

BAB II TINJAUAN PUSTAKA

A. Klasifikasi Penyakit Paru-paru Menggunakan *Deep Learning* pada Citra *X-ray*

Deteksi otomatis penyakit paru-paru dari citra *X-ray* toraks telah menjadi fokus riset yang intensif sejak pandemi COVID-19. Berbagai penelitian menunjukkan potensi *deep learning* dalam meningkatkan akurasi dan efisiensi diagnosis. Pendekatan berbasis *deep learning* mampu menangkap pola patologi halus yang sulit dikenali oleh mata manusia, sehingga membantu tenaga medis dalam pengambilan keputusan klinis yang lebih cepat dan andal [7].

Sistem berbasis *deep learning* telah dievaluasi untuk deteksi *tuberculosis* paru menggunakan citra *X-ray* toraks sebagai tes *triage* dengan melibatkan 392 pasien dengan kultur positif [6]. Algoritma *deep learning* mencapai sensitivitas 95% dan spesifisitas 81% untuk mendeteksi *tuberculosis* aktif. Penelitian ini membuktikan potensi AI sebagai alat skrining awal yang efektif, namun keterbatasan utamanya terletak pada evaluasi yang terbatas pada satu jenis penyakit paru-paru tanpa membandingkan performa berbagai arsitektur CNN.

Evaluasi komprehensif terhadap lima algoritma AI berbeda untuk deteksi *tuberculosis* dari *X-ray* toraks dalam *setting* beban *tuberculosis* tinggi telah dilakukan dengan melibatkan 17.629 partisipan [7]. Studi ini membandingkan lima produk komersial AI dan menunjukkan bahwa CAD4TB versi 7 mencapai AUC tertinggi 0.89 untuk deteksi *tuberculosis* kultur-positif. Temuan penting dari penelitian ini adalah variabilitas performa signifikan antara algoritma dengan AUC berkisar 0.84 – 0.89. Hal ini mengindikasikan bahwa pemilihan arsitektur dan konfigurasi model sangat memengaruhi hasil akhir.

Pendekatan *multi-label deep learning* untuk klasifikasi manifestasi *tuberculosis* berdasarkan zona paru menggunakan *X-ray* toraks telah dikembangkan dengan menggunakan arsitektur DenseNet121 yang dimodifikasi dengan *attention mechanism* [14]. Penelitian ini mengklasifikasikan enam zona paru secara simultan dan mencapai akurasi *zone-wise* 92,3% serta *patient-level accuracy* 89,7%. Integrasi *attention mechanism* terbukti meningkatkan interpretabilitas model.

dengan memvisualisasikan region yang berkontribusi terhadap prediksi. Penelitian ini terbatas pada satu jenis penyakit dan tidak melakukan perbandingan sistematis dengan arsitektur CNN lainnya seperti EfficientNetB2 atau ResNet50, serta tidak mengeksplorasi dampak optimasi hiperparameter terhadap performa klasifikasi.

B. Perbandingan Arsitektur *Deep Learning* untuk *Medical Imaging*

Pemilihan arsitektur CNN yang tepat memiliki dampak signifikan terhadap kinerja klasifikasi citra medis. Berbagai studi komparatif telah dilakukan untuk mengidentifikasi arsitektur optimal dalam konteks citra medis. Studi dilakukan dengan mempertimbangkan perbedaan mendasar antara data medis dan citra alami seperti dominasi warna *grayscale*, variasi anatomi yang tinggi, serta keterbatasan ukuran *dataset*.

Studi komparatif komprehensif terhadap 16 arsitektur CNN berbeda pada *dataset* CheXpert dan *COVID-19 Image Data Collection* untuk klasifikasi radiografi toraks telah dilakukan untuk mengevaluasi performa berbagai arsitektur mulai dari model ringan hingga model sangat dalam [9]. Hasil menunjukkan bahwa jaringan yang lebih dalam seperti ResNet152, DenseNet161, dan ResNet50 mencapai AUROC gabungan tertinggi antara 0.881 – 0.882, sementara model yang lebih ringan seperti DenseNet121-121 menunjukkan *trade-off* terbaik antara performa dengan AUROC 0.874 dan efisiensi komputasi dengan 7,98 juta parameter. Temuan penting yang perlu disoroti adalah bahwa arsitektur yang dirancang untuk ImageNet dengan struktur kompleks tidak selalu lebih unggul untuk klasifikasi citra medis yang memiliki karakteristik berbeda. Hal ini mengindikasikan pentingnya evaluasi empiris untuk *task-specific medical imaging*.

Kajian komprehensif terhadap penerapan *Dense Convolutional Network* dalam analisis citra medis telah mengidentifikasi bahwa pola konektivitas padat DenseNet121 memfasilitasi penggunaan kembali fitur yang superior dan propagasi gradien yang lebih baik dibandingkan arsitektur konvensional seperti VGG dan ResNet50 *original* [10]. *Review* sistematis ini menganalisis berbagai varian DenseNet121 yang diaplikasikan pada berbagai modalitas citra medis termasuk *CT scan*, MRI, dan *X-ray* dengan fokus pada efisiensi parameter dan kemampuan ekstraksi fitur hierarkis. DenseNet121 terbukti efektif untuk *dataset* medis berukuran kecil hingga menengah karena mekanisme *feature reuse*-nya

mengurangi risiko *overfitting*.

Arsitektur CNN yang mengoptimalkan *accuracy* dan *efficiency* melalui *compound scaling method* telah diperkenalkan untuk menskalakan *depth*, *width*, dan *resolution* secara bersamaan dengan rasio tetap [15]. Evaluasi pada ImageNet menunjukkan EfficientNetB7 mencapai *top-1 accuracy* 84,3% dengan parameter 8,4 kali lebih sedikit dan *inference speed* 6,1 kali lebih cepat dibandingkan model terbaik CNN. *Compound scaling approach* terbukti lebih efektif dibandingkan *scaling* dimensi tunggal dalam meningkatkan *representational capacity* model. Penelitian ini membuka paradigma baru dalam desain arsitektur CNN yang mengutamakan *efficiency*, sangat relevan untuk aplikasi *medical imaging* yang sering beroperasi dengan *computational resource* terbatas dan memerlukan *inference time* yang cepat untuk *clinical deployment*.

Studi komparatif menggunakan EfficientNetB2 untuk deteksi COVID-19 dari citra *X-ray* toraks telah dilakukan dengan membandingkan performa berbagai varian EfficientNetB2 dari B0 hingga B7 [16]. Penelitian ini menunjukkan bahwa EfficientNetB3 mencapai keseimbangan optimal antara *accuracy* 97,89% dan *computational efficiency* untuk klasifikasi empat kelas, yaitu COVID-19, *Normal*, *Viral Pneumonia*, dan *Lung Opacity*. Model ini menggunakan *transfer learning* dengan *fine-tuning* seluruh *layer* dan *data augmentation strategy* yang mencakup *horizontal flip*, *rotation*, dan *brightness adjustment*. Hasil *ablation study* menunjukkan bahwa *pre-training ImageNet weights* meningkatkan *accuracy* sebesar 4,2% dibandingkan *training from scratch*, mengonfirmasi efektivitas *transfer learning* untuk *medical imaging* meskipun terdapat *domain shift* antara *natural images* dan *radiographic images*.

C. Optimasi Hyperparameter Menggunakan Bayesian Optimization

Optimasi *hyperparameter* merupakan faktor kritis dalam *deep learning* karena secara langsung memengaruhi kinerja model. Metode tradisional seperti *Grid Search* dan *Random Search* memiliki keterbatasan efisiensi komputasi, mendorong adopsi *Bayesian optimization* sebagai pendekatan yang lebih *sample-efficient*. *Bayesian optimization* memanfaatkan informasi dari evaluasi sebelumnya untuk memandu pencarian menuju konfigurasi yang lebih menjanjikan, sehingga mengurangi kebutuhan evaluasi yang mahal [8].

Studi komparatif terhadap berbagai *tools* optimasi *hyperparameter* telah dilakukan untuk mengevaluasi performa *Grid Search*, *Random Search*, *Bayesian Optimization*, dan *Genetic Algorithms* pada berbagai algoritma *machine learning* menggunakan *dataset* UCI *Machine Learning Repository* [8]. Hasil eksperimen menunjukkan bahwa *Bayesian optimization* mencapai konvergensi tercepat dengan rata-rata 40% lebih sedikit evaluasi dibandingkan *Random Search* untuk mencapai performa yang setara. *Framework* seperti Optuna dan Hyperopt yang mengimplementasikan *Tree-structured Parzen Estimator* menunjukkan efisiensi superior pada *search space* berdimensi tinggi dengan *computational budget* terbatas. Temuan ini mengonfirmasi keunggulan pendekatan *sequential model-based optimization* dibandingkan metode *sampling* konvensional, terutama untuk tugas *deep learning* di mana setiap evaluasi memerlukan waktu *training* yang lama.

Framework hyperparameter optimization generasi baru telah diperkenalkan dengan desain yang mengutamakan fleksibilitas dan efisiensi [13]. Optuna mengimplementasikan berbagai algoritma *sampling* termasuk TPE, CMA-ES, dan *Grid/Random Search* dalam arsitektur modular yang mendukung *define-by-run* API. Hal ini memungkinkan *dynamic construction of search space* yang tidak dimungkinkan oleh *framework* tradisional. Fitur *pruning* otomatis Optuna mampu menghentikan *trial* yang tidak menjanjikan secara dini, menghemat hingga 60% *computational resource* dalam eksperimen *benchmark* pada *dataset* MNIST dan CIFAR-10. Optuna juga menyediakan *visualization tools* yang komprehensif untuk *monitoring convergence history*, *hyperparameter importance analysis*, dan *parallel coordinate plots* yang memfasilitasi interpretasi hasil optimasi. Keunggulan Optuna dalam *terms of ease-of-use*, *scalability*, dan *integration* dengan *deep learning frameworks* menjadikannya pilihan populer untuk penelitian *medical imaging* yang memerlukan *extensive hyperparameter tuning*.

Implementasi *Bayesian optimization* menggunakan *Gaussian Process* untuk *hyperparameter tuning* CNN dalam *task* klasifikasi citra medis telah dilakukan dengan menggunakan *Gaussian Process* sebagai *surrogate model* untuk memodelkan hubungan antara konfigurasi hiperparameter dan performa model [17]. *Expected Improvement* digunakan sebagai *acquisition function* untuk menentukan titik evaluasi berikutnya. Eksperimen pada *dataset medical imaging*

menunjukkan bahwa *Bayesian optimization* dengan *Gaussian Process* mencapai *optimal configuration* dalam rata-rata 45 *trials*, signifikan lebih sedikit dibandingkan 200 lebih *trials* yang diperlukan *Random Search*. Penelitian ini juga menganalisis *uncertainty quantification* yang disediakan *Gaussian Process*, menunjukkan bahwa *confidence intervals* dapat membantu mengidentifikasi *region search space* yang memerlukan eksplorasi lebih lanjut versus *region* yang sudah cukup di-exploit.

Analisis mendalam terhadap algoritma *Tree-structured Parzen Estimator* telah dilakukan dengan mengusulkan perbaikan pada komponen-komponen kunci TPE [18]. Studi ini mengidentifikasi bahwa parameter kontrol seperti *gamma* untuk memisahkan *good* versus *bad trials*, *number of EI candidates*, dan *prior weight* memiliki dampak signifikan terhadap *convergence speed* dan *quality of final solution*. Eksperimen pada 57 *benchmark functions* menunjukkan bahwa konfigurasi TPE yang dioptimalkan dapat meningkatkan *performance* hingga 23% dibandingkan *default settings* pada *high-dimensional search spaces* dengan lebih dari 20 dimensi. Penelitian ini memberikan *guideline* praktis untuk *setting* parameter TPE berdasarkan karakteristik masalah seperti *dimensionality*, *budget*, dan *noise level* dalam *objective function evaluations*.

D. *Tree-structured Parzen Estimator (TPE)* dan *Gaussian Process (GP)* dalam *Hyperparameter Optimization*

Tree-structured Parzen Estimator dan *Gaussian Process* merepresentasikan dua paradigma berbeda dalam *Bayesian optimization* dengan karakteristik komplementer. TPE merupakan metode yang banyak diadopsi dalam *framework* seperti Hyperopt dan Optuna karena efisiensinya pada *search space* berdimensi tinggi. GP menawarkan *global modeling* dengan *uncertainty quantification* yang teliti, namun terbatas skalabilitasnya pada jumlah *trial* besar.

TPE memodelkan distribusi probabilitas konfigurasi hiperparameter menggunakan dua *Gaussian Mixture Models* terpisah untuk konfigurasi yang menghasilkan performa baik dan performa buruk [18]. Pendekatan ini berbeda fundamental dengan *Gaussian Process* yang memodelkan langsung hubungan antara hiperparameter dan performa. TPE menggunakan *Expected Improvement* sebagai *acquisition function* yang dapat diaproksimasi sebagai rasio antara *density*

model, menjadikan komputasi signifikan lebih efisien pada *high-dimensional spaces* karena tidak memerlukan *matrix inversion* yang mahal secara komputasi. Eksperimen *benchmark* menunjukkan TPE mempertahankan kompleksitas komputasi linear per iterasi dibandingkan kubik untuk GP, memberikan keuntungan signifikan ketika anggaran evaluasi mencapai ratusan *trials*.

Perbandingan performa TPE dan *GP-based Bayesian optimization* pada berbagai *hyperparameter optimization benchmarks* telah dilakukan dengan mengevaluasi kedua metode pada 15 *task machine learning* dengan dimensi yang bervariasi dari 4–48 dimensi dan batasan anggaran 50–500 evaluasi [19]. Hasil menunjukkan bahwa GP mencapai performa akhir yang lebih baik pada masalah berdimensi rendah dengan kurang dari 10 dimensi dengan anggaran yang cukup lebih dari 200 evaluasi, sementara TPE lebih robust pada masalah berdimensi tinggi dan skenario anggaran terbatas. Analisis pola konvergensi menunjukkan GP memiliki eksplorasi yang lebih baik di tahap awal berkat pemodelan ketidakpastian yang eksplisit, sedangkan TPE lebih efektif dalam fase eksploitasi dengan estimasi densitas adaptif. Temuan ini mengindikasikan bahwa pemilihan metode optimasi seharusnya mempertimbangkan karakteristik masalah seperti dimensionalitas dan anggaran komputasi yang tersedia.

Konsep dasar *Tree-structured Parzen Estimator* telah diperkenalkan sebagai alternatif efisien untuk *Gaussian Process* dalam *hyperparameter optimization* [20]. Makalah ini mendemonstrasikan bahwa dengan memodelkan probabilitas konfigurasi *given* performa alih-alih performa *given* konfigurasi, TPE dapat menghindari hambatan komputasi yang melekat dalam inferensi GP. Implementasi TPE menggunakan *adaptive parzen window estimators* yang secara iteratif disempurnakan berdasarkan konfigurasi baik dan buruk yang diamati. Eksperimen pada *hyperparameter tuning* jaringan saraf menunjukkan TPE mencapai performa yang sebanding dengan metode berbasis GP sambil mengurangi waktu komputasi hingga 5 – 10 kali pada masalah dengan lebih dari 15 hiperparameter. Penelitian ini menjadi fondasi untuk adopsi luas TPE dalam *framework AutoML* modern.

Studi empiris komprehensif pada *Black-Box Optimization Challenge 2020* telah dilakukan untuk membandingkan berbagai metode *Bayesian optimization*

termasuk *GP-based BO*, TPE, dan *Random Search* [21]. Studi ini menggunakan 24 fungsi *black-box* dengan karakteristik yang bervariasi seperti konveks, multimodal, *noisy*, dan berdimensi tinggi serta batasan anggaran 100 – 1000 evaluasi. Hasil menunjukkan bahwa untuk *hyperparameter tuning medical imaging* secara khusus, metode berbasis GP unggul pada fase eksplorasi awal dengan pemahaman global yang lebih baik terhadap lanskap objektif, sementara TPE menunjukkan kemampuan eksploitasi yang superior pada tahap selanjutnya. Untuk *task* dengan hiperparameter *mixed integer* dan kontinu yang khas dalam *tuning CNN* seperti *learning rate*, *batch size*, *dropout rate*, dan jumlah unit, TPE menunjukkan konvergensi yang lebih stabil. Penelitian ini merekomendasikan strategi optimasi dua fase dengan menggunakan GP untuk eksplorasi luas awal kemudian beralih ke TPE untuk eksploitasi yang lebih halus.

E. Attention Mechanism: Convolutional Block Attention Module (CBAM)

Mekanisme *attention* telah menjadi teknik yang *powerful* untuk meningkatkan kapabilitas representasi dalam CNN dengan cara mengarahkan fokus model pada fitur-fitur yang paling informatif. CBAM merupakan modul *attention* ringan yang secara berurutan menyimpulkan peta *attention* sepanjang dimensi kanal dan spasial, kemudian mengalikannya dengan peta fitur input untuk penyempurnaan fitur adaptif [11]. CBAM dapat diintegrasikan secara mulus ke berbagai arsitektur CNN dengan *overhead* komputasi yang dapat diabaikan dan dapat dilatih secara *end-to-end*. Modul ini menunjukkan peningkatan konsisten pada berbagai *benchmark dataset* termasuk ImageNet-1K, MS COCO *object detection*, dan VOC 2007 *detection* dengan peningkatan rata-rata 1 – 2% pada akurasi tanpa menambah parameter signifikan.

Implementasi *CBAM-based attention gate* dalam *framework* U-Net untuk segmentasi mikroaneurisma pada citra fundus retina telah dilakukan dengan mengintegrasikan CBAM di setiap *encoder block* untuk penekanan fitur berbasis kanal dan spasial [12]. Arsitektur ini dikombinasikan dengan *attention gates* pada *skip connections* untuk fokus selektif pada region lesi. Evaluasi pada *dataset public* IDRiD dan Messidor menunjukkan performa superior dengan IoU 0.758, *Dice Coefficient* 0.865, dan AUC-ROC 0.996, mengungguli U-Net standar dengan *Dice*

0.782 dan *attention* U-Net dengan *Dice* 0.821. CBAM di *encoder* terbukti efektif mengidentifikasi fitur spasial dan *channel-wise* yang paling kritis untuk lokalisasi lesi mikroskopis, sementara *attention gates* menyempurnakan fitur spesifik lesi dan meminimalkan *noise* dari *background* pembuluh darah yang kompleks. Penelitian ini mendemonstrasikan bahwa CBAM tidak hanya meningkatkan akurasi segmentasi, tetapi juga memberikan interpretabilitas melalui visualisasi peta *attention*.

Integrasi CBAM ke dalam arsitektur U-Net untuk segmentasi tumor otak dari *MRI scans* telah dikembangkan dengan mengimplementasikan modul CBAM pada *skip connections* antara *encoder* dan *decoder* untuk propagasi fitur yang lebih halus [22]. Evaluasi pada *dataset* BraTS menunjukkan peningkatan *Dice score* sebesar 3,7% dari 0.861 menjadi 0.898 dibandingkan U-Net dasar. Analisis visualisasi peta *attention* mengungkapkan bahwa *channel attention* berhasil menekan fitur *background* yang tidak relevan dan meningkatkan pola spesifik tumor, sementara *spatial attention* secara efektif melokalisasi batas tumor dengan presisi lebih tinggi. Penelitian ini juga melakukan *ablation study* yang menunjukkan bahwa *attention* kanal-kemudian-spasial secara berurutan mengungguli konfigurasi paralel atau spasial-kemudian-kanal, mengonfirmasi pilihan desain makalah CBAM asli.

Arsitektur ResNet50-CBAM untuk klasifikasi COVID-19 dari citra *X-ray* toraks telah dikembangkan dengan mengintegrasikan modul CBAM setelah setiap *residual block* dalam ResNet50 untuk diskriminasi fitur yang lebih baik [23]. Model ini mencapai akurasi 98,32% pada *dataset* COVIDx dengan 13.975 gambar *chest X-ray*, mengungguli ResNet50 dasar dengan 96,84% dan DenseNet121 dengan 97,15%. Visualisasi *Gradient-weighted Class Activation Mapping* menunjukkan bahwa model yang ditingkatkan CBAM fokus lebih akurat pada region perifer paru dan *ground-glass opacities* yang karakteristik untuk COVID-19, dibandingkan model dasar yang *attention*-nya lebih tersebar. Penelitian ini membuktikan bahwa CBAM secara efektif meningkatkan performa kuantitatif dan interpretabilitas kualitatif untuk *task* klasifikasi citra medis.

F. Keterbatasan Penelitian Terdahulu dan Kontribusi Penelitian

Berdasarkan tinjauan pustaka yang telah dipaparkan, teridentifikasi beberapa keterbatasan

pada penelitian-penelitian terdahulu yang menjadi motivasi penelitian ini. Tabel menyajikan perbandingan sistematis antara penelitian ini dengan penelitian-penelitian terdahulu berdasarkan tujuh aspek kunci, yaitu arsitektur model, metode optimasi hiperparameter, *scope* klasifikasi, integrasi *attention mechanism*, strategi *transfer learning*, penanganan *class imbalance*, dan metrik evaluasi.

Tabel 1 Perbandingan Penelitian

No	Author	Metode	Hyperparameter Tuning	Poin of Interest	Hasil	Keterbatasan
1	Vaiyapuri et al.(2025) [24]	U-Net untuk segmentasi brain MRI dengan Optuna TPE	- Tree-structured Parzen Estimator (TPE) via Optuna - Bayesian Optimization dengan adaptive pruning - Optimasi: filters, optimizer, learning rate, dropout - Segmentasi tumor otak	- Pendekatan TPE-only untuk segmentasi citra medis - Optimasi arsitektur U-Net - Dataset brain MRI (TCGA-LGG) 3.929 axial slices dari 110 pasien - Pencarian hyperparameter komprehensif	- Mean Dice Coefficient: 0,941 - Mean IoU: 0,8763 - Mean SSIM: 0,983 - Mengungguli baseline Grid Search dan PSO - TPE efektif untuk segmentasi tumor otak	- TPE saja tanpa fase eksplorasi GP - Risiko konvergensi prematur ke local optima - Tugas segmentasi (bukan klasifikasi) - Single modality (hanya MRI) - Tidak membandingkan dengan pendekatan hybrid TPE-GP
2	Hamza et al.(2023)[25]	Deep Bayesian Optimization + Fusion (D2BOF-COVIDNet) untuk klasifikasi COVID-19	- Bayesian Optimization dengan Gaussian Process - Fusi Chest X-ray dan MRI - Ekstraksi deep feature - Pencarian hyperparameter berbasis GP	- Bayesian optimization berbasis GP untuk fusi multi-modal - Kombinasi CXR dan MRI - Framework deep feature fusion - Transfer learning dari pre-trained CNNs	- Performa klasifikasi tinggi pada deteksi COVID-19 - Fusi multi-modal efektif - Akurasi diagnostik meningkat	- GP saja, tidak hybrid dengan TPE - Fokus hanya pada COVID-19 - Kebutuhan multi-modal (tidak praktis untuk pengaturan dengan sumber daya terbatas) - Tidak ada perbandingan sistematis dengan TPE
3	Ali et al. (2023)[26]	Random Search + Grid Search untuk hyperparameter tuning algoritma ML	- Random Search + Grid Search (metode tradisional) - Ant Bee Colony (ABC) algorithm - Particle Swarm Optimization (PSO) - Perbandingan dengan metode tradisional	- Perbandingan metode optimasi tradisional - ABC dan PSO untuk pencarian hyperparameter - Analisis kompleksitas komputasi - Evaluasi multiple algoritma ML	- PSO menunjukkan hasil menjanjikan namun konvergensi lebih lambat - ABC efektif untuk skenario spesifik - Baseline Grid/Random Search telah ditetapkan	- Bukan metode Bayesian (tidak ada pemodelan probabilistik) - Komputasi mahal - Konvergensi lambat dibanding pendekatan Bayesian - Tidak ada kuantifikasi ketidakpastian
4	Chihaoui et al. (2024)[27]	Genetic Algorithm (GA) untuk optimasi CNN dan VGG16 pada deteksi pneumonia	- Genetic Algorithm - Evolusi berbasis populasi - Operator selection, crossover, mutation - Optimasi arsitektur CNN + hyperparameter - Dataset: 5.856 citra chest X-ray	- Pendekatan evolusioner (GA) untuk optimasi CNN - Architecture search + hyperparameter tuning - Deteksi pneumonia dari chest X-rays - Optimasi multi-objective	- CNN custom: 97% akurasi training, 94% akurasi testing, loss 0,072 - VGG16-GA: 90,90% akurasi training, 90,90% akurasi testing, loss 0,11 - GA efektif untuk penemuan hyperparameter	- GA memerlukan evaluasi masif (berbasis populasi) - Tidak ada pemodelan probabilistik - Komputasi intensif - Tidak membandingkan dengan metode Bayesian (TPE/GP) - Klasifikasi biner (pneumonia vs Normal)
5	Subaşi.(2024)[28]	Perbandingan komprehensif Grid Search vs Random Search	- Grid Search vs Random Search - Kombinasi grid exhaustive - Random sampling dari ruang hyperparameter - Multiple model ML (SVM, NN) - Multiple dataset (MNIST, dll.)	- Perbandingan komprehensif Grid vs Random - Analisis efisiensi - Evaluasi skalabilitas - Trade-off biaya komputasi - Benchmark pada dataset standar	- Grid Search: MNIST SVM 92,44% akurasi (3349s) - Random Search: MNIST SVM 90,5% akurasi (2581s) - Random Search 23% lebih cepat dengan trade-off akurasi 1,94% - Grid komprehensif namun mahal	- Metode tradisional non-Bayesian - Grid Search: kompleksitas eksponensial dengan dimensi - Random Search: tidak ada panduan cerdas - Tidak ada pemodelan ketidakpastian - Tidak membandingkan dengan pendekatan Bayesian modern (TPE/GP) approaches (TPE/GP)

Berdasarkan Tabel 1, teridentifikasi beberapa karakteristik distingtif penelitian ini dibandingkan penelitian terdahulu dalam konteks optimasi hiperparameter untuk klasifikasi citra medis. Pendekatan *TPE-only* melalui Optuna untuk segmentasi tumor otak mencapai *Mean Dice Coefficient* 0.941, namun terbatas pada metode optimasi tunggal tanpa fase eksplorasi dari GP [24]. Pendekatan optimasi Bayesian berbasis GP untuk klasifikasi COVID-19 dengan fusi multimodal tidak mengintegrasikan TPE untuk eksploitasi yang lebih halus [25]. Metode tradisional seperti PSO, ABC, dan *Genetic Algorithm* bersifat mahal secara komputasi dan tidak menyediakan pemodelan probabilistik [26]. Perbandingan *Grid Search* versus *Random Search* pada *dataset* MNIST menunjukkan akurasi 92,44% versus 90,5%, namun kedua metode ini tidak memiliki panduan cerdas dan kuantifikasi ketidakpastian [28].

Dari aspek metodologi optimasi hiperparameter, penelitian ini mengimplementasikan optimasi otomatis dua fase yang ketat menggunakan integrasi GP-TPE dengan 20 *trials* eksplorasi awal dan 30 *trials* eksploitasi yang lebih halus. Strategi dua fase ini memanfaatkan kekuatan komplementer dari kedua metode, yaitu GP untuk pemahaman global yang lebih baik pada tahap awal dan TPE untuk optimasi lokal yang efisien pada tahap akhir. Pendekatan metode tunggal pada penelitian terdahulu berisiko mengalami konvergensi prematur ke optima lokal atau kurang efisien pada tahap optimasi akhir [24]. Metode evolusioner seperti PSO dan GA memerlukan evaluasi yang sangat banyak tanpa pemodelan probabilistik [27]. Studi komparatif menunjukkan *Grid Search* membutuhkan 3.349 detik dibandingkan *Random Search* 2.581 detik pada MNIST, mengonfirmasi inefisiensi komputasi dari metode pencarian ekshaustif [28].

Dari aspek cakupan arsitektur, penelitian ini melakukan perbandingan sistematis terhadap tiga *backbone state-of-the-art* dengan dua konfigurasi masing-masing, menghasilkan total enam model kandidat yang lebih komprehensif dibandingkan penelitian terdahulu. Pemilihan ketiga arsitektur didasarkan pada representasi paradigma berbeda dalam desain CNN, yaitu *compound scaling* untuk EfficientNetB2, *dense connectivity* untuk DenseNet121, dan *residual learning* untuk ResNet50. Penelitian terdahulu hanya fokus pada arsitektur U-Net untuk tugas segmentasi atau membandingkan *custom* CNN dengan VGG16 untuk

klasifikasi biner pneumonia [27]. Pendekatan multimodal dengan fusi dari CNN yang telah dilatih sebelumnya tidak melakukan perbandingan arsitektur secara sistematis. Penelitian ini secara sistematis mengintegrasikan mekanisme *attention* CBAM pada semua arsitektur untuk menilai manfaat generalisasi, berbeda dengan penelitian terdahulu yang tidak mengeksplorasi performa lintas arsitektur dari mekanisme *attention*.

Dari aspek cakupan penyakit dan aplikabilitas klinis, penelitian ini mengklasifikasikan lima kategori penyakit yang mencakup diagnosis diferensial yang menantang antara Pneumonia Bakterial dan Pneumonia Viral yang kritis secara klinis. Penelitian terdahulu fokus pada segmentasi tumor otak dengan 3.929 potongan aksial dari 110 pasien atau spesifik untuk deteksi COVID-19 dengan persyaratan multimodal yang tidak praktis untuk pengaturan dengan sumber daya terbatas [25]. Klasifikasi biner pneumonia versus *Normal* mencapai akurasi 94% pada pengujian, namun tidak menangani kompleksitas diagnosis diferensial [27]. Klasifikasi lima kelas dalam penelitian ini mencakup skenario diagnosis diferensial yang menantang yang sering dijumpai dalam praktik klinis, dengan beberapa kelas memiliki kesamaan visual tinggi yang memerlukan daya diskriminatif maksimal dari model. Penelitian dengan *tuning* manual atau optimasi strategi tunggal rentan terhadap konfigurasi suboptimal untuk skenario multikelas dengan kesamaan antarkelas yang tinggi.

Dari aspek penanganan ketidakseimbangan dan *preprocessing* data, penelitian ini menerapkan strategi ganda berupa pembobotan kelas dan augmentasi adaptif yang lebih canggih dibandingkan pendekatan strategi tunggal pada penelitian terdahulu. Kelas minoritas dengan sampel kurang dari 5.000 mendapat augmentasi lebih agresif dengan rotasi, pergeseran, dan penyesuaian kecerahan, sementara bobot kelas dihitung menggunakan strategi seimbang untuk memastikan model tidak bias terhadap kelas mayoritas. Penelitian terdahulu menggunakan augmentasi standar tanpa strategi adaptif untuk *dataset brain* MRI yang tidak seimbang [24]. *Dataset* deteksi pneumonia dengan 5.856 citra *X-ray* dada tidak secara eksplisit menangani ketidakseimbangan kelas [27]. Dengan karakteristik-karakteristik ini, penelitian ini berkontribusi pada pemahaman empiris mengenai kombinasi optimal arsitektur-hiperparameter untuk klasifikasi penyakit paru

multikelas dan menyediakan kerangka metodologi yang dapat diadaptasi untuk tugas pencitraan medis lainnya dengan karakteristik serupa seperti ketidakseimbangan kelas, ukuran *dataset* terbatas, dan kesamaan visual antarkelas yang tinggi.

