Optimasi Hyperparameter CNN Untuk Klasifikasi Penyakit Thyroid Berbasis USG

1st Dhia Farras Roshandi
Prodi Teknik Informatika
Telkom University Purwokerto
Purwokerto, Indonesia
dhiafarrasroshandi@student.telkomuniversity.ac.id

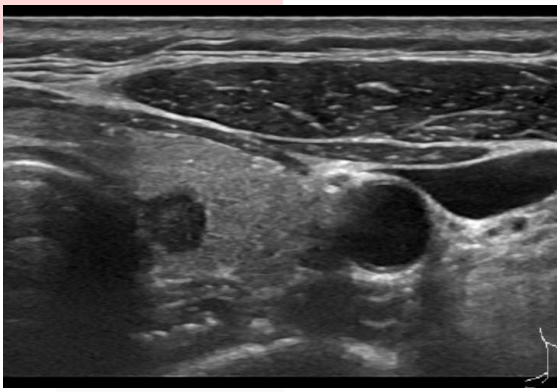
2nd Ridwan Pandiya
Fakultas Informatika
Telkom University Purwokerto
Purwokerto, Indonesia
ridwanp@telkomuniversity.ac.id

***Abstrak* — Dalam praktik antar dokter, interpretasi citra ultrasonografi (USG) kelenjar *thyroid* mungkin berbeda. Untuk membantu proses diagnosis, beberapa arsitektur *Deep Learning* seperti *Convolutional Neural Network* (CNN) diaplikasikan sebagai sistem pendukung keputusan dalam proses klasifikasi dan deteksi nodul *thyroid*. Namun, performa arsitektur CNN bergantung pada pemilihan *hyperparameter*. Karena kombinasinya yang luas dan beratnya upaya komputasi membuat metode *trial-and-error* digunakan, meskipun penentuan *hyperparameter* menjadi tidak efisien. Untuk mengatasi permasalahan tersebut, penelitian ini melibatkan metode optimasi dalam pemilihan *hyperparameter* optimum dengan tujuan meningkatkan kemampuan generalisasi model CNN (*EfficientNet-B3*). Dalam penelitian ini akan diimplementasikan tiga metode optimasi, yaitu *Particle Swarm Optimization* (PSO), *Genetic Algorithm* (GA), dan *Bayesian Optimization* (BO), yang diterapkan pada satu model CNN yang sama. Dataset yang digunakan adalah TN5000, yaitu dataset citra USG siap pakai yang dilengkapi dengan ROI *(Region of Interest)* anotasi *bounding box* nodul *thyroid*. Hasil uji coba menunjukkan bahwa ketiga metode optimasi mempengaruhi kemampuan generalisasi model CNN. Metode GA meraih skor AUC tertinggi sebesar 0.9486, disusul oleh BO sebesar 0.9479, dan PSO sebesar 0.9448. Meskipun model teroptimasi belum mampu melampaui skor metrik evaluasi utama model dasar, penelitian ini berkontribusi sebagai referensi terkait performa dan pengaruh optimasi *hyperparameter* pada model *Deep Learning* dalam sistem pendukung keputusan diagnosis citra USG *thyroid.*

Kata kunci*— CNN, PSO, GA, BO, klasifikasi, *thyroid

I. PENDAHULUAN

Penyakit *thyroid* merupakan gangguan kesehatan yang disebabkan oleh kelainan bentuk atau fungsi pada kelenjar *thyroid*, yang berperan penting sebagai produsen hormon pengatur metabolisme tubuh [1, 2]. Kelainan pada area tersebut dapat memicu gangguan hormonal sistemik, sehingga deteksi dini sangat diperlukan untuk mencegah risiko berkembangnya keganasan atau kanker [1, 3]. Salah satu metode penapisan awal non-invasif yang umum dianjurkan adalah diagnosis menggunakan citra ultrasonografi (USG) [3]. Metode diagnosis citra USG ini mampu menampilkan karakteristik fisik kelenjar *thyroid* secara mendetail, meliputi ukuran, bentuk, keberadaan nodul, hingga batas jaringan sebagai pertimbangan awal untuk mengidentifikasi struktur yang abnormal [7, 11, 22]. Representasi visual dari karakteristik nodul *thyroid* beserta visualisasi kompleksitasnya dapat diamati pada Gambar 1.1.



GAMBAR 1
(Citra ultrasonografi (USG) kelenjar *thyroid* yang memiliki noise bawaan (speckle noise) dan berkontras rendah)

Meskipun diagnosis citra USG mampu menampilkan informasi struktur nodul secara terperinci, proses analisisnya sangat bergantung pada keahlian dan pengalaman tenaga medis atau dokter spesialis radiologi [4]. Kendati demikian, faktor kelelahan fisik dan keterbatasan manusiawi sering kali memicu ketidkkonsistenan interpretasi hasil, baik oleh dokter yang sama di waktu berbeda maupun antar-dokter yang berbeda [4, 10]. Guna meminimalkan risiko variabilitas diagnosis tersebut, implementasi Machine Learning dan Deep Learning dikembangkan sebagai sistem pendukung keputusan (decision support system) untuk tugas klasifikasi citra medis [4, 7]. Perlu digarisbawahi bahwa sistem pendukung ini tidak bertujuan menggantikan peran dokter sepenuhnya, melainkan berfungsi sebagai alat penguat validitas diagnosis sebelum vonis akhir dijatuhkan [4, 7].

Di antara berbagai arsitektur kecerdasan buatan, Convolutional Neural Network (CNN) menjadi pilihan utama karena performanya yang bersaing dalam mengekstraksi fitur spasial kompleks serta kemampuannya menangani speckle noise (bintik noise) bawaan pada citra USG [7, 14, 15, 21]. Melalui lapisan konvolusinya, CNN mampu mengekstraksi fitur penting seperti tekstur dan batas pola nodul tanpa memerlukan konfigurasi manual [7]. Namun, tingkat keberhasilan arsitektur CNN sangat dipengaruhi oleh

pemilihan kombinasi hyperparameter yang digunakan [7, 14]. Secara praktis, penentuan hyperparameter ini sering kali masih mengandalkan metode trial-and-error, yang tidak efisien karena membutuhkan waktu komputasi yang besar dan belum tentu menghasilkan kombinasi yang optimal [3, 7, 14, 16].Untuk mengatasi kelemahan tersebut, penggunaan algoritma optimasi seperti Particle Swarm Optimization (PSO), Genetic Algorithm (GA), dan Bayesian Optimization (BO) diusulkan untuk mengeksplorasi ruang parameter secara otomatis guna memaksimalkan fungsi fitness [31, 32]. Beberapa penelitian terdahulu telah berupaya mengintegrasikan metode optimasi ini pada domain citra medis [31, 32]. Sebagai contoh, penelitian terdahulu menerapkan PSO untuk mengoptimalkan CNN pada klasifikasi citra USG [6]. Pendekatan GA juga telah digunakan untuk menyetel hyperparameter Ensemble Transfer Learning pada deteksi berbasis ultrasonografi [8], sementara BO dieksplorasi untuk merancang arsitektur CNN optimal secara otomatis pada klasifikasi lesi citra USG [12]. Secara umum, intervensi metode optimasi terbukti mampu menekan risiko overfitting serta meningkatkan kemampuan generalisasi model [9, 31, 32].

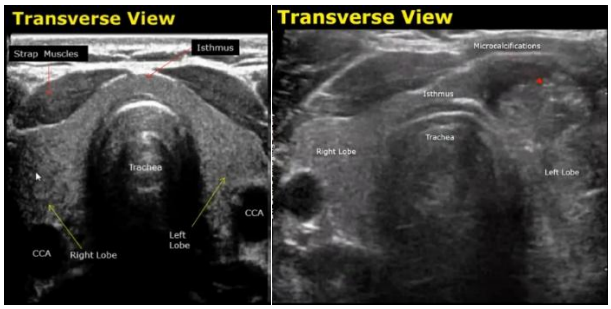
Namun, sebagian besar studi terdahulu masih terfokus pada implementasi satu metode optimasi tertentu, sehingga minim akan studi perbandingan yang setara [31, 32]. Berdasarkan celah tersebut, penelitian ini dilakukan untuk mengisi research gap melalui studi komparasi performa dan efisiensi antara algoritma PSO, GA, dan BO yang diterapkan pada satu model CNN dasar yang sama dalam konteks klasifikasi *thyroid* memanfaatkan dataset TN5000 [11]. Penelitian ini diharapkan dapat berkontribusi sebagai referensi ilmiah dalam pemilihan metode optimasi hyperparameter yang efektif pada domain analisis citra USG *thyroid*, sekaligus mendukung pengembangan sistem bantu analisis medis yang lebih akurat dan konsisten [1].

II. KAJIAN TEORI

Bab ini memaparkan konsep dasar dan landasan teoritis penelitian yang meliputi karakteristik citra USG kelenjar thyroid, arsitektur CNN, metode optimasi (PSO, GA, dan BO), serta deskripsi dataset yang digunakan.

A. Citra USG kelenjar *thyroid*

Kelenjar *thyroid* merupakan kelenjar endokrin di bagian leher yang berfungsi menjaga metabolisme tubuh melalui produksi hormon T3 dan T4 [1, 2]. Karena hormon tersebut terhubung dengan berbagai reseptor di organ vital seperti jantung, otak, dan ginjal, gangguan pada kelenjar ini dapat memicu risiko klinis sistemik yang serius [1, 3]. Mengingat kelainan nodul yang sering kali tidak disadari dan berpotensi berkembang menjadi kanker, deteksi dini melalui interpretasi citra ultrasonografi (USG) sangat direkomendasikan [3, 22]. Diagnosis klinis dilakukan dengan menganalisis karakteristik visual nodul seperti ukuran, bentuk, tingkat kecerahan, dan pola kalsifikasinya. Perbandingan visual antara kondisi kelenjar normal dan yang memiliki abnormalitas nodul dapat diamati pada Gambar 2.1.



GAMBAR 2
(Kondisi normal (kiri); terdapat abnormalitas (kanan))

Analisis visual yang biasa dilakukan oleh dokter spesialis radiologi tersebut kemudian ditransformasikan ke dalam arsitektur CNN agar model AI dapat mempelajari karakteristik klinis nodul secara mandiri [4, 7]. Karakteristik geometris seperti bentuk dan ukuran diekstraksi secara kuantitatif melalui koordinat *bounding box* (x_min, y_min, x_max, y_max) dari anotasi XML untuk menghitung dimensi *width*, *height*, *larger_size*, serta *sqrt_larger_size* [11]. Di sisi lain, karakteristik klinis seperti intensitas kecerahan dan pola kalsifikasi termanifestasi langsung di dalam matriks piksel mentah citra USG [7]. Data piksel tersebut didistribusikan ke *backbone* CNN untuk proses ekstraksi fitur otomatis [10]. Melalui lapisan konvolusinya, model mempelajari fitur visual berupa kontras warna, ketajaman tepi, hingga tekstur rumit untuk mengidentifikasi bintik mikrokalsifikasi. Integrasi antara koordinat geometris dan fitur visual piksel ini memungkinkan model meniru pola pengambilan keputusan klinis seorang radiolog secara valid [4, 7].

B. *EfficientNet-B3*

Arsitektur CNN EfficientNet-B3 digunakan sebagai model utama untuk menganalisis pola visual citra USG thyroid secara efektif [7]. Keunggulan arsitektur ini terletak pada mekanisme *compound scaling* yang menyeimbangkan tiga metrik utama (*depth, width*, dan resolusi) secara simultan sehingga model dapat konvergen dengan lebih efisien [27]. Dalam implementasinya, model dirancang untuk menerima masukan citra USG dari dataset TN5000 yang telah melalui tahap *preprocessing* dan *cropping* ROI *bounding box* [11]. Lapisan konvolusi pada EfficientNet-B3 secara otomatis mengekstraksi fitur kompleks berupa tekstur dan pola batas nodul tanpa rekayasa fitur manual [7, 14].

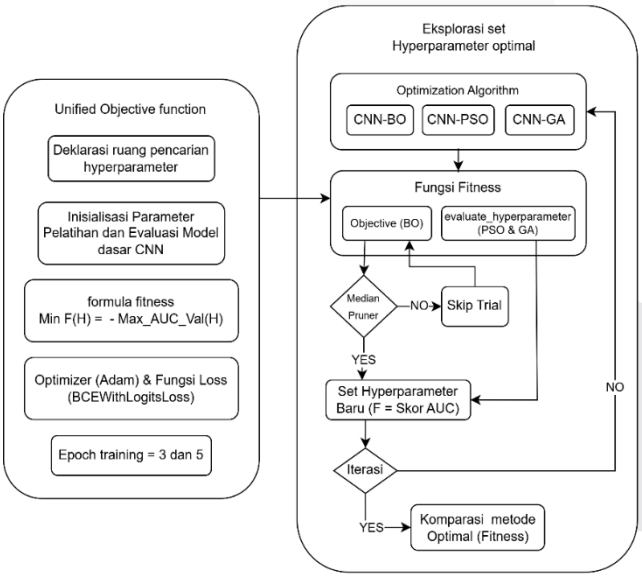
Pada bagian ekstraksi fitur, penerapan strategi *fine-tuning* penuh lewat konfigurasi freeze_features=FALSE dilakukan untuk memodifikasi kemampuan ekstraktor fitur bawaan (*pretrained weights*) agar adaptif terhadap karakteristik unik citra USG [10, 29]. Selanjutnya, struktur *fully connected layer* di bagian akhir arsitektur dirancang ulang menjadi *classifier hidden layer* kustom yang mengintegrasikan lapisan Linear, Layer Normalization, fungsi aktivasi GELU, dan *Dropout* [12]. Modifikasi ini memungkinkan model dasar bekerja lebih maksimal dalam klasifikasi biner dibandingkan menggunakan pengaturan bawaan arsitektur asli [12]. Dengan konfigurasi kustom tersebut, model dasar siap diintegrasikan dengan metode optimasi metaheuristik untuk mencari kombinasi *hyperparameter* paling optimal [31, 32].

C. Metode Optimasi (PSO, GA, dan BO)

Proses pencarian kombinasi *hyperparameter* optimal pada *classifier hidden layer* model *EfficientNet-B3* dilakukan secara komparatif menggunakan tiga algoritma dengan strategi eksplorasi yang berbeda:

1. *Particle Swarm Optimization* (PSO): Mengoptimalkan nilai *hyperparameter* dengan meniru pergerakan kawanan (*particles*) [6]. Posisi setiap partikel diperbarui secara iteratif berdasarkan koordinat terbaik lokal (pBest) dan global (gBest) untuk mencapai konvergensi cepat [6, 23, 25-26].
2. *Genetic Algorithm* (GA): Menerapkan prinsip seleksi alam melalui siklus generasi yang melibatkan tahapan evaluasi nilai fitness, seleksi individu terbaik, *crossover*, dan mutasi acak untuk menghindari jebakan lokal optimum [8, 13].
3. *Bayesian Optimization* (BO): Menggunakan pendekatan statistik probabilitas yang membangun surrogate model dan menggunakan fungsi akuisisi untuk memfokuskan pencarian secara efisien hanya pada ruang parameter yang berpotensi tinggi [9, 12].

Ketiga algoritma ini dibandingkan karena memiliki kesamaan sifat pendekatan, yaitu mengeksplorasi nilai optimal tanpa informasi gradien arsitektur CNN, beroperasi dalam ruang pencarian yang dibatasi, serta menggunakan optimasi nilai *fitness* sebagai tolak ukur [31, 32]. Studi komparatif ini penting untuk menunjukkan keunggulan masing-masing strategi eksplorasi secara objektif dalam kasus klasifikasi citra USG *thyroid*.



GAMBAR 3
(Ilustrasi skema *benchmarking* UOF untuk metode optimasi *hyperparameter)*

Untuk menjamin keadilan evaluasi, proses *benchmarking* dilakukan secara seragam menggunakan *Unified Objective Function* (UOF) sebagai fungsi *fitness* utama, seperti yang diilustrasikan pada Gambar 2.2. Melalui skema ini, setiap kombinasi parameter dinilai berdasarkan fungsi objektif terpadu yang menggabungkan skor metrik AUC (*Area under the ROC Curve*) dan nilai *loss* model dasar [17]. Pemilihan skor AUC didasarkan pada adanya masalah ketidakseimbangan kelas pada dataset citra *thyroid* [11]. Metrik yang bergantung pada jumlah distribusi data (*accuracy*) maupun ambang batas probabilitas (*precision, recall*, dan *F1-Score*) berisiko bias dan tidak cocok untuk pelatihan siklus pendek (3–5 epoch). Sebaliknya, metrik AUC memiliki kemampuan evaluasi diskriminatif yang kuat tanpa bergantung pada ambang batas probabilitas [17, 31]. Hasil akhir *benchmarking* ini menjadi umpan balik bagi ketiga algoritma untuk menentukan metode optimasi yang paling optimum dan stabil dalam meningkatkan generalisasi model.

D. Dataset Citra USG

Penelitian ini memanfaatkan dataset TN5000 dengan format anotasi PASCAL VOC untuk tugas deteksi dan klasifikasi nodul thyroid [11]. Dataset ini telah dilengkapi koordinat *Region of Interest* (ROI) berupa *bounding box* dari ahli medis, sehingga mempermudah tahapan *preprocessing* melalui ekstraksi objek (*cropping*) secara presisi tanpa perlu segmentasi awal [11]. Meski demikian, dataset ini menyimpan tantangan berupa pola tekstur nodul yang identik disertai *speckle noise* bawaan citra USG yang menuntut optimasi model secara presisi [3, 5, 7, 30]. Selain itu, terdapat kendala ketidakseimbangan kelas (*class imbalance*) akibat dominasi sampel nodul *malignant* (1) dibandingkan kelas *benign* (0), yang menjadi tantangan tersendiri dalam menjaga objektivitas performa klasifikasi model [24, 28]. Mempertimbangkan kompleksitas tersebut, dataset TN5000 sangat relevan untuk mengevaluasi ketangguhan metode optimasi *hyperparameter* pada klasifikasi citra USG thyroid.

III. METODE

Eksperimen berbasis komputasi dalam penelitian ini mencakup tahapan *preprocessing* dataset, konfigurasi arsitektur CNN, proses optimasi *hyperparameter*, hingga evaluasi komparatif model dasar dan model teroptimasi.

A. Tahap *preprocessing* dataset

Tahap awal difokuskan pada penyiapan citra ultrasonografi (USG) dari dataset TN5000 [11]. Pemanfaatan anotasi *bounding box* bawaan format PASCAL VOC pada dataset ini memungkinkan ekstraksi objek atau *Region of Interest* (ROI) dilakukan secara langsung melalui proses pemotongan (*cropping*) presisi tanpa memerlukan tahapan segmentasi ataupun deteksi awal [11]. Guna meminimalkan risiko *overfitting*, citra ROI diolah melalui serangkaian operasi *preprocessing* yang meliputi *resizing*, normalisasi nilai piksel, dan teknik augmentasi data dinamis. Citra kemudian dibagi menjadi tiga subset data independen dengan rasio 7:1:2, alokasinya yaitu 70% untuk *training set*, 10% untuk *validation set* (evaluasi internal), dan 20% untuk *testing set* (pengujian akhir) demi menjamin objektivitas evaluasi kemampuan generalisasi model.

B. Pengaturan model dasar CNN

Model dasar yang diimplementasikan adalah arsitektur EfficientNet-B3 karena efisiensi komputasinya yang ringan namun tetap mampu mempertahankan akurasi yang terkontrol pada domain citra medis [27]. Skema *transfer learning* diadopsi dengan memuat bobot bawaan ImageNet (*pretrained weights*) sebagai inisialisasi awal guna mengatasi

keterbatasan jumlah sampel data medis, mempercepat konvergensi, serta menekan beban komputasi [7, 27].

Model dikonfigurasi untuk menerima masukan citra berdimensi 3×224×224 piksel. Penelitian ini menerapkan strategi *fine-tuning* penuh dengan menonaktifkan pembekuan lapisan konvolusi (***freeze_features=FALSE***) agar seluruh filter dapat beradaptasi secara dinamis terhadap karakteristik visual unik nodul *thyroid* [29]. Modifikasi dilakukan khusus pada bagian *classifier head* dengan merancang struktur lapisan tersembunyi (*classifier hidden layer*) kustom yang terdiri atas lapisan Linear (128 neuron), *Layer Normalization*, fungsi aktivasi GELU, dan lapisan *Dropout*. Model dioptimasi menggunakan algoritma Adam dengan fungsi kerugian *Binary Cross Entropy with Logits Loss* (BCEWithLogitsLoss). Arsitektur kustom yang tervalidasi stabil ini ditetapkan sebagai model dasar (*base model*) untuk fungsi objektif pada tahapan optimasi *hyperparameter*.

C. Tahap optimasi *hyperparameter*

Untuk menjamin perbandingan performa yang adil (*fair comparison*) antaralgoritma metaheuristik, penelitian ini menerapkan metode *benchmarking* terpadu berbasis *Unified Objective Function* (UOF). Skema UOF mengondisikan agar Particle Swarm Optimization (PSO) [6], Genetic Algorithm (GA) [8], dan Bayesian Optimization (BO) [12] bergerak pada lingkungan komputasi, standarisasi fungsi, dan formula *fitness* yang identik. Ruang pencarian (*search space*) dibatasi di sekitar parameter model dasar untuk memfokuskan eksplorasi pada wilayah parameter berpotensi tinggi [31, 32], seperti dijabarkan pada Tabel 3.1.

TABEL 1
(Batas Rentang Ruang Pencarian Hyperparameter)

Hyperparameter	Nilai Model Dasar	Batas Rentang (Search Space)
Learning Rate	1×10^{-4}	1×10^{-5} - 1×10^{-3}
Batch Size	16	8 - 32
Dropout Rate	0.3	0.2 - 0.4
Weight Decay Rate	1×10^{-4}	1×10^{-5} - 1×10^{-3}

Nilai *fitness* dari setiap kombinasi parameter dievaluasi berdasarkan skor metrik *Area under the ROC Curve* (AUC) negatif [17]. Pendekatan minimasi fungsi objektif ini diadopsi agar algoritma optimasi mencari titik koordinat terendah, yang berkorelasi linear dengan pencapaian skor AUC tertinggi untuk memaksimalkan performa klasifikasi model:

$$\min f(H) = -best_AUC_val(H) \quad (1)$$

Di mana *f*(H) merupakan nilai fungsi objektif yang diminimalkan oleh algoritma metaheuristik, sedangkan AUC adalah nilai metrik evaluasi klasifikasi biner dari model dasar. Persamaan (1) ini diterapkan pada UOF secara seragam demi menjaga keadilan saat komparasi performa.

Sebagai strategi efisiensi waktu komputasi, durasi pelatihan pada fungsi objektif dibatasi sebesar 3 epoch untuk metode PSO [20] dan GA [18] demi menyeimbangkan luasnya populasi agen dengan durasi evaluasi per individu. Sementara itu, durasi pelatihan untuk metode BO yang berjalan pada *framework* Optuna [19] ditingkatkan menjadi 5 epoch dengan dukungan mekanisme pemangkasan otomatis (*pruning*) untuk memperkaya data statistik grafik performa secara akurat.

D. Fase pelatihan dan pengujian model CNN

Proses pelatihan model dasar dilakukan menggunakan citra ROI *thyroid* dari dataset TN5000. Namun, distribusi dataset ini memiliki kendala ketidakseimbangan kelas (*class imbalance*) yang signifikan akibat dominasi sampel label *malignant* (1) dibandingkan kelas *benign* (0). Untuk mengatasi potensi bias klasifikasi ini tanpa mengubah struktur format asli PASCAL VOC, metode sampling WeightedRandomSampling diterapkan sebelum data masuk ke dalam *batch* pelatihan [24, 28]. Metode ini memberikan pembobotan nilai yang lebih besar pada kelas minoritas (*benign*), sehingga probabilitas kemunculan sampel *benign* meningkat secara seimbang pada setiap iterasi pelatihan.

Alur komputasi pada setiap iterasi mencakup proses *forward pass* untuk menghasilkan nilai *logits* mentah, yang kemudian dievaluasi langsung oleh fungsi kerugian BCEWithLogitsLoss. Hasil evaluasi nilai *loss* tersebut ditransmisikan kembali melalui proses *backpropagation* untuk menghitung gradien *error*. Pembaruan bobot arsitektur model dasar dilakukan secara iteratif menggunakan optimizer Adam guna meminimalkan nilai *loss* menuju titik konvergen. Seluruh parameter pengontrol eksternal pada setiap algoritma metaheuristik diatur secara seragam guna menjaga *fairness* saat perbandingan efisiensi komputasi dan performa final dilakukan.

E. Evaluasi model dasar CNN dan model CNN teroptimasi

Setelah seluruh proses pencarian parameter selesai dijalankan, kombinasi *hyperparameter* terbaik dari metode yang menghasilkan skor AUC tertinggi akan diadopsi ke dalam arsitektur CNN untuk membentuk model teroptimasi final (*Optimized CNN*). Untuk membuktikan bahwa peningkatan performa pada model final ini benar-benar stabil dan bukan hasil kebetulan medis secara statistik, tahapan evaluasi akhir dibagi ke dalam dua pengujian utama:

1. **Uji Komprehensif Performa:** Menggunakan skema pemantauan kurva generalisasi di mana tiga subset data (*train, validation,* dan *test set*) diuji secara simultan pada setiap epoch untuk mendeteksi tanda-tanda *overfitting* atau *underfitting* serta menganalisis kestabilan grafik akurasi model.
2. **Uji Validasi Statistika:** Menggunakan pendekatan pelatihan berulang dan uji non-parametrik **Wilcoxon Signed-Rank Test**. Model dasar dan model teroptimasi akan dilatih ulang secara mandiri dari awal sebanyak ***n_repeat = 4*** kali. Skor AUC yang diperoleh dari setiap replikasi dikumpulkan sebagai sampel data untuk divalidasi lewat uji signifikansi Wilcoxon guna membuktikan konsistensi peningkatan performa secara matematis.

Sebagai analisis tambahan, metrik *Confusion Matrix* diimplementasikan pada data uji untuk memetakan distribusi *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Visualisasi ini digunakan untuk mengkaji secara mendalam pengaruh dari pergeseran nilai

hyperparameter terhadap sensitivitas dan spesifisitas model CNN dalam membedakan karakteristik klinis nodul *thyroid*.

IV. HASIL DAN PEMBAHASAN

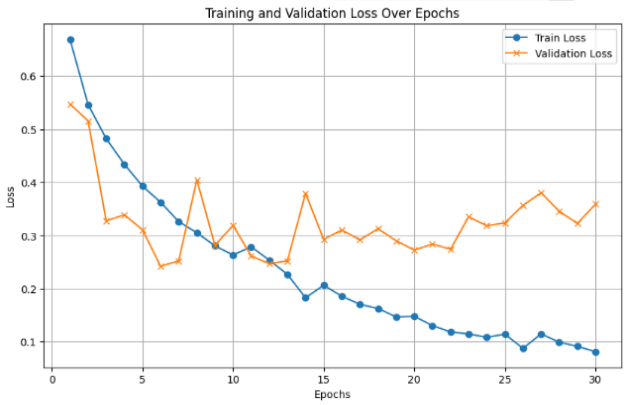
Bab ini menyajikan analisis objektif dan interpretasi data terhadap penemuan eksperimen, serta generalisasi hasil dari perbandingan performa model klasifikasi nodul thyroid.

A. Evaluasi performa awal model dasar CNN (*baseModel*)

Tahap awal penelitian berfokus pada pengujian arsitektur dasar EfficientNet-B3 yang dipilih karena efisiensinya dalam mengekstraksi fitur kompleks citra medis tanpa membebani komputasi secara berlebihan. Melalui skema *transfer learning* dari dataset ImageNet, bagian *classifier head* dikustomisasi menjadi lapisan linier (128 neuron), *Layer Normalization*, fungsi aktivasi GELU, dan *Dropout* = 0,3. Kustomisasi ini dirancang untuk mengatasi karakteristik citra USG nodul *thyroid* yang memiliki *speckle noise*, kabur, serta batas tekstur samar. Penggunaan fungsi aktivasi GELU berbasis probabilitas memberi kesempatan bagi fitur minor di bawah standar untuk tetap dihitung sehingga menekan risiko misklasifikasi. Untuk menangani ketidakseimbangan kelas pada format anotasi PASCAL VOC, metode *WeightedRandomSampling* diterapkan agar nodul minoritas (*benign*) mendapatkan frekuensi uji yang adil dalam setiap *batch* tanpa memicu *overfitting*. Hasil performa kuantitatif model dasar ini dirangkum dalam Tabel 4.1.

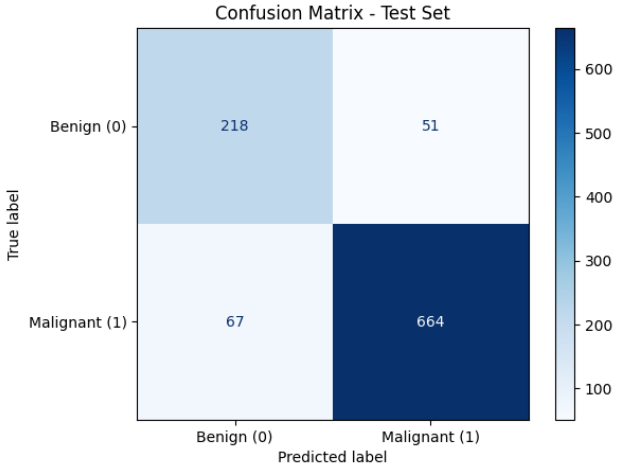
TABEL 2
(Hasil evaluasi awal model dasar CNN (*baseModel*))

	Loss	***AUC_Score***
val	0.2739	**0.9589**
train	**0.1183**	-
test	0.4141	0.9376



GAMBAR 4
(Grafik *loss* pada set pelatihan dan validasi citra USG (model dasar CNN))

Merujuk Tabel 4.1 dan Gambar 4.1, model dasar menunjukkan performa yang cukup tinggi dengan skor AUC validasi 0,9589 dan data uji 0,9376. Namun, visualisasi pelacakan *loss* menunjukkan kemampuan generalisasi model masih rendah; kurva *validation loss* mengalami fluktuasi tidak stabil dan stagnan sejak epoch ke-15, sementara kurva *train loss* terus menurun secara konsisten hingga epoch ke-30. Fenomena ini mengindikasikan adanya bias akibat model cenderung menghafal data latih (*overfitting*).



GAMBAR 5
(*Confusion Matrix-Test Set* model dasar CNN (*baseModel*))

Analisis performa awal klasifikasi biner (*benign* vs *malignant*) dipetakan secara detail menggunakan *confusion matrix* pada Gambar 4.2. Hasil deteksi pada data uji menunjukkan nilai *True Positive* sebesar 664, namun masih menyisakan 67 kasus *False Negative* dan 51 kasus *False Positive*. Dari sudut pandang medis, kekeliruan *False Negative* berakibat fatal karena menunda penanganan keganasan kanker, sedangkan *False Positive* memicu kecemasan pasien serta tindakan medis sekunder yang tidak diperlukan. Keterbatasan fungsional ini memperkuat urgensi dilakukannya optimasi nilai *hyperparameter*.

B. Implementasi dan komparasi metode optimasi *hyperparameter*

Proses pencarian kombinasi parameter dilakukan secara terbatas (*constrained optimization*) untuk menjaga stabilitas konvergensi model dengan rentang ruang pencarian sesuai metode bab sebelumnya [31, 32]. Pengujian performa ketiga algoritma metaheuristik—PSO, GA, dan BO—diintegrasikan melalui *Unified Objective Function* (UOF) dengan nilai *fitness* yang dievaluasi berdasarkan skor metrik AUC negatif [17].

Konfigurasi internal diatur berimbang: PSO berjalan dengan 25 partikel dan 10 iterasi, sedangkan GA menggunakan ukuran populasi 15 selama 10 generasi. Keduanya dibatasi sebesar 3 epoch per individu di dalam UOF untuk menyeimbangkan beban komputasi [18, 20]. Di sisi lain, BO yang berjalan sekuensial pada *framework* Optuna dikonfigurasi sebanyak 30 *trial* dengan durasi 5 epoch per *trial*, dibantu mekanisme *Median Pruner* untuk memotong *trial* yang tidak potensial [19]. Perbandingan parameter hasil optimasi disajikan pada Tabel 4.2 dan 4.3.

TABEL 3
(Perbandingan Parameter Teroptimasi dan Skor AUC Data Uji)

	baseModel	PSO
AUC_Score	0.9376	**0.9546**
Learning_rate	0.0001	0.00075
Batch_size	16	30
Dropout_rate	0.3	0.2203

Weight_decay	0.0001	0.00017

TABEL 4
(Perbandingan Parameter Teroptimasi dan Skor AUC Data Uji (contd.)

	BO	GA
AUC_Score	0.9474	0.9503
Learning_rate	0.000585	0.000292
Batch_size	19	20
Dropout_rate	0.2018	0.2406
Weight_decay	0.00004	0.000357

Merujuk Tabel 4.2 dan 4.3, metode PSO menempati peringkat pertama dengan skor AUC tertinggi sebesar 0,9546. Karakteristik penelusuran PSO membuktikan pergerakan kurva *best cost* yang melandai konsisten tanpa fluktuasi tajam selama ±12 jam eksplorasi, didorong oleh presisi penyesuaian vektor kecepatan partikel dan memori kelompok (*gbest*) [6, 23, 25, 26]. Metode GA berada di posisi kedua dengan skor AUC 0,9503 melalui waktu pencarian ±6 jam, mengandalkan evolusi biologi (*crossover* dan mutasi) yang luas namun dibebani oleh lambatnya konvergensi populasi penuh [8, 13]. Sementara itu, BO mencatatkan skor AUC paling rendah yaitu 0,9474, namun menjadi metode paling efisien dengan waktu eksplorasi hanya ±2 jam berkat fitur *pruning* yang sukses mengeliminasi titik performa rendah [9, 12, 19]. Analisis Optuna menegaskan bahwa *learning rate* memiliki tingkat signifikansi tertinggi (0,69), disusul oleh *weight decay* (0,17) dalam mempengaruhi performa akhir model.

C. Evaluasi akhir dan uji validasi (model dasar dan model teroptimasi)

Tahap konfirmasi dilakukan dengan melatih ulang model dasar secara penuh dan membandingkannya terhadap arsitektur hasil optimasi (CNN-PSO, CNN-GA, CNN-BO) menggunakan pengujian siklus penuh tiga set data. Untuk memperkuat validitas ilmiah, dilakukan uji statistik *Wilcoxon Signed-Rank* dengan skema pelatihan berulang (*n_repeat=4*). Rangkuman perbandingan metrik evaluasi akhir dipaparkan pada Tabel 4.4 dan 4.5.

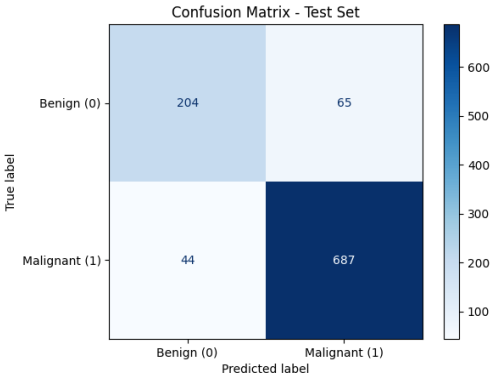
TABEL 5
(Hasil evaluasi model dasar CNN (*baseModel*) serta model teroptimasi dalam pelatihan dan evaluasi 3 set data (contd.))

	Loss	***AUC_Score***
baseModel	0.3263	0.9466
CNN-PSO	**0.2884**	0.9448
CNN-GA	0.3085	**0.9486**
CNN-BO	0.3635	0.9479

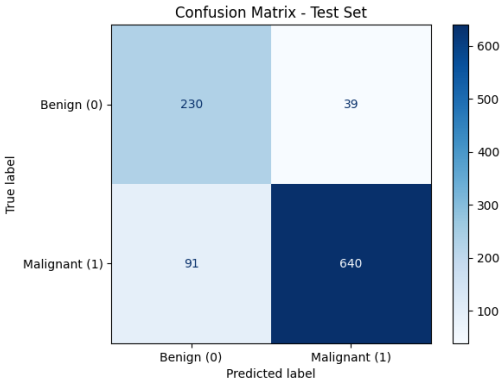
TABEL 6
(Hasil evaluasi model dasar CNN (*baseModel*) serta model teroptimasi dalam pelatihan dan evaluasi 3 set data)

	Accuracy	***Precsion***	***Recall***	***F1_Score***
baseModel	**0.9000**	0.9376	**0.9248**	**0.9311**
CNN-PSO	0.8860	0.9351	0.9070	0.9208
CNN-GA	0.8760	**0.9470**	0.8796	0.9121
CNN-BO	0.8670	0.9463	0.8673	0.9051

Hasil pengujian pada Tabel 4.4 dan 4.5 memperlihatkan *trade-off* metrik yang menarik. Model teroptimasi CNN-PSO mencatatkan nilai kekeliruan terkecil dengan capaian *test loss* terendah sebesar 0,2884. Kestabilan kurva performa CNN-PSO membuktikan kemampuannya dalam menekan tingkat *error* dan meredam gejala *overfitting* yang sebelumnya mendominasi model dasar. Kendati demikian, pencapaian metrik *recall* (0,9070) dan akurasi (0,8860) milik CNN-PSO belum mampu melampaui nilai milik baseModel. Fenomena penurunan metrik *recall* ini juga dialami secara lebih ekstrem oleh varian CNN-GA dan CNN-BO.



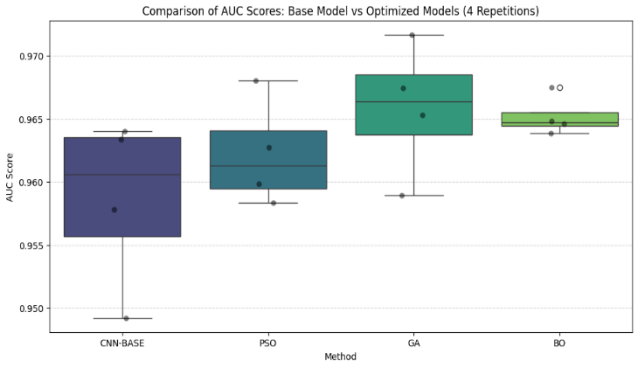
GAMBAR 6
(*Confusion matrix* model dasar)



GAMBAR 7
(*confusion matrix* model teroptimasi)

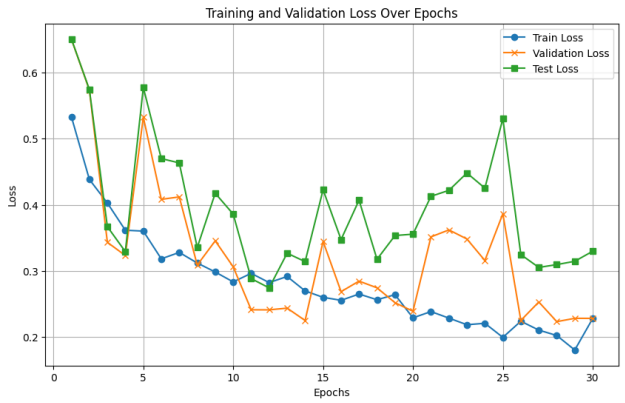
Berdasarkan *confusion matrix* evaluasi akhir dari Gambar 4.3 dan 4.4, model dasar menghasilkan nilai *False Negative* sebanyak 44 dan *False Positive* sebanyak 65. Sebaliknya, model CNN-PSO menghasilkan nilai *False Positive* yang lebih kecil (39) tetapi memicu lonjakan angka *False Negative* hingga 91. Ketimpangan sebaran ini terjadi karena tahapan optimasi *hyperparameter* melalui UOF lebih memprioritaskan pemisahan kelas dalam sebaran data (skor AUC) daripada akurasi *threshold* prediksi label bawaan 0,5 [11, 17, 31]. Oleh sebab itu, diperlukan kalibrasi ulang pasca-pelatihan pada nilai *threshold* prediksi probabilitas untuk

memangkas *False Negative* ke tingkat minimum sebelum model diintegrasikan ke sistem nyata.



GAMBAR 8
(rentang sebaran data performa seluruh model CNN setelah pelatihan sebanyak *n_repeat* = 4 kali)

Hasil uji validasi *Wilcoxon Signed-Rank* menghasilkan nilai *p-value* > 0,05 pada seluruh skenario perbandingan silang antarmodel. Secara inferensial, nilai tersebut menunjukkan tidak adanya perbedaan signifikansi yang mutlak, yang berarti seluruh varian model secara statistik berada pada koridor tingkat performa yang setara akibat batasan pengulangan sampel iterasi. Namun, visualisasi sebaran grafik *boxplot* pada Gambar 4.5 menangkap bukti krusial: sebaran nilai AUC milik baseModel memiliki rentang sangat lebar dengan batas bawah paling anjlok, menandakan kerentanan stabilitas yang tinggi. Implementasi optimasi terbukti sukses mendongkrak batas bawah tersebut. CNN-GA memimpin dengan raihan median nilai tertinggi, sedangkan CNN-BO mencatatkan konsistensi sebaran data yang paling rapat dan stabil.



GAMBAR 9
(Grafik performa *loss* dari pelatihan menyeluruh model CNN-PSO selama 30 *epoch*)

Di sisi lain, peningkatan performa pada Gambar 4.6 menguatkan posisi model CNN-PSO sebagai model terbaik. Meskipun belum mampu melampaui skor metrik evaluasi utama model dasar pada *threshold* standar, model CNN-PSO terbukti sukses menangani *overfitting* bawaan dari model dasar tanpa mengorbankan performa metrik *recall* dan *accuracy* secara drastis [24, 28]. Dengan demikian, arsitektur CNN-PSO dinyatakan unggul dalam menekan masalah *overfitting* dan memberikan jaminan pemisahan kelas yang konsisten untuk sistem pendukung keputusan diagnosis dini nodul tiroid ganas.

V. KESIMPULAN

Optimasi *hyperparameter* menggunakan *Unified Objective Function* (UOF) efektif meningkatkan performa dan mengatasi *overfitting* pada model CNN klasifikasi nodul thyroid. Hasil pengujian menunjukkan PSO menjadi metode paling optimal dengan skor AUC tertinggi 0,9546, disusul GA (AUC 0,9503), sementara BO memberikan efisiensi terbaik dengan durasi pencarian hanya ±2 jam (AUC 0,9474). Uji *Wilcoxon Signed-Rank* mengonfirmasi seluruh model teroptimasi memiliki stabilitas yang lebih baik daripada model dasar, di mana arsitektur CNN-PSO sukses menekan *error* dengan *test loss* terendah 0,2884. Lonjakan *False Negative* akibat *threshold* standar 0,5 dapat diatasi melalui kalibrasi ulang nilai *threshold* prediksi probabilitas pasca-pelatihan untuk meminimalkan risiko misdiagnosis klinis. Secara keseluruhan, tujuan merancang model klasifikasi citra USG thyroid yang optimal dan stabil telah sepenuhnya tercapai.

REFERENSI

[1] M. Armstrong, Edinen Asuka, dan A. Fingeret, “Physiology, Thyroid Function,”. *Treasure Island,* FL: *StatPearls Publishing, 2023.*

[2] Y. Pirahanchi, I. Jialal, dan F. Toro, “Physiology, Thyroid Stimulating Hormone (TSH),”. *Treasure Island,* FL: *StatPearls Publishing, 2023.*

[3] A. H. AlSaedi, D. S. Almalki, dan R. M. ElKady, “Approach to Thyroid Nodules: Diagnosis dan Treatment,” *Curēus*, vol. 16, 2024.

[4] Y. Xu, M. Xu, Z. Geng, Y. Zhao, Q. Zhang, dan C. Li, "Thyroid nodule classification in ultrasound imaging using deep transfer learning," BMC Cancer, vol. 25, 2025.

[5] O. Karaali, “Learning-Based Ultrasound Image Super-Resolution,” Sep. 2023.

[6] S. B. Porna, M. A. Rahman, S. Banik, M. H. Khan, R. Uddin, A. Basit, H. H. Uddin, G. D. Bhavani., “Novel CNN-Based Models for Breast Ultrasound Image Classification: Leveraging EEHO, PSO, dan BSA Optimizers,” *In Proc. 2025 International Conference on Knowledge Engineering and Communication Systems (ICKECS)*, Karnataka, India, 2025, pp. 1–6.

[7] C.-L. Cao, Q.-L. Li, J. Tong, L.-N. Shi, W.-X. Li, Y. Xu, J. Cheng, T.-T. Du, J. Li, X.-W. Cui, “Artificial Intelligence in Thyroid Ultrasound,” *Frontiers in Oncology*, vol. 13, 2023.

[8] A. Ghosh dan K. Srinivasan, “EffiDenseGenOp: Ensemble Transfer Learning with Hyperparameter Tuning Using Genetic Algorithm Optimization for PCOS Detection from Ultrasound Sonography Images,” *IEEE Access*, pp. 1–1, 2025.

[9] J. Hou, “Enhanced Automated Prostate Segmentation in Ultrasound Images Based on Diverse Pre-Processing Strategies and Multi- Input Architectures,” 2025.

[10] E. A. Agyekum, Z. Yuzhi, Y. Fang, J. M. S. Nyantakyi, dan S. Agbo-Quashie, "Ultrasound-based classification of follicular thyroid Cancer using deep convolutional neural networks with transfer learning," Sci. Rep., vol. 15, 2025.

[11] H. Zhang, Q. Liu, X. Han, L. Niu, dan W. Sun, “TN5000: an Ultrasound Image Dataset for Thyroid Nodule Detection and Classification,” *Scientific Data*, vol. 12, 2025.

[12] M. Ahmed, H. Du, dan A. AlZoubi, “ENAS-B: Combining ENAS with Bayesian Optimization for Automatic Design of Optimal CNN Architectures for Breast Lesion Classification from Ultrasound Images,” *Ultrasonic Imaging*, vol. 46, pp. 17–28, 2023.

[13] E. Michael, H. Ma, H. Li, dan S. Qi, “An Optimized Framework for Breast Cancer Classification Using Machine Learning,” *BioMed Research International*, vol. 2022, pp. 1–18, 2022.

[14] P. Purwono, A. Ma’arif, W. Rahmaniar, H. I. K. Fathurrahman, A. Z. K. Frisky, dan Q. M. ul Haq, “Understanding of Convolutional Neural Network (CNN): a Review,” *International Journal of Robotics and Control Systems*, vol. 2, pp. 739–748, 2023.

[15] J. A. Ayeni, “Convolutional Neural Network (CNN): the Architecture and Applications,” *Applied Journal of Physical Science*, vol. 4, pp. 42–50, 2022.

[16] Ohood Fadil Alwan, Qusay Kanaan Kadhim, Rasha Basim Issa, dan Shaymaa Taha Ahmed, “Early Detection and Segmentation of Ovarian Tumor Using Convolutional Neural Network with Ultrasound Imaging,” *Revue d’intelligence artificielle*, vol. 37, pp. 1503–1509, 2023.

[17] J. Li, "Area under the ROC Curve has the most consistent evaluation for binary classification," PLOS ONE, vol. 19, no. 12, p. e0316019, Des. 2024. https://doi.org/10.1371/journal.pone.0316019

[18] F.-A. Fortin, F.-M. De-Rainville, M.-A. Gardner, M. Parizeau, dan C. Gagne, “DEAP: Evolutionary Algorithms Made Easy Félix-Antoine Fortin François-Michel De Rainville Christian Gagné,” *Journal of Machine Learning Research*, vol. 13, pp. 2171–2175, 2012.

[19] T. Akiba, S. Sano, T. Yanase, T. Ohta, dan M. Koyama, “Optuna: A Next-generation Hyperparameter Optimization Framework,” *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019.

[20] L. James V. Miranda, “PySwarms: a Research Toolkit for Particle Swarm Optimization in Python,” *The Journal of Open Source Software*, vol. 3, p. 433, 2018.

[21] X. Xiao, J. Zhang, Y. Shao, J. Liu, K. Shi, C. He, dan D. Kong, "Deep Learning-Based Medical Ultrasound Image and Video Segmentation Methods: Overview, Frontiers, and Challenges," Sensors (Basel), vol. 25, p. 2361, 2025.

[22] J. Czajkowska, J. Juszczyk, L. Piejko, dan M. Glenc-Ambroży, "High-Frequency Ultrasound Dataset for Deep Learning-Based Image Quality Assessment," Sensors (Basel), vol. 22, p. 1478, 2022.

[23] D. R. Kandati dan T. R. Gadekallu, "Federated learning approach for early detection of chest lesion caused by COVID-19 infection using particle swarm optimization," Electronics, vol. 12, p. 710, 2023.

[24] D. Vamsikrishna, G. Reshma, N. B. Mustare, B. S. Rao, D. Sriramsetti, dan E. Muniyandy, "AHDNAM: An Optimised Hybrid Deep Learning and Swarm Intelligence Framework for Breast Ultrasound Diagnosis," Journal of Theoretical and Applied Information Technology, vol. 104, 2026.

[25] A. A. Alhussan, M. M. Eid, S. K. Towfek, dan D. S. Khafaga, "Breast Cancer Classification Depends on the Dynamic Dipper Throated Optimization Algorithm," Biomimetics (Basel), vol. 8, p. 163, 2023.

[26] A. Yonar, "A swarm intelligence-driven hybrid framework for brain tumor classification with enhanced deep features," Sci. Rep., vol. 15, p. 37543, 2025.

[27] A. AlZoubi, F. Lu, Y. Zhu, T. Ying, M. Ahmed, dan H. Du, "Classification of breast lesions in ultrasound images using deep convolutional neural networks: transfer learning versus automatic architecture design," Med. Biol. Eng. Comput., vol. 62, pp. 135-149, Jan. 2024.

[28] M. S. Sundari, N. V. Sailaja, D. Swapna, S. Vikkurty, V. C. Jadala, K. Durga, dan P. Thottempudi, "Transfer learning-enhanced CNN model for integrative ultrasound and biomarker-based diagnosis of polycystic ovarian disease," Sci. Rep., vol. 15, p. 34519, 2025.

[29] A. Zifan, K. Zhao, M. Lee, Z. Peng, L. J. Roney, S. Pai, J. T. Weeks, M. S. Middleton, A. E. Kaffas, J. B. Schwimmer, dan S. B. Sirlin, "Adaptive Evolutionary Optimization of Deep Learning Architectures for Focused Liver Ultrasound Image Segmentation," Diagnostics (Basel), vol. 15, p. 117, 2025.

[30] C. Cruz-Ramos, O. García-Avila, J. A. Almaraz-Damian, V. Ponomaryov, R. Reyes-Reyes, dan S. Sadovnychiy, "Benign and Malignant Breast Tumor Classification in

Ultrasound and Mammography Images via Fusion of Deep Learning and Handcraft Features," Entropy (Basel), vol. 25, p. 991, 2023.

[31] M. A. K. Raiaan, S. Sakib, N. M. Fahad, A. A. Mamun, M. A. Rahman, S. Shatabda, dan M. S. H. Mukta, "A systematic review of hyperparameter optimization techniques in Convolutional Neural Networks," Decision Analytics Journal, vol. 11, p. 100470, 2024.

[32] D. Kilichev dan W. Kim, "Hyperparameter Optimization for 1D-CNN-Based Network Intrusion Detection Using GA and PSO," Mathematics, vol. 11, no. 17, p. 3724, 2023.