukan terpisah untuk setiap kelas menjadi 3.000 citra latih, 950 citra validasi, dan 250 citra uji. Proporsi TB pada subset latih, validasi, dan uji masing-masing sekitar 16,7%, 15,8%, dan 20,0%. Tabel I menyajikan sebaran citra hasil pembagian tersebut.

TABEL I
DISTRIBUSI CITRA MENURUT SUBSET DAN KELAS

Subset	Normal	TB	Total
Latih	2.500	500	3.000
Validasi	800	150	950
Uji	200	50	250
Total	3.500	700	4.200



(a)

(b)

Gambar 2. citra X-ray toraks dari Tuberculosis (TB) Chest X-ray Database: (a) kelas Normal, (b) kelas TB.

C. Pra-pemrosesan dan Pipeline Data

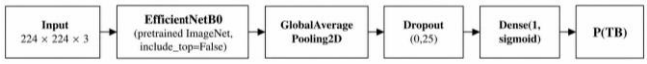
Setiap citra diubah ukurannya menjadi 224 × 224 piksel dengan 3 kanal RGB, lalu dikonversi menjadi tensor float32. Normalisasi memakai fungsi preprocessing bawaan EfficientNet. Augmentasi dibatasi pada pembalikan horizontal dan hanya diterapkan pada data latih, agar orientasi anatomi tetap terjaga dan model tidak mempelajari artefak augmentasi yang tidak bermakna klinis.

Pipeline data dibangun dengan tf.data. Data latih diacak pada setiap *epoch*, dikelompokkan dalam batch berukuran 32, dan di-prefetch supaya pembacaan berkas dari disk tidak menghambat komputasi GPU.

D. Arsitektur Model

Kedua model memakai EfficientNetB0 dengan konfigurasi include_top=False dan input_shape=(224,

224, 3). Bobot awal berasal dari pra-pelatihan pada ImageNet. Di atas blok pengekstrak fitur dipasang lapisan *GlobalAveragePooling2D*, *Dropout* dengan laju 0,25, dan satu unit Dense beraktivasi sigmoid yang menghasilkan probabilitas kelas TB pada rentang 0 sampai 1.



Gambar 3. Arsitektur EfficientNetB0 yang digunakan pada kedua skenario pelatihan.

E. Skenario Pelatihan

Dua skenario pelatihan dirancang untuk mengukur pengaruh pembobotan kelas. Skenario baseline dilatih tanpa pembobotan kelas dengan fungsi rugi Binary Crossentropy standar. Skenario cost-sensitive menambahkan pembobotan kelas yang dihitung melalui Persamaan (1):

$$w_c = \frac{N}{2\,N_c} \tag{1}$$

Pada Persamaan (1), N adalah jumlah total citra latih dan N_c adalah jumlah citra latih kelas c. Dengan 3.000 citra latih yang terdiri atas 2.500 citra Normal dan 500 citra TB, bobot kelas Normal bernilai 0,60 dan bobot kelas TB bernilai 3,00. Bobot kelas TB lima kali lebih besar daripada bobot kelas Normal sehingga kesalahan pada kelas TB diperhitungkan lebih berat oleh fungsi rugi.

F. Metrik Evaluasi

Setiap prediksi digolongkan ke dalam empat kategori, yaitu true positive (TP), true negative (TN), false positive (FP), dan false negative (FN). Metrik kinerja dihitung melalui Persamaan (2) sampai (6) [17], [18]:

$$\text{Akurasi} = (TP+TN)/(TP+TN+FP+FN) \tag{2}$$
$$\text{Sensitivitas} = TP/(TP+FN) \tag{3}$$
$$\text{Spesifisitas} = TN/(TN+FP) \tag{4}$$
$$\text{Presisi} = TP/(TP+FP) \tag{5}$$
$$F1 = 2\times(\text{Presisi}\times\text{Sensitivitas})/(\text{Presisi}+\text{Sensitivitas}) \tag{6}$$

Nilai *ROC-AUC* dihitung dari probabilitas keluaran model. Ambang keputusan diuji pada rentang 0,30 sampai 0,70 untuk memeriksa kestabilan keputusan model di luar nilai standar 0,50.

G. Interpretasi Visual dan Antarmuka

Grad-CAM dihasilkan pada lapisan konvolusi terakhir EfficientNetB0 untuk memvisualisasikan area yang paling memengaruhi keputusan model. Hasil evaluasi disimpan dalam berkas CSV dan JSON untuk analisis lanjutan. Antarmuka berbasis Tkinter memungkinkan pengguna mengunggah satu citra dan memperoleh probabilitas prediksi kedua model beserta peta Grad-CAM-nya secara berdampingan. Antarmuka memakai berkas model, pra-pemrosesan, dan ambang keputusan yang sama dengan eksperimen evaluasi. Keluaran antarmuka merupakan hasil prediksi model, bukan diagnosis klinis.

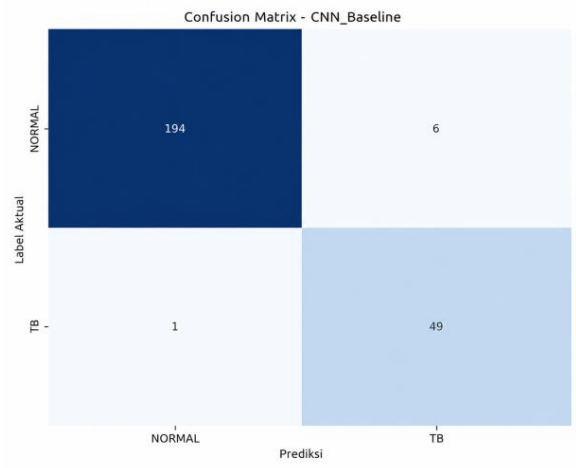
IV. HASIL DAN PEMBAHASAN

A. Gambaran Umum

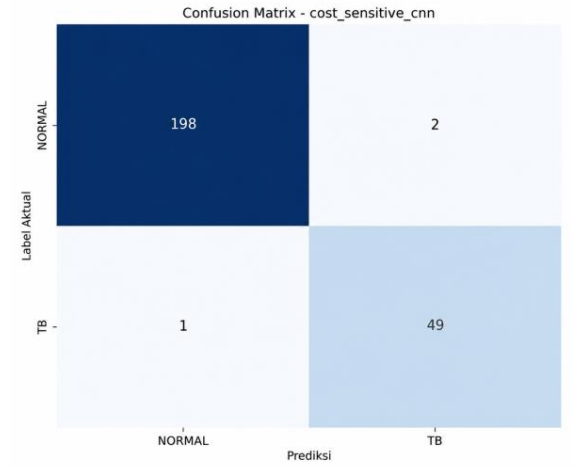
Penelitian ini membandingkan *CNN Baseline* dan *Cost-Sensitive CNN* untuk membedakan kelas Normal dan TB. Konfigurasi model dan pembagian data mengikuti Tabel I dan Tabel II. Perbedaan kedua model hanya pada strategi pelatihan. *CNN Baseline* dilatih tanpa pembobotan kelas. *Cost-Sensitive CNN* memakai bobot kelas 0,60 untuk Normal dan 3,00 untuk TB.

B. Hasil Evaluasi pada Data Uji

Tabel III membandingkan metrik kedua model pada 250 citra uji, sedangkan Gambar 4 menyajikan matriks konfusinya.



(a)



(b)

Gambar 4. Matriks konfusi pada 250 citra uji: (a) *CNN Baseline* (TN = 194, FP = 6, FN = 1, TP = 49), (b) *Cost-Sensitive* CNN (TN = 198, FP = 2, FN = 1, TP = 49).

Jurnal Elektrika Vol. 18 No. 2 2026 P-ISSN 2085-0565 | E-ISSN 2580-8486
Received 23 Oct 2026, Revised 29 Oct 2026, Accepted 30 Oct 2026
DOI 10.26623/elektrika.v18i2.16277

CNN Baseline mengenali 49 dari 50 citra TB (sensitivitas 98%) dan 194 dari 200 citra Normal (spesifisitas 97%). Terdapat 7 kesalahan, yaitu 6 citra Normal terdeteksi sebagai TB dan 1 citra TB terlewat sebagai Normal (Gambar 4a).

Presisi 89,09% lebih rendah daripada sensitivitas dan spesifisitas karena kelas TB hanya 20% dari data uji. 6 FP dibandingkan 49 TP berarti sekitar 11% prediksi TB adalah alarm palsu. Akurasi 97,20% tampak tinggi, tetapi F1-Score 93,33% menggambarkan kualitas deteksi kelas TB dengan lebih jujur.

Cost-Sensitive CNN mengenali 49 dari 50 citra TB (sensitivitas 98%) dan 198 dari 200 citra Normal (spesifisitas 99%). Terdapat 3 kesalahan, yaitu 2 citra Normal terdeteksi sebagai TB dan 1 citra TB terlewat sebagai Normal (Gambar 4b). Akurasi mencapai 98,80%, presisi 96,08%, dan F1-Score 97,03%. Sekitar 4% prediksi TB model ini adalah alarm palsu (2 FP dibandingkan 49 TP).

C. Perbandingan Kinerja

Cost-Sensitive CNN unggul pada akurasi, spesifisitas, presisi, dan F1-Score, sedangkan sensitivitas kedua model sama (98,00%). Selisih terbesar ada pada presisi (6,99 poin persentase), diikuti F1-Score (3,70 poin), spesifisitas (2,00 poin), dan akurasi (1,60 poin). Selisih ini berasal dari penurunan FP dari 6 pada baseline menjadi 2 pada *Cost-Sensitive* CNN, sedangkan FN sama-sama berjumlah 1.

D. Pola Kesalahan Baseline

Baseline memiliki sensitivitas tinggi (98%) tetapi FP-nya lebih banyak daripada FN. Dalam konteks skrining, FN lebih berbahaya karena kasus TB terlewat, sedangkan FP masih dapat dikoreksi melalui pemeriksaan lanjutan [5]. Baseline cukup baik pada sensitivitas, tetapi kurang efisien pada presisi.

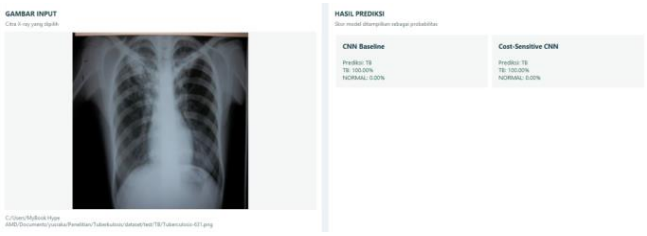
Cost-Sensitive CNN menurunkan FP dari 6 menjadi 2, tetapi tetap melewatkan 1 citra TB seperti baseline. Dengan demikian, risiko klinis yang paling penting dalam skrining, yaitu kasus TB yang terlewat, tidak berkurang pada ambang 0,50.

E. Pengaruh Class Weight

Class weight membuat kesalahan pada kelas TB berkontribusi lebih besar pada loss. Pada ambang 0,50, sensitivitas kedua model sama (98,00%), sehingga pembobotan kelas tidak meningkatkan sensitivitas pada penelitian ini. Perbedaan yang muncul ada pada FP, yang turun dari 6 menjadi 2, sehingga spesifisitas naik dari 97,00% ke 99,00% dan presisi naik dari 89,09% ke 96,08%. Hasil ini tidak sejalan dengan perkiraan umum bahwa pembobotan kelas menaikkan sensitivitas dengan risiko FP bertambah. Karena itu, penurunan FP tidak dapat langsung dikaitkan dengan pembobotan kelas. Penyebabnya dapat berasal dari inisialisasi bobot, urutan data, augmentasi, dan proses optimasi. Seed diatur satu kali sebelum kedua pelatihan, bukan diulang sebelum masing-masing model. Dengan satu kali pelatihan, selisih ini belum dapat dibedakan dari variasi acak.

F. Probabilitas Keluaran pada Antarmuka

Gambar 5 dan Gambar 6 menunjukkan tampilan antarmuka untuk dua citra uji. Pada citra TB (Gambar 5), kedua model memprediksi TB dengan probabilitas yang tampil sebagai 100,00%. Pada citra Normal (Gambar 6), kedua model memprediksi Normal, tetapi probabilitas TB *baseline* sebesar 25,72% sedangkan *Cost-Sensitive* CNN sebesar 4,03%. Pada citra *ini Cost-Sensitive* CNN memberi skor TB yang lebih jauh dari ambang 0,50. Contoh ini hanya satu citra sehingga tidak dapat digeneralisasi. Perbandingan sistematis dapat dilakukan dengan membandingkan distribusi skor TB kedua model pada seluruh data uji.

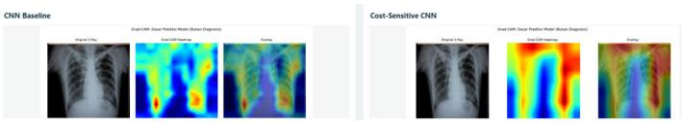


Gambar 5. Antarmuka pada citra uji TB (Tuberculosis-631): CNN Baseline dan Cost-Sensitive CNN sama-sama memprediksi TB.

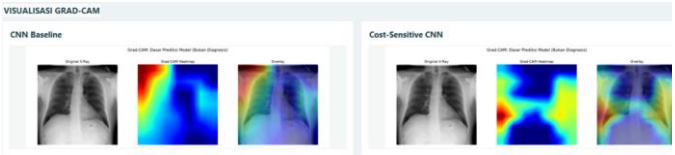
Gambar 6. Antarmuka pada citra uji Normal (Normal-2510): kedua model memprediksi Normal, dengan probabilitas TB 25,72% (*CNN Baseline*) dan 4,03% *(Cost-Sensitive CNN*).

G. Visualisasi Grad-CAM

Gambar 7 dan Gambar 8 menampilkan Grad-CAM kedua model pada citra TB dan citra Normal yang sama dengan Gambar 5 dan Gambar 6. Gambar 9 memuat contoh tambahan hasil Grad-CAM.



Gambar 7. Grad-CAM pada citra uji TB (Tuberculosis-631): (kiri) CNN Baseline, (kanan) Cost-Sensitive CNN. Setiap panel terdiri atas citra asli, peta panas, dan overlay.

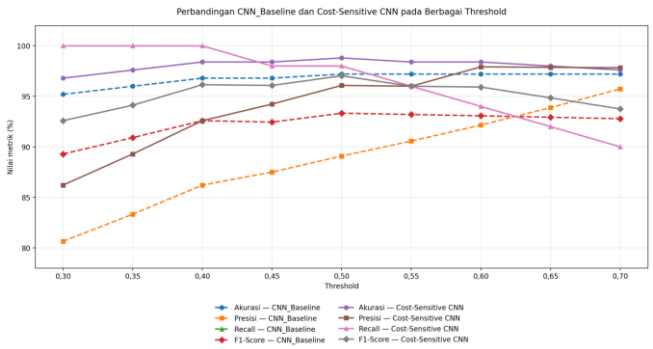


Gambar 8. Grad-CAM pada citra uji Normal (Normal-2510): (kiri) CNN Baseline, (kanan) Cost-Sensitive CNN.

Peta panas pada kedua model tidak selalu berada di lapang paru. Pada Gambar 7 dan Gambar 8, area aktivasi tinggi sebagian berada di tepi citra, di sekitar bahu, dan di luar lapang paru. Temuan kualitatif ini menunjukkan bahwa prediksi yang benar belum tentu bertumpu pada area paru. Karena itu Grad-CAM dalam penelitian ini diperlakukan sebagai alat bantu telaah, bukan bukti bahwa model memakai ciri medis yang benar. Analisis yang lebih sistematis, misalnya dengan anotasi lesi oleh radiolog atau pemotongan area paru sebelum klasifikasi, diperlukan pada penelitian lanjutan.

H. Threshold dan Probabilitas Keluaran

Kedua model menghasilkan probabilitas TB melalui neuron sigmoid, dan nilainya dapat berbeda untuk citra yang sama, seperti terlihat pada Gambar 6. Perbedaan kecil di sekitar 0,50 dapat membalik kelas akhir. Karena kedua model memiliki kesalahan pada ambang 0,50, metrik keduanya diperkirakan berubah pada ambang lain. Menaikkan ambang cenderung menurunkan FP tetapi dapat menambah FN. Tabel IV melaporkan hasil analisis ambang 0,30 sampai 0,70 untuk kedua model.



Gambar 9. Grafik pada Berbagai Ambang Keputusan

Gambar 9 menyajikan tren seluruh metrik terhadap ambang, pada kedua model menaikkan ambang menurunkan FP tetapi menambah FN. F1-Score tertinggi baseline (93,33%) dicapai pada ambang 0,50. F1-Score tertinggi Cost-Sensitive CNN (97,03%) dan akurasi tertingginya (98,80%) juga dicapai pada ambang 0,50.

Pada ambang 0,30 sampai 0,40, kedua model mencapai sensitivitas 100%, tetapi baseline menghasilkan lebih banyak FP (12, 10, dan 8) daripada Cost-Sensitive CNN (8, 6, dan 4). Pada ambang 0,45 sampai 0,70, sensitivitas kedua model identik dan turun dari 98% menjadi 90%. Pada rentang ini, FN kedua model bertambah dari 1 menjadi 5. Cost-Sensitive CNN memiliki F1-Score lebih tinggi daripada baseline pada seluruh ambang 0,30 sampai 0,70, dengan selisih antara 0,97 dan 3,70 poin persentase. Selisih terkecil terjadi pada ambang 0,70 dan selisih terbesar pada ambang 0,50. Keunggulan ini berasal dari FP yang lebih sedikit, bukan dari sensitivitas yang lebih tinggi.

Hasil ini menunjukkan bahwa kinerja Cost-Sensitive CNN pada ambang 0,50 tidak sempurna dan berubah seiring ambang. Jika prioritas skrining adalah tidak melewatkan kasus TB, ambang 0,30 sampai 0,40 menghasilkan FN = 0 pada kedua model, dengan FP yang lebih sedikit pada Cost-Sensitive CNN. Pemilihan ambang untuk penggunaan klinis sebaiknya ditentukan pada data validasi, bukan pada data uji.

V. KESIMPULAN

CNN Baseline mencapai akurasi 97,20%, sensitivitas 98,00%, spesifisitas 97,00%, presisi 89,09%, dan F1-Score 93,33%, dengan TP = 49, FN = 1, TN = 194, FP = 6. *Cost-Sensitive CNN* mencapai akurasi 98,80%, sensitivitas 98,00%, spesifisitas 99,00%, presisi 96,08%, dan F1-Score 97,03%, dengan TP = 49, FN = 1, TN = 198, FP = 2. Pada data uji ini, *Cost-Sensitive CNN* lebih unggul, terutama dalam menekan FP sehingga presisi dan F1-Score meningkat, sedangkan sensitivitas sama dengan *baseline*. Keunggulan F1-Score ini konsisten pada ambang 0,30 sampai 0,70. Telaah Grad-CAM menunjukkan bahwa area aktivasi tidak selalu berada di lapang paru. Klaim keunggulan tersebut perlu diperkuat dengan pengulangan pelatihan, uji statistik, validasi eksternal, dan analisis interpretabilitas yang lebih sistematis.

DAFTAR PUSTAKA

[1] WHO, *Global Tuberculosis Report 2024 (2024th ed)*. 2024. https://iris.who.int/server/api/core/bitstreams/7292c91e-ffb0-4cef-ac39-0200f06961ea/content

[2] H. Lee, J. Kim, J. Kim, and Y.-J. Park, "Review of the Global Burden of Tuberculosis in 2023: Insights from the WHO Global Tuberculosis Report 2024," *Jugan geon-gang gwa jilbyeong*, vol. 18, no. 11 Suppl, pp. 55–69, Mar. 2025, doi: 10.56786/PHWR.2025.18.11suppl.5.

[3] Suryanti and I. A. Ahmed, "Recent Review Tuberculosis in Indonesia: Burden and the Challenge of Under-Reporting.," *Iranian journal of public health*, vol. 54, no. 2, pp. 445–446, Feb. 2025. https://pmc.ncbi.nlm.nih.gov/articles/PMC11992917/

[4] K. L. Tobing *et al.*, "Under-reporting of Childhood Tuberculosis Cases in Indonesia and Associated Factors," *Journal of Preventive Medicine and Public Health*, vol. 59, no. 3, pp. 289–297, 2026, doi: 10.3961/jpmph.25.894.

[5] World Health Organization. WHO consolidated guidelines on tuberculosis. Module 2: screening –

Jurnal Elektrika Vol. 18 No. 2 2026 P-ISSN 2085-0565 | E-ISSN 2580-8486
Received 23 Oct 2026, Revised 29 Oct 2026, Accepted 30 Oct 2026
DOI 10.26623/elektrika.v18i2.16277

systematic screening for tuberculosis disease. Geneva: World Health Organization; 2021. https://www.who.int/publications/i/item/9789240022676.

[6] Z. Z. Qin *et al.*, "Tuberculosis detection from chest x-rays for triaging in a high tuberculosis-burden setting: an evaluation of five artificial intelligence algorithms," *The Lancet Digital Health*, vol. 3, no. 9, pp. e543–e554, 2021, doi: 10.1016/S2589-7500(21)00116-3.

[7] S. Hansun, A. Argha, S. T. Liaw, B. G. Celler, and G. B. Marks, "Machine and Deep Learning for Tuberculosis Detection on Chest X-Rays: Systematic Literature Review," *Journal of Medical Internet Research*, vol. 25, 2023, doi: 10.2196/43154.

[8] M. Oloko-Oba and S. Viriri, "A Systematic Review of Deep Learning Techniques for Tuberculosis Detection From Chest Radiograph," *Frontiers in Medicine*, vol. 9, 2022, doi: 10.3389/fmed.2022.830515.

[9] E. Showkatian, M. Salehi, H. Ghaffari, R. Reiazi, and N. Sadighi, "Deep learning-based automatic detection of tuberculosis disease in chest X-ray images," *Polish Journal of Radiology*, vol. 87, no. 1, pp. 118–124, 2022, doi: 10.5114/pjr.2022.113435.

[10] A. Mirugwe, L. Tamale, and J. Nyirenda, "Improving Tuberculosis Detection in Chest X-Ray Images Through Transfer Learning and Deep Learning: Comparative Study of Convolutional Neural Network Architectures," *JMIRx Med*, vol. 6, pp. e66029–e66029, 2025, doi: 10.2196/66029.

[11] D. Witarsyah, H. Almohab, R. S. Sianturi, and D. Syamsuar, "Enhancing Tuberculosis Detection from Chest X-Ray Images Using Deep Learning: Evaluating Multi-Architecture Performance and Efficiency," *Engineering, Technology and Applied Science Research*, vol. 16, no. 2, pp. 34365–34372, 2026, doi: 10.48084/etasr.16321.

[12] V. T. Q. Huy and C. M. Lin, "An Improved Densenet Deep Neural Network Model for Tuberculosis Detection Using Chest X-Ray Images," *IEEE Access*, vol. 11, pp. 42839–42849, 2023, doi: 10.1109/ACCESS.2023.3270774.

[13] V. Pal, H. Pabari, S. Indoria, S. Patel, D. Krishnan, and V. Ravi, "Multifaceted Disease Diagnosis: Leveraging Transfer Learning with Deep Convolutional Neural Networks on Chest X-Rays for COVID-19, Pneumonia, and Tuberculosis," *The Open Bioinformatics Journal*, vol. 17, no. 1, p. 55, 2024, doi: 10.2174/0118750362303182240516043224.

[14] K. K. Goswami, R. Kumar, R. Kumar, A. J. Reddy, and S. K. Goswami, "Deep Learning Classification of Tuberculosis Chest X-rays," *Cureus*, 2023, doi: 10.7759/cureus.41583..

[15] M. Owda, A. Abumihsan, A. Y. Owda, and M. Abumohsen, "A Lightweight Hybrid Deep Learning Model for Tuberculosis Detection from Chest X-Rays.," *Diagnostics (Basel, Switzerland)*, vol. 15, no. 24, Dec. 2025, doi: 10.3390/diagnostics15243216..

[16] K. K. Mujeeb Rahman, S. Zulaikha, B. Dhafer, and R. Ahmed, "Advancing tuberculosis screening: A tailored CNN approach for accurate chest X-ray analysis and practical clinical integration," *Intelligence-Based Medicine*, vol. 11, p. 100196, 2025, doi: https://doi.org/10.1016/j.ibmed.2024.100196.

[17] A. Supriyanto, R. R. Isnanto, and O. D. Nurhayati, "Klasifikasi penyakit daun kopi robusta menggunakan metode SVM dengan ekstraksi ciri GLCM", *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 12, no. 4, hal. 241-248, Nov. 2023. doi: https://doi.org/10.22146/jnteti.v12i4.8044

[18] Y. D. Al Iman, S. E. Putri, and A. Supriyanto, "Klasifikasi Pneumonia Pada Citra X-Ray Dada Menggunakan Ekstraksi Ciri Convolutional Neural Network dan Algoritma K-Nearest Neighbor", *Elektrika*, vol. 18, no. 1, pp. 31–39, Apr. 2026, doi: 10.26623/elektrika.v18i1.14211.