

BAB II

TINJAUAN PUSTAKA

A. *Ensemble Learning* dengan *Hyperparameter*

Penelitian mengenai deteksi dan klasifikasi penyakit retina berbasis citra *Optical Coherence Tomography* (OCT) dengan pendekatan *ensemble learning* telah memberikan kontribusi yang signifikan dalam meningkatkan efisiensi serta akurasi diagnosis klinis. *Ensemble learning* dikenal karena kemampuannya menggabungkan kekuatan beberapa model misalnya berbagai varian *Convolutional Neural Network*, *Gradient Boosting*, atau *Random Forest* sehingga mampu memperoleh variasi tekstur mikroskopis pada lapisan retina dan tetap tangguh terhadap *noise speckle* khas *Optical Coherence Tomography* (OCT). Beragam studi telah mengeksplorasi potensi teknik seperti *bagging*, *boosting*, dan *stacking* dalam memperbaiki performa deteksi lesi retina, menunjukkan peningkatan sensitivitas dan spesifisitas yang konsisten dibanding model tunggal, sekaligus mempercepat alur kerja skrining oftalmologi secara otomatis, berikut hasil *review* penelitian klasifikasi citra menggunakan *Hyperparameter* dan *Ensemble Learning*.

Penelitian [23] berfokus pada perbaikan sistem klasifikasi citra angka tulisan tangan MNIST melalui otomatisasi penyetelan *hyperparameter* yang selama ini masih bersifat coba-coba dan rawan *over-fitting*. Permasalahan utama karena akurasi tinggi sangat dibutuhkan, misalnya pada pengenalan karakter untuk cek bank atau pengurutan kode pos, sementara beban komputasi dan waktu eksperimen manual terus meningkat seiring kompleksitas arsitektur CNN. Penelitian tersebut menggabungkan *Convolutional Neural Network* dengan Algoritma Genetika dalam skema *stacking* yang disebut sebagai *Ensemble Genetic Algorithm-CNN* (EGACNN). *Genetic Algorithm* (GA) secara evolusioner mencari kombinasi terbaik jumlah lapisan ($2 \times \text{Conv } 3 \times 3$ ber-filter 32–64 diikuti Dense 128), *learning rate* 0,001; *batch size* 32 dan *dropout* 0,5. Hasil eksperimen menunjukkan peningkatan signifikan yaitu *Ensemble Genetic Algorithm-CNN* (EGACNN) mencapai akurasi 99,91 %, presisi 98,79 %, recall 98,78 %, dan F1-score 98,78 %

hanya dalam lima *epoch*, melampaui sembilan arsitektur pembandingan termasuk ResNet, VGG, dan CSNN. Temuan ini menegaskan efektivitas *Genetic Algorithm* (GA) dalam mereduksi usaha manual dan mempercepat konvergensi model tanpa mengorbankan akurasi.

Penelitian [24]. bertujuan mengatasi tantangan identifikasi *hard exudates* (HE) dan klasifikasi *disorganization of retinal inner layers* (DRIL) pada citra OCT mata dengan *diabetic macular edema* (DME), yang selama ini mengandalkan inspeksi manual lambat dan subjektif. Penelitian menggunakan 442 citra OCT (6,847 HE teranotasi) dan merancang *pipeline* dua-cabang: (1) deteksi HE menggunakan YOLOv7 dengan penyesuaian *hyperparameter input resize* 1024×1024 px, *batch* 16, *momentum* 0,937 serta *learning-rate* adaptif 0,01 sampai dengan 0,10 disertai pasca proses *IOU-merging* dan *edge-cleaning*; (2) klasifikasi DRIL lewat *ensemble soft-voting* dari tiga model (ConvNeXt-tiny 1: epoch 50, LR 0,01; ConvNeXt-tiny 2: epoch 100, LR 0,001; RegNetX_016: epoch 50, LR 0,001; seluruhnya *batch* 16, resolusi 434×1024 px). Pendekatan ini unggul karena memadukan deteksi objek kecil dan klasifikasi citra penuh, sekaligus menjaga waktu inferensi di bawah satu detik per gambar sehingga berpotensi dipakai di klinik. Hasilnya menunjukkan kinerja yang unggul untuk DRIL akurasi 91,1 % (CI 95 % 0,852–0,969), sensitivitas atau spesifisitas 91,1 % dengan AUC = 0,91 serta AUPR = 0,91 dan koefisien Kappa 0,82 ($p < 0,05$). Sedangkan deteksi HE mencapai *Average Precision* = 34,4 % *precision* 48,7 % *recall* 43,1 % serta korelasi jumlah HE models–ahli $r = 0,804$ ($p < 0,05$), performa yang wajar mengingat ukuran lesi yang sangat kecil. Kesimpulannya kombinasi YOLOv7 ter-tuning (*batch* 16, LR 0,01–0,10) dengan *ensemble* ConvNeXt/RegNet (epoch 50–100; LR 0,001–0,01) menghasilkan sistem berbasis kecerdasan yang cepat dan andal, serta sangat akurat untuk DRIL dan cukup efektif menandai HE. Solusi ini berpotensi menjadi alat bantu objektif bagi *oftalmologi* dalam evaluasi *biomarker* OCT pada DME, sekaligus dasar pengembangan penelitian lanjutan dengan dataset lebih besar.

Penelitian [25] berupaya meningkatkan akurasi segmentasi tiga jenis cairan retina *intra-retinal fluid* (IRF), *sub-retinal fluid* (SRF), dan *pigment epithelial*

detachment (PED) pada citra OCT, masalah yang krusial karena inspeksi manual *RETOUCH* bersifat lambat dan rentan bias. Tim mengusulkan *deep ensemble* CNN, UNet + peta jarak antar-lapisan retina untuk IRF, *Structured-Dropout* UNet (SD-UNet) ramping untuk SRF, serta UNet berpadu augmentasi data untuk PED. Setiap model dilatih 200 epoch dengan *learning rate* 3×10^{-4} dan *loss Binary Cross Entropy* pada patch 256×512 px, hasil ketiganya digabung dalam blok prediktor. Uji pada dataset *RETOUCH* (70 volume publik, tiga vendor OCT) menunjukkan peningkatan rata-rata Dice multi-kelas menjadi 71,86 %, naik 1,8 % dibanding model *benchmark* SFU; khusus SRF, SD-UNet menyumbang kenaikan 5 %. Visualisasi pada citra *Cirrus*, *Spectralis*, dan *Topcon* menampilkan batas lesi lebih bersih dengan *false-positive* lebih sedikit, menegaskan keunggulan pendekatan *ensemble* berbasis *tuning hyperparameter* yang spesifik untuk cairan retina.

Penelitian [21] berfokus pada otomasi klasifikasi empat penyakit retina *choroidal neovascularization* (CNV), *diabetic macular edema* (DME), *Drusen*, dan mata normal menggunakan citra *Optical Coherence Tomography* (OCT). Tantangan klinisnya adalah gejala keempat kategori sering tumpang-tindih sehingga pemeriksaan manual memakan waktu dan rawan salah diagnosis. Tim merancang skema *deep ensemble* berbasis *transfer-learning*: tiga jaringan pralatih (VGG16, VGG19, DenseNet201) dimodifikasi dengan blok AveragePooling2D → Flatten → Dense 32 (ReLU) → Dropout 0,20 → Dense 4 (*soft-max*). Masing-masing model dilatih pada 10.336 citra OCT seimbang (2.584 per kelas) dengan *learning-rate* 3×10^{-4} , *batch* 32, dan augmentasi (shift 20 %, rotasi 15°, *flipping*) setelah normalisasi min-max. Prediksi tiga model digabung melalui *weighted-average ensembling*; bobot optimal (0,4: 0,5: 0,3) diperoleh via pencarian grid. Hasilnya, *ensemble* mencapai akurasi 83 % naik 5-22 poin persentase dibanding model tunggal terbaik dengan *f1-score* tertinggi pada CNV (0,88) dan DME (0,85). *Confusion-matrix* menunjukkan penurunan salah klasifikasi antar-penyakit, menegaskan bahwa kombinasi beberapa CNN ringan yang disetel *hyperparameter* secara cermat dapat meningkatkan reliabilitas deteksi OCT tanpa arsitektur kompleks.

Penelitian [26] berfokus pada otomatisasi deteksi infeksi COVID-19 dari citra CT dada dengan mengatasi keterbatasan pendekatan tunggal yang rentan salah

klasifikasi. Penelitian tersebut merancang *ensemble novel* yang memadukan delapan model pralatih VGG16, VGG19, ResNet50, ResNet50V2, InceptionV3, Xception, MobileNet, dan InceptionResNetV2 melalui *fusi* probabilitas *Softmax* pada sebuah *composite neural network* dengan dua lapis *fully connected* ter-regulasi. Setiap kandidat diperlakukan dengan pembekuan konvolusi dasar serta penambahan kepala klasifikasi *dropout* dan regularisasi L1/L1L2 guna mencegah *overfitting*, sedangkan *ensembel* dioptimasi terpisah dengan *learning-rate* lebih kecil untuk menjaga kestabilan konvergensi. Metode ini diuji pada dua dataset *benchmark*, SARS-CoV-2 (2 482 citra) dan COVID-CT (746 citra), menggunakan skema *hold-out* 80/20, Hasilnya menunjukkan peningkatan kinerja yang signifikan dibandingkan model tunggal maupun teknik *ensembel* klasik, pada dataset SARS-CoV-2, akurasi mencapai 98,99 %, presisi 99,02 %, dan recall 98,97 %; sedangkan pada COVID-CT, akurasi 93,33 % dengan presisi 93,60 % dan recall 92,97 %. Hasil tersebut menunjukkan efektivitas strategi fusi probabilitas dalam mereduksi kesalahan klasifikasi dan menjadikannya salah satu performa terbaik yang dilaporkan untuk deteksi COVID-19 berbasis CT hingga saat ini. Hasil review dari penelitian ditunjukkan pada Tabel 1.

Tabel 1. Review Klasifikasi Citra dengan *Hyperparameter* dan *Ensemble Learning*

No	Author	Metode	Hyperparameter tuning	Point of interest	Hasil
1	W. Hussain et al. [23]	• <i>Transfer-learning</i> VGG16, VGG19, DenseNet201 (ImageNet weights) • Tambahan layers: AvgPooling2D → Flatten → Dense 32 (ReLU) → Dropout → Dense 4 (soft-max) • Prediksi digabung weighted-average ensembling	• <i>Learning rate</i> = 3×10^{-4} (Adam); Hidden layer = 32 neuron; Dropout = 0,20; Batch size = 32 • Kombinasi bobot <i>ensemble</i> dicari grid ($w_1, w_2, w_3 \in 0,1 \dots 0,6$)	• Dataset Kermany OCT (10 336 cuplikan, seimbang 4 kelas: CNV, DME, Drusen, Normal). • Normalisasi min-max + augmentasi (shift 0,2; flip; rotasi 15°; rescale 1/255). • Banding 9 arsitektur; VGG16, VGG19, DenseNet201 paling stabil; <i>ensemble</i> untuk <i>robust</i> & kurangi <i>over-fit</i>.	<i>Ensemble</i> memperoleh akurasi 83 % (prec/rec/f1: CNV 0,88; DME 0,85; Drusen 0,76; Normal 0,83) – lebih tinggi $\pm$ 5–22 pp dari model tunggal terbaik. Confusion-matrix menunjukkan perbaikan umum dan pengurangan salah klasifikasi antar-penyakit.
2	L. Toto et al. [24]	• Model IRF → UNet + relative-layer distance map	• Epoch = 200 • <i>Learning rate</i> = 3×10^{-4} • <i>Optimizer</i> = Adam (implicit)	• Dataset RETOUCH (70 volume publik = 6 754 B-scan)	Rata-rata Dice multi-kelas pada 3 split: 71,86 % vs 70 % (SFU) → +1,8 % Per-model: UNet-IRF

No	Author	Metode	Hyperparameter tuning	Point of interest	Hasil
		- Model SRF → Structured-Dropout UNet (SD-UNet) - Model PED → UNet + data-augmentation Hasil ketiga digabung (majority/ <i>concatenate</i>)	- Loss = <i>Binary Cross Entropy</i> - Input ukuran setelah crop = 256×512 px - Patch overlap diuji (0,65–0,72) DropBlock pada SD-UNet untuk cegah <i>overfitting</i>	28 eksperimen banding 27 arsitektur → pilih terbaik per jenis cairan <i>Ensemble one-to-one</i> meningkatkan robustnes dibanding single UNet SFU (<i>state-of-art</i>)	+1 %; SD-UNet-SRF +5 %; UNet-PED +2 % dibanding SFU terlatih masing-masing. Deteksi visual menunjukkan lebih sedikit <i>false-positive</i> pada vendor Cirrus, Spectralis, Topcon.
3	M Rahil et al. [25]	- Branch 1 – YOLOv7 untuk object detection hard exudates (HE) - Branch 2 – <i>Ensemble</i> (ConvNeXt-tiny ×2 + RegNetX_016, soft voting) untuk klasifikasi DRIL	YOLOv7 (HE): - Input resize = 1024×1024 px - Batch size = 16 - Momentum = 0,937 <i>Learning rate</i> = 0,01 – 0,10 ConvNeXt-tiny #1 (DRIL): - Epoch = 50, LR = 0,01, Batch = 16, Input = 434×1024 px - ConvNeXt-tiny #2: Epoch = 100, LR = 0,001, Batch = 16 RegNetX_016: Epoch = 50, LR = 0,001, Batch = 16	- Dataset = 442 OCT mata DME (annotasi 6 847 HE & DRIL yes/no). <i>Pipeline</i> ganda: deteksi HE + klasifikasi DRIL → digabung (tool pendukung klinisi). - Edge-cleaning & IOU-merging pasca-YOLO; <i>soft voting ensemble</i> untuk <i>robust</i> DRIL. - Inference < 1 detik/gambar (workstation V100).	HE detection: Average Precision = 34,4 %; Precision = 48,7 %; Recall = 43,1 %; korelasi hitung HE vs ahli $r = 0,804$ ($p < 0,05$). DRIL classification (<i>ensemble</i>): Akurasi = 91,1 % (CI95 % 0,852–0,969), Sens = Spes = 91,1 %, AUC = 0,91, Kappa = 0,82 (“almost perfect”).
4	Uddin et al. [21]	EGACNN = CNN + Genetic Algorithm (stacking)	- Jumlah layer CNN: 2 Conv ($32 \times 3 \times 3$, $32 \times 3 \times 3$) + 1 Dense 128 - Kernel size: 3×3 - Jumlah filter puncak: 64 <i>Learning rate</i>: 0,001 - Batch size: 32 - Dropout rate: 0,50	GA otomatis mencari kombinasi di atas—mengurangi trial-and-error manual dan mencegah <i>overfitting</i> .	Akurasi 99,91 % (epoch 5); presisi 98,79 %, recall 98,78 %, F1 98,78 %. Mengungguli sembilan arsitektur pembandingan.
5	N. S. Shaik and T. K. Cherukuri [26]	Transfer-learning + novel ensemble: delapan CNN pralatih (VGG16, VGG19, ResNet50, ResNet50V2, InceptionV3, Xception, MobileNet, InceptionResNetV2) dengan konvolusi dasar dibekukan → kepala klasifikasi ber-regulasi (Dense 512-512-256 + Dropout 0,3) → probabilitas <i>Softmax</i> tiap model dipadukan dalam “composite neural network” ($2 \times \text{FC-256} + \text{Dropout } 0,2$) untuk keputusan akhir.	<i>Candidate models</i>: LR 0,003, batch 32, epoch 500, Dropout 0,3, Adam, L1 (0,01) / L1L2 bias. <i>Ensemble</i>: LR 0,001, batch 32, epoch 100, Dropout 0,2, Adam.	- Fusi probabilitas (bukan majority-vote) mengurangi salah klasifikasi. - Validasi hold-out 80/20 pada dua dataset (SARS-CoV-2 & COVID-CT). Menetapkan <i>state-of-the-art</i> untuk deteksi COVID-19	Akurasi 98,99 %, Presisi 99,02 %, Recall 98,97 % pada SARS-CoV-2; Akurasi 93,33 %, Presisi 93,60 %, Recall 92,97 % pada COVID-CT — unggul atas Max Voting, Bagging, Boosting, dan model individu.

B. Ensemble Learning dengan Hyperparameter Tuning

Analisis komparatif terhadap penelitian-penelitian terkini menunjukkan bahwa model *ensemble* yang diusulkan dalam penelitian ini memiliki keunggulan yang jelas dalam metodologi fusi, dan pendekatan optimasi. Sementara model-model lain seperti yang diusulkan [24] dan [25] dirancang untuk deteksi biomarker atau segmentasi cairan, model ini memiliki keunggulan diagnostik yang lebih luas. Dengan kemampuannya melakukan klasifikasi *end-to-end* terhadap empat kelas penyakit utama, model ini menjawab pertanyaan klinis yang lebih fundamental dan komprehensif, serupa dengan tujuan pada penelitian [21].

Secara metodologi pendekatan fusi pada level fitur (*feature-level fusion*) yang diimplementasikan merupakan kemajuan dibandingkan metode penggabungan pada level prediksi, seperti *weighted-average* oleh [21] atau fusi probabilitas oleh [26]. Dengan menggabungkan vektor fitur mentah yang kaya informasi sebelum direduksi menjadi skor probabilitas, model ini mampu mempelajari korelasi antar-fitur yang lebih kompleks dari kelima backbone secara bersamaan, sebuah kapabilitas yang secara teoritis hilang pada fusi level prediksi.

Keunggulan lainnya pendekatan optimasi *hyperparameter* yang holistik dan sistematis. Berbeda dengan sebagian besar penelitian referensi yang cenderung menentukan *hyperparameter* secara manual atau hanya mengoptimalkan sebagian kecil parameter [21], [23], [24], [25], [26], penelitian ini menerapkan pencarian stokastik (*RandomSearch*) yang ekstensif terhadap 12 parameter kritis. Proses ini tidak hanya menyetel parameter pelatihan, tetapi juga parameter struktural arsitektur itu sendiri, seperti jumlah neuron di setiap cabang dan di classifier fusi. Ini menunjukkan bahwa performa tinggi yang dicapai bukanlah hasil dari arsitektur yang statis seperti pada penelitian [23] yang mengoptimalkan satu CNN dasar melainkan dari sebuah proses penemuan untuk menemukan konfigurasi *ensemble* yang paling optimal secara empiris. Kombinasi dari cakupan diagnostik yang general, strategi fusi di level fitur yang lebih kaya informasi, dan metodologi optimasi yang mendalam ini menempatkan model yang diusulkan sebagai sebuah

solusi yang secara konseptual lebih maju, lebih fleksibel, dan terbukti sangat andal untuk klasifikasi penyakit retina.

