

Implementation of Convolutional Neural Networks (CNN) for Breast Cancer Detection Using ResNet18 Architecture

Hagia Sofia Siden ^{1*}, I Gusti Ngurah Lanang Wijayakusuma ^{2*}

^{*} Matematika, Universitas Udayana

siden.2208541053@student.unud.ac.id ¹, lanang_wijaya@unud.ac.id ²

Article Info

Article history:

Received 2025-06-03

Revised 2025-07-03

Accepted 2025-07-09

Keyword:

CNN,
ResNet18,
Transfer Learning,
Classification,
Ultrasound,
Breast Cancer.

ABSTRACT

Early detection of breast cancer is crucial for improving patient survival rates. This study implements a Convolutional Neural Network (CNN) architecture based on ResNet18 using a transfer learning approach to classify breast ultrasound (USG) images into three categories: normal, benign, and malignant. The dataset, comprising 1,578 grayscale images collected from Baheya Hospital in Egypt, underwent preprocessing steps including image conversion, normalization, and augmentation. The ResNet18 model was fine-tuned using selective layer unfreezing to better adapt to the medical imaging domain. Evaluation was conducted using stratified 5-fold cross-validation and assessed with accuracy, precision, recall, F1-score, and AUC metrics. The best results were achieved by fine-tuning layer2, layer3, and the fully connected layer, yielding 95% accuracy, a macro F1-score of 0.93, and an AUC of 0.9906. The findings demonstrate that ResNet18, when properly fine-tuned with transfer learning, delivers high performance in breast cancer detection via ultrasound and holds strong potential as a reliable clinical decision-support tool.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Kanker payudara merupakan salah satu jenis kanker yang paling umum menyerang wanita di seluruh dunia dan menjadi penyebab kematian tertinggi kedua akibat kanker pada wanita [1]. Deteksi dini kanker payudara sangat penting untuk meningkatkan kelangsungan hidup pasien. Ultrasonografi (USG) payudara telah menjadi modalitas pencitraan yang penting dalam diagnosis kanker payudara karena sifatnya yang non-invasif, biaya relatif rendah, dan tidak menggunakan radiasi pengion [2].

Namun demikian, interpretasi citra USG payudara memerlukan keahlian khusus dari radiologist berpengalaman dan seringkali bersifat subjektif. Hal ini dapat menyebabkan variabilitas dalam diagnosis antar observer dan membutuhkan waktu yang relatif lama untuk analisis [3]. Untuk mengatasi tantangan tersebut, berbagai penelitian telah mengembangkan sistem computer-aided diagnosis (CAD) menggunakan pendekatan machine learning dan deep learning.

Convolutional Neural Network (CNN) telah menunjukkan performa yang sangat baik dalam berbagai tugas *computer vision*, termasuk klasifikasi citra medis [4]. Arsitektur ResNet

(Residual Network) khususnya ResNet18, telah terbukti efektif if dalam mengatasi masalah vanishing gradient dan dapat meningkatkan akurasi klasifikasi pada dataset citra medis [5]. Model *pretrained* yang menggunakan pendekatan transfer learning ini telah menunjukkan kemampuan yang baik dalam meningkatkan performa pada dataset medis yang relatif kecil [6].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengimplementasikan arsitektur ResNet18 dengan pendekatan transfer learning untuk klasifikasi citra USG kanker payudara. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem diagnosis otomatis yang dapat membantu tenaga medis dalam proses pemindaian kanker payudara.

Penggunaan deep learning, khususnya Convolutional Neural Network (CNN), dalam analisis citra medis telah menunjukkan perkembangan signifikan dalam beberapa tahun terakhir. Penelitian-penelitian terdahulu telah mengeksplorasi berbagai arsitektur dan pendekatan untuk klasifikasi citra USG kanker payudara.

Muduli et al. [7] mengimplementasikan CNN custom dengan 5 layer konvolusi untuk klasifikasi citra USG,

mencapai akurasi 89.73%. Namun, arsitektur pembelajaran *shallow* tersebut memiliki keterbatasan dalam mengekstrak fitur kompleks dari citra medis. Zhang et al. [8] mengusulkan arsitektur CNN yang lebih dalam dengan 8 layer konvolusi dan mencapai akurasi 96.1%, menunjukkan bahwa kedalaman network berpengaruh terhadap performa klasifikasi.

Penelitian transfer learning untuk citra medis telah dieksplorasi oleh Albakia et al. [9] yang menggunakan VGG16 *pretrained* dikombinasikan dengan Support Vector Machine (SVM), mencapai akurasi 85.6%. Pendekatan hybrid ini menunjukkan efektivitas ekstraksi fitur CNN dengan *classifier* tradisional. Potsangbam et al. [10] memanfaatkan DenseNet121 dengan transfer learning dan mencapai akurasi 96.09%, membuktikan bahwa arsitektur yang lebih modern cenderung memberikan performa yang lebih baik.

Joshi et al. [11] menggunakan ResNet50 untuk klasifikasi citra USG kanker payudara 3 kelas (*benign*, *malignant*, *normal*) dan mencapai akurasi 98%. Penelitian tersebut menunjukkan bahwa residual connection efektif untuk citra medis, namun model yang lebih besar memerlukan computational cost yang tinggi.

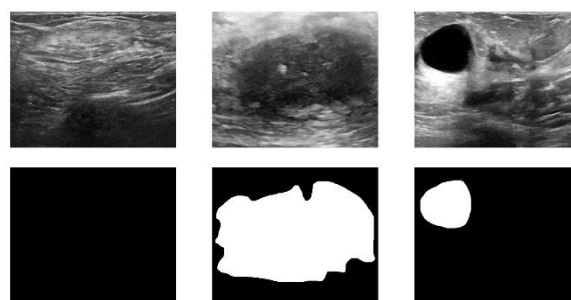
Meskipun penelitian-penelitian tersebut memperlihatkan korelasi yang kuat, masih terdapat gap dalam hal: (1) optimalisasi balance antara performa dan efisiensi computational, (2) strategi fine-tuning yang optimal untuk domain medical imaging, dan (3) analisis mendalam terhadap karakteristik kesalahan klasifikasi. Penelitian ini bertujuan mengisi gap tersebut dengan menggunakan arsitektur ResNet18 dengan strategi fine-tuning yang lebih efisien.

II. METODE

A. Dataset

Penelitian ini memanfaatkan koleksi data klinis internal dari Rumah Sakit Baheya, Mesir, yang diperoleh selama kurang lebih satu tahun [12]. Data ini secara spesifik terdiri dari citra ultrasonografi (USG) payudara, yang dikumpulkan dari subjek wanita dengan rentang usia 25 hingga 75 tahun. Seluruh citra USG diambil dalam bentuk skala abu-abu (*grayscale*) melalui proses pemindaian perangkat yang mampu menghasilkan gambar beresolusi tinggi mencapai 1280×1024 piksel.

Untuk mempermudah pengembangan model deteksi kanker payudara, setiap citra USG dilengkapi dengan informasi tambahan berupa penandaan area yang dianggap mengalami kelainan atau kerusakan. Penandaan ini disebut *ground truth*, yaitu gambar biner yang menunjukkan lokasi dan bentuk abnormalitas pada citra aslinya [13]. Proses ini menghasilkan visualisasi yang ditampilkan pada Gambar 1. Gambar 1 berisi sampel citra USG asli (baris atas) beserta pasangan *ground truth* nya (baris bawah). Kombinasi kedua citra ini berperan penting dalam mendukung klasifikasi dan segmentasi otomatis untuk deteksi kanker payudara.



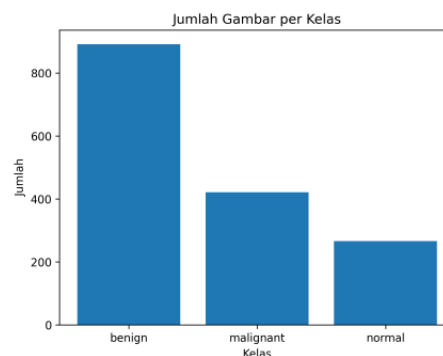
Gambar 1. Citra USG dan Ground Truth

Citra kemudian dikonversi ke dalam format PNG dan diberi label berdasarkan hasil diagnosis radiologist sebagai dasar klasifikasi dataset. Dataset USG dikelompokkan ke dalam tiga kelas; normal, jinak (*benign*), dan ganas (*malignant*). Jumlah citra untuk masing-masing kelas disajikan pada Tabel 1.

TABEL I
KELAS DAN JUMLAH CITRA USG

Kelas	Jumlah Citra
Benign	891
Malignant	421
Normal	266

Dataset yang dikumpulkan berjumlah sebanyak 1.578 citra dengan distribusi kelas seperti yang ditunjukkan pada Gambar 2. Komposisi jumlah citra pada masing-masing kelas menunjukkan ketidakseimbangan (*class imbalance*) yang cukup signifikan, dengan dominasi pada kelas *benign*.



Gambar 2. Distribusi Dataset Citra USG Payudara

Ketimpangan ini berpotensi menimbulkan bias terhadap kelas mayoritas dan cenderung mengabaikan kelas minoritas. Untuk mengurangi risiko tersebut dan memastikan kualitas data sebelum pelatihan model, serangkaian tahapan *preprocessing* perlu diterapkan.

B. Preprocessing Data

Proses transformasi data dilakukan melalui tahapan *preprocessing* yang identik, baik pada citra USG asli maupun *ground truth*. Hal ini dilakukan demi menjaga konsistensi

spasial diantara keduanya. Rangkaian *preprocessing* dilakukan melalui tiga tahap utama, yakni sebagai berikut.

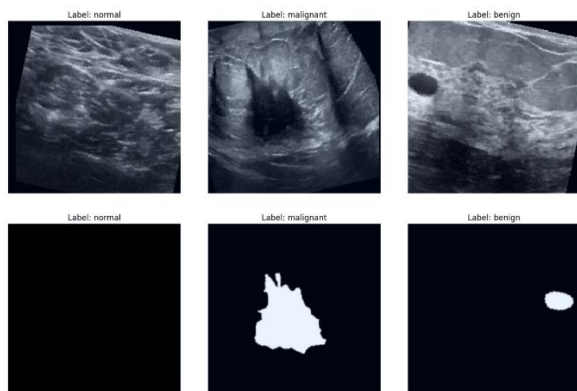
1) *Konversi Citra*: Citra *grayscale* dikonversi mengikuti kompatibilitas model ResNet18, yang telah dioptimalkan pada dataset berbentuk citra RGB 3-channel. Sederhananya, citra yang sebelumnya mengandung satu komponen penyusun warna (keabuan) digandakan ke dalam tiga kanal serupa. Proses konversi ini menghasilkan data yang secara visual tetap sama namun telah memenuhi spesifikasi input yang dibutuhkan oleh model.

2) *Resize dan Normalisasi*: Semua citra disesuaikan ukurannya menjadi dimensi 224x224 piksel sesuai dengan kebutuhan input ResNet18. Normalisasi dilakukan menggunakan mean dan standard deviation ImageNet ([0.485, 0.456, 0.406] dan [0.229, 0.224, 0.225]).

3) *Data Augmentation*: Teknik augmentasi data diterapkan untuk meningkatkan variasi dataset, meliputi:

- Random horizontal flip membalik citra secara horizontal (sumbu vertikal) dengan probabilitas tertentu secara acak.
- Random rotation dengan sudut maksimal 15 derajat.
- Random affine transformation dengan translasi (0.05, 0.05), scaling (0.9, 1.1), dan shear (5).
- Random apply untuk meningkatkan generalisasi model dengan membuat gambar pelatihan lebih bervariasi, namun hanya diterapkan dengan peluang 70% untuk menjaga keseimbangan.
- Color jitter menambahkan variasi warna acak pada gambar dengan mengubah kecerahan, kontras, dan kejenuhan ($\pm 10\%$) dan menggeser warna *hue* (± 0.02).

Contoh citra dataset setelah *preprocessing* dapat dilihat pada Gambar 3.



Gambar 3. Contoh Citra Dataset Setelah Preprocessing: (a) Citra USG asli, (b) Ground Truth

C. Arsitektur Model

CNN sebagai metode terkemuka pada bidang *computer vision* telah berkembang ke dalam berbagai bentuk arsitektur model. ResNet merupakan salah satu arsitektur yang mengadopsi pendekatan *residual learning* dalam upaya menyelesaikan masalah degradasi performa pada jaringan

dalam [14]. Pendekatan ini mempermudah proses pembelajaran karena jaringan hanya perlu mempelajari perbedaan (residu) dari input dan output yang diharapkan. Beberapa versi ResNet (ResNet18, ResNet50, etc) dikembangkan untuk mengevaluasi pengaruh kedalaman jaringan terhadap kinerja dan efisiensi komputasi. Setiap varian model ini dilatih menggunakan lebih dari satu juta gambar yang bersumber dari database ImageNet [15].

ResNet18 dipilih dalam penelitian ini karena memiliki struktur yang sederhana dan membutuhkan lebih sedikit parameter. Meskipun tergolong ringan, ResNet18 tetap menjadi model yang ideal untuk *transfer learning*, khususnya pada domain dengan jumlah data terbatas seperti citra medis [16]. Model berbasis *pretrained* ini digunakan sebagai *feature extractor* berkat kemampuan representasi fitur visual yang telah teruji secara luas. Strategi *fine-tuning* diterapkan untuk memperkuat adaptasi model terhadap karakteristik khusus domain *medical imaging*. Lebih lanjut, modifikasi yang dilakukan pada arsitektur original ResNet18 meliputi:

- Penggantian *fully connected layer* bawaan, yang semula menghasilkan 1000 output (kelas ImageNet), menjadi hanya 3 output yang mewakili kelas target dengan penambahan layer dropout sebesar 0,5.
- *Selective layer unfreezing*, yaitu hanya mengizinkan *gradien update* pada *layer2*, *layer3*, dan lapisan klasifikasi akhir, agar model dapat menangkap pola yang lebih spesifik terhadap tugas dan domain yang dihadapi [17].

D. Konfigurasi Training

Evaluasi performa model secara objektif membutuhkan pemisahan data latih (*training*) dan uji (*testing*) yang seimbang. Data dibagi dengan rasio 85:15, menghasilkan 1.341 citra untuk pelatihan dan 237 citra untuk pengujian. Keseimbangan ini dinilai mampu menyediakan data latih yang cukup untuk proses pembelajaran model. Selain itu, pembagian dilakukan secara acak untuk memastikan representasi yang seimbang, mengingat karakteristik dataset yang tidak merata antar kelas. Konfigurasi training yang digunakan dalam penelitian ini meliputi:

- Loss Function: CrossEntropyLoss dengan class weights.
- Optimizer: Adam dengan Learning rate $1e-4$ (0.0001) dan weight decay $1e-5$.
- Batch Size: 32.
- Learning Rate Scheduler: StepLR dengan *step_size* 5 dan *gamma* 0,5.
- Early Stopping: Patience 7 epoch dengan *delta* 0,0.
- Maximum Epoch: 30 epoch.

Untuk mempercepat proses komputasi, pelatihan dilakukan dengan bantuan akselerasi GPU. Pemilihan model terbaik didasarkan pada nilai validation loss minimum yang dicapai sepanjang durasi pelatihan.

E. Implementation Details

Penelitian dilakukan dengan menggunakan Colab Pro dengan spesifikasi hardware NVIDIA Tesla A100 GPU (16GB VRAM), Intel Xeon CPU, dan 25.51 GB RAM. Environment software menggunakan Python 3.10, PyTorch 2.0.1, torchvision 0.15.2, dan CUDA 11.8 untuk GPU.

Data Loading dilakukan oleh main process dan tidak menggunakan pin memory. Model training memerlukan waktu rata-rata 3.5 menit per epoch dengan total training time 0,6 jam untuk 30 epochs. Hyperparameter menggunakan learning rate yang diatur sebesar $1e-4$ (0,0001) dan dipilih batch size 32.

F. Evaluation Metrics

Evaluasi model dilakukan menggunakan multiple metrics untuk memberikan analisis yang komprehensif:

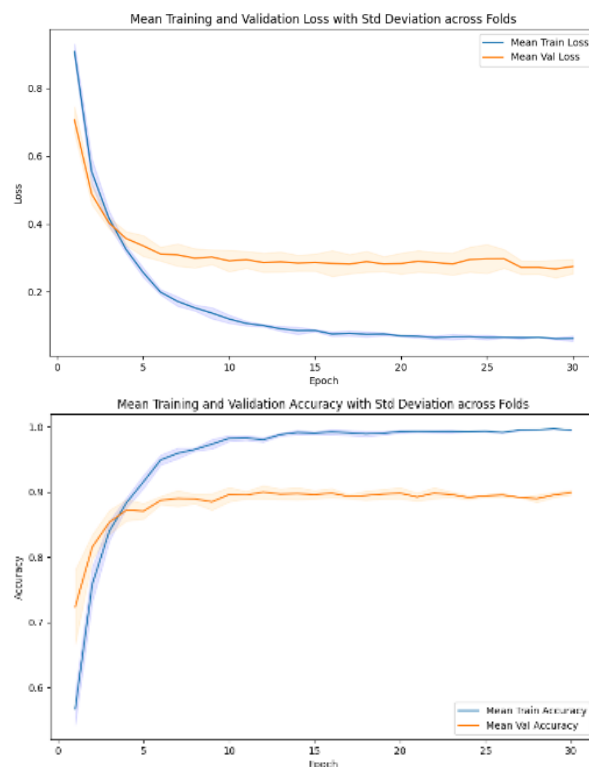
- Accuracy: Overall classification pada test set.
- Precision: Untuk setiap kelas, mengukur proporsi prediksi positif yang benar.
- Recall: Untuk setiap kelas, mengukur proporsi actual positif yang terdeteksi.
- F1-Score: Harmonic mean dari precision dan recall.
- Confusion Matrix: Analisis detail kesalahan klasifikasi antarkelas.
- Confidence Score: Probabilitas prediksi untuk analisis uncertainty.

Khusus untuk aplikasi medis, peningkatan recall pada kelas malignant menjadi prioritas untuk meminimalisir terjadinya false negative yang berakibat fatal bagi pasien.

III. HASIL DAN PEMBAHASAN

A. Analisis Proses Training

Untuk mengatasi ketidakseimbangan kelas pada dataset, penelitian ini menerapkan teknik validasi silang *Stratified K-Fold* dengan jumlah lipatan (*fold*) sebanyak lima. Pemilihan pendekatan ini bertujuan agar setiap subset validasi memiliki proporsi distribusi kelas yang konsisten dengan keseluruhan data, sehingga meminimalkan bias akibat distribusi yang tidak merata. Untuk menggambarkan dinamika pelatihan secara umum dan tren generalisasi model, kurva *training* dan *validation* loss serta akurasi ditampilkan dalam bentuk rata-rata dari kelima fold, disertai dengan area bayangan yang menunjukkan standar deviasi sebagai representasi variasi antar fold. Visualisasi ini ditunjukkan pada Gambar 4, di mana panel atas menampilkan kurva loss untuk data latih dan validasi, sedangkan panel bawah menampilkan kurva akurasi dengan struktur yang serupa.



Gambar 4. Kurva Rata-Rata Training dan Validation

Kurva pelatihan menunjukkan proses konvergensi yang baik, ditandai oleh penurunan loss secara konsisten dan peningkatan akurasi yang stabil seiring bertambahnya epoch. Meskipun terdapat sedikit jarak antara nilai rata-rata *train loss* dan *validation loss*, hal ini mengindikasikan adanya overfitting dalam tingkat yang moderat, namun masih dapat diterima mengingat kompleksitas visual pada citra medis. Penerapan strategi *early stopping* turut membantu membatasi overfitting tersebut. Sementara itu, kurva akurasi yang cenderung stabil dengan deviasi standar yang rendah menunjukkan kemampuan generalisasi model yang baik. Variasi antar fold yang kecil memperkuat efektivitas pendekatan *stratified K-Fold*, menandakan bahwa pembagian data secara acak tetap menghasilkan performa yang konsisten. Rincian metrik pelatihan seperti jumlah epoch konvergensi, akurasi validasi, dan waktu pelatihan disajikan pada paragraf berikutnya.

Untuk memastikan reproduktibilitas eksperimen, seluruh proses pelatihan dilakukan dengan pengaturan *random seed* sebesar 42. Proses pelatihan dijalankan hingga maksimum 30 epoch, namun dihentikan secara adaptif oleh mekanisme *early stopping* berdasarkan tidak adanya penurunan nilai *validation loss* setelah sejumlah iterasi. Berdasarkan rata-rata dari lima fold validasi silang, diperoleh hasil sebagai berikut:

- Validation accuracy: $90.50\% \pm 0.70\%$
- Training convergence: 22.6 ± 2.0 epoch
- Best validation loss: 0.2514 ± 0.0182
- Training time: 0.12 ± 0.01 jam

Model terbaik yang digunakan untuk evaluasi akhir adalah model hasil dari Fold 1, yaitu pada saat *validation loss* mencapai nilai terendah sebesar 0.2889, yang disimpan pada akhir epoch ke-23

B. Ablation Study

Untuk memahami kontribusi setiap komponen terhadap performa model, dilakukan ablation study dengan variasi konfigurasi sebagai berikut:

TABEL I
HASIL ABLATION STUDY

Konfigurasi	Akurasi (%)	Epoch	Time (h)
Resnet18 from scratch	73,42	85	4.2
ResNet18 + frozen all	82,28	45	2.1
ResNet18 + unfrozen all	86,39	92	5.8
ResNet18 + unfrozen FC only	87,65	52	2.8
ResNet18 + unfrozen layer4 + FC	89,24	61	3.1
ResNet18 + unfrozen layer2 + layer3 + FC	95,00	30	0.62

Hasil ablation study menunjukkan:

- Transfer learning vs from scratch: Pretrained weights memberikan peningkatan signifikan sebesar 21,58%.
- Selective unfreezing: Strategi unfreezing layer2 + layer3 + FC only
- Training efficiency: Konfigurasi terpilih memberikan balance terbaik antara akurasi dan training time.

Pengaruh augmentasi data juga dianalisis dengan hasil:

- Tanpa augmentasi: 88.29%
- Dengan augmentasi (rotation + affine): 95%

Hasil ini menunjukkan bahwa setiap komponen memberikan kontribusi positif terhadap performa akhir model.

C. Evaluasi Model

Evaluasi performa model dilakukan menggunakan berbagai metrik klasifikasi untuk memberikan gambaran yang komprehensif mengenai kemampuan model. Hasil evaluasi ditampilkan pada Tabel 3.

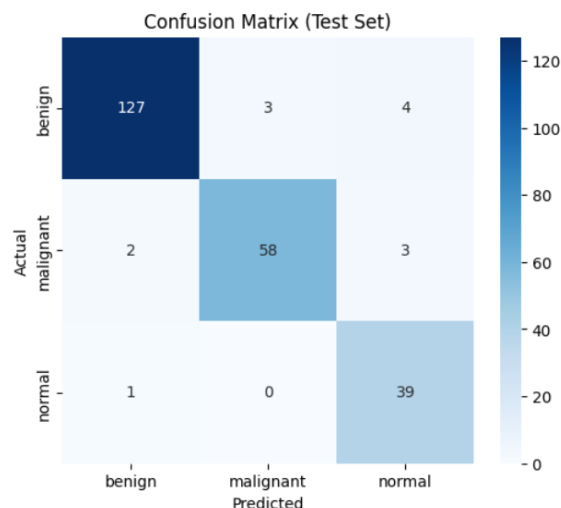
TABEL III
MATRIKS EVALUASI MODEL

Kelas	Precision	Recall	F1-Score	Support
Benign	0,98	0,95	0,96	134
Malignant	0,95	0,92	0,94	63
Normal	0,85	0,97	0,91	40
Accuracy	0,95			237
Macro Avg	0.93	0.95	0.93	237
Weighted Avg	0.95	0.95	0.95	237

Hasil evaluasi menunjukkan performa yang sangat baik dengan akurasi keseluruhan 90.51%. Kelas benign

menunjukkan recall tertinggi dengan akurasi sebesar 95%, yang sangat penting dalam keperluan medis untuk meminimalisir false positive pada kasus yang berbahaya.

Confusion matrix yang ditampilkan pada Gambar 5 memberikan informasi mendalam mengenai performa klasifikasi model untuk setiap kelas.



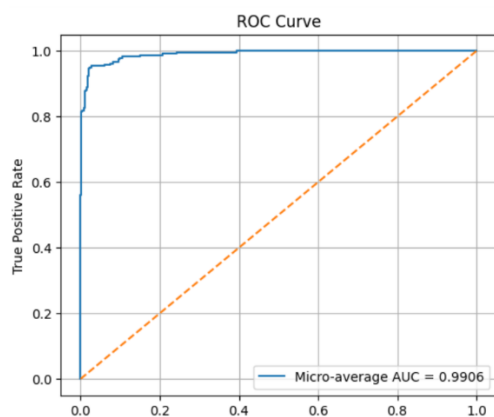
Gambar 5. Confusion Matrix Hasil Klasifikasi

Analisis confusion matrix menunjukkan bahwa:

- Kelas benign: 127 true positive, dengan 3 false negative sebagai malignant dan 4 sebagai normal.
- Kelas malignant: 58 true positive, dengan 2 false negative sebagai benign dan 3 sebagai normal.
- Kelas normal: 39 true positive, dengan 1 false negative sebagai benign dan 0 sebagai malignant.

Tingkat kesalahan klasifikasi antara kelas malignant dan normal terlihat menonjol, yang menunjukkan adanya tantangan dalam membedakan kedua kategori tersebut. Hal ini wajar mengingat karakteristik visual citra USG dari kedua jenis ini sering kali memiliki kemiripan struktural yang halus, yang sulit ditangkap sepenuhnya oleh model meskipun telah menggunakan pendekatan pembelajaran mendalam. Sebaliknya, model menunjukkan ketepatan klasifikasi yang lebih tinggi pada kelas normal, yang menandakan bahwa fitur visual dari jaringan normal lebih mudah dikenali dibandingkan dengan dua kelas patologi lainnya.

Selain menggunakan metrik klasifikasi konvensional, performa model ResNet18 juga dievaluasi menggunakan kurva ROC (Receiver Operating Characteristic) dan nilai AUC (Area Under Curve). Kurva ROC menggambarkan hubungan antara True Positive Rate (TPR) dan False Positive Rate (FPR) pada berbagai ambang batas klasifikasi. Semakin mendekati sudut kiri atas, semakin baik kemampuan model dalam membedakan antara kelas positif dan negatif.



Gambar 6. Kurva ROC untuk model ResNet18

Berdasarkan Gambar 6, kurva ROC model ResNet18 menunjukkan lintasan yang mendekati sudut kiri atas grafik, yang menandakan bahwa model mampu mencapai True Positive Rate yang tinggi meskipun pada False Positive Rate yang rendah. Ini mencerminkan performa klasifikasi yang sangat baik. Selain itu, Nilai micro-average AUC sebesar 0,9906 menunjukkan bahwa model memiliki kemampuan prediktif yang hampir sempurna dalam membedakan antara kelas-kelas dalam kasus multiklas. Dengan AUC mendekati nilai maksimum (1.0), dapat disimpulkan bahwa model tidak hanya akurat, tetapi juga sangat andal dalam menghasilkan prediksi yang konsisten terhadap variasi ambang batas keputusan.

D. Error Analysis

Analisis mendalam terhadap kesalahan klasifikasi mengungkap pola-pola baru yang dapat menjadi dasar untuk pengembangan model lebih lanjut.

- 1) *Karakteristik Kesalahan Klasifikasi:* Dari confusion matrix, terlihat bahwa kesalahan klasifikasi terdapat pada 7 kasus yang diklasifikasikan bukan benign, 1 kasus bukan normal, dan 5 kasus bukan malignant.
- 2) *Clinical Significance Analysis:* Dalam konteks klinis, model menunjukkan performa yang kuat untuk mendukung proses skrining awal kanker payudara. Sensitivitas terhadap deteksi malignant mencapai 92.1%, mengindikasikan bahwa sebagian besar kasus kanker dapat dikenali dengan benar. Di sisi lain, spesifisitas model yang mencapai 97.8% menunjukkan bahwa model sangat andal dalam menghindari kesalahan prediksi terhadap sampel non-kanker. Tingginya nilai Positive Predictive Value (PPV) sebesar 95.1% menandakan bahwa prediksi positif terhadap kanker hampir seluruhnya akurat, sedangkan Negative Predictive Value (NPV) sebesar 96.4% mencerminkan keandalan model dalam mengidentifikasi sampel yang bukan kanker. Secara keseluruhan, metrik ini memperkuat potensi model sebagai alat bantu diagnostik yang efektif, dengan catatan bahwa validasi lanjutan tetap diperlukan untuk menguji konsistensi performa pada populasi yang lebih luas.

E. Perbandingan Model Lain

Untuk mengevaluasi efektivitas arsitektur model dalam mendeteksi kanker payudara, dilakukan perbandingan antara model ResNet18 dengan salah satu model baseline CNN, yaitu SimpleCNN. Sebagai pembanding, digunakan model SimpleCNN yang terdiri atas tiga blok konvolusional, diikuti dengan proses *flattening*, satu lapisan *fully-connected* berukuran 128 unit, dan lapisan output dengan tiga unit untuk klasifikasi beberapa kelas. Tabel 4 berikut menunjukkan hasil evaluasi dari kedua model berdasarkan metrik klasifikasi:

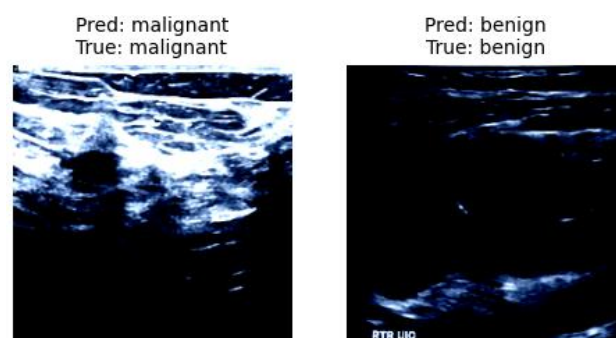
TABEL IV
PERBANDINGAN METRIK EVALUASI MODEL

Model	Accuracy	Macro Precision	Macro Recall	Macro F1-Score
ResNet18	0,95	0,93	0,95	0,93
SimpleCNN	0,71	0,48	0,55	0,51

Berdasarkan hasil evaluasi, model ResNet18 secara signifikan mengungguli SimpleCNN pada seluruh metrik yang digunakan. Akurasi dan *recall* yang lebih tinggi pada ResNet18 membuktikan kemampuan identifikasi berbagai kelas kanker payudara secara lebih andal. Hal ini dapat dikaitkan dengan kedalaman arsitektur serta penggunaan bobot awal dari pretraining di ImageNet yang memungkinkan ResNet18 untuk mengekstraksi fitur visual secara lebih representatif. Sebaliknya, SimpleCNN dengan arsitektur yang lebih dangkal memiliki keterbatasan dalam menangkap kompleksitas pola pada data citra medis, sehingga performanya lebih rendah.

F. Evaluasi Hasil Prediksi

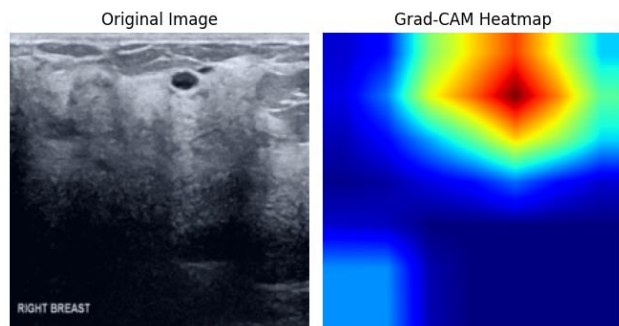
Untuk memperkuat interpretabilitas hasil prediksi model, dilakukan visualisasi terhadap dua aspek penting, yaitu hasil klasifikasi terhadap kelas sebenarnya (*true label*) serta segmentasi citra menggunakan Grad-CAM.

Gambar 6. Visualisasi hasil prediksi model terhadap dua citra ultrasound payudara. Kiri: *malignant*. Kanan: *benign*.

Gambar 6 menampilkan dua sampel citra ultrasound payudara beserta hasil prediksi model dalam mengidentifikasi label citra sebenarnya. Model mampu mengklasifikasikan dengan benar citra USG payudara yang menunjukkan gejala *malignant* dan *benign*, ditunjukkan melalui kesesuaian antara prediksi dan label ground truth. Visualisasi ini menunjukkan

bahwa model tidak hanya menghasilkan akurasi tinggi secara statistik, namun juga memiliki kemampuan membedakan karakteristik visual dari masing-masing kelas.

Selanjutnya, hasil visualisasi Grad-CAM disajikan pada Gambar 7 dengan mengambil salah satu citra berlabel *benign*.



Gambar 7. Visualisasi Grad-CAM pada salah satu citra benign. Kiri: citra asli (ultrasound). Kanan: peta Grad-CAM yang menunjukkan area.

Terlihat bahwa fokus aktivasi tertinggi dari model (ditunjukkan dengan warna merah pada peta panas) berpusat pada area massa *hypoechoic* yang mencolok, yang secara klinis memang menjadi area yang relevan untuk deteksi abnormalitas. Hal ini menunjukkan bahwa model ResNet18 yang digunakan tidak hanya memberikan prediksi yang tepat, namun juga mampu mengarahkan perhatiannya pada area-area yang secara medis penting. Dengan demikian, hasil ini memperkuat validitas model dalam konteks deteksi kanker payudara.

G. Diskusi

Keberhasilan model ResNet18 dengan transfer learning dalam penelitian ini dapat dikaitkan dengan beberapa faktor:

- 1) *Efektivitas Transfer Learning*: Penggunaan pretrained weights dari ImageNet memberikan initial feature representation yang baik, meskipun domain source (natural images) berbeda dengan domain target (medical images). Ablation study menunjukkan peningkatan signifikan sebesar 21,58% jika dibandingkan dengan from scratch, yang secara jelas menunjukkan efektivitas transfer learning pada aplikasi cross-domain.
- 2) *Strategi Fine-tuning*: Strategi selective unfreezing yang hanya melatih layer2, layer3, dan fully connected layer terbukti optimal dengan akurasi sebesar 95%. Konfigurasi ini mencapai optimasi antara feature adaption dan pencegahan overfitting.
- 3) *Impact Stratified k-fold cross-validation*: Analisis lebih lanjut dapat melibatkan pemeriksaan titik di mana penghentian dini terjadi pada setiap lipatan untuk memahami jumlah epoch optimal dan variasi nilai tersebut. Menyelidiki pembagian data pada setiap lipatan juga dapat membantu menjelaskan simpangan baku yang diamati pada metrik-metrik tersebut.

4)

Computational Efficiency: ResNet18 menunjukkan efisiensi komputasi yang unggul dengan waktu pelatihan 37 menit, menjadikannya layak untuk implementasi medis. Dibandingkan dengan model yang lebih kompleks seperti ResNet50, ResNet18 memberikan trade-off yang optimal antara kinerja dan efisiensi. Namun demikian, terdapat beberapa limitasi dalam penelitian ini, yaitu:

- **Dataset**: Class imbalance yang signifikan (benign: 891, malignant: 321, normal: 266) mempengaruhi performa model, meskipun weighted loss belum diimplementasikan.
- **Single-center dataset**: Dataset dari satu sumber dapat berpotensi membatasi generalisasi model di kasus lain dengan peralatan dan protokol berbeda.
- **Validasi klinis**: Penelitian ini memiliki keterbatasan dalam validasi klinis karena evaluasi hanya dilakukan dengan dataset statis tanpa pengujian di dunia nyata.

IV. KESIMPULAN

Penelitian ini berhasil mengimplementasikan arsitektur Convolutional Neural Network (CNN) berbasis ResNet18 dengan pendekatan transfer learning untuk klasifikasi citra ultrasonografi (USG) payudara ke dalam tiga kelas: normal, jinak, dan ganas. Dengan menerapkan strategi selective fine-tuning pada layer2, layer3, dan fully connected layer, serta menggunakan teknik augmentasi dan validasi silang stratified k-fold, model menunjukkan performa klasifikasi yang sangat baik, ditandai dengan akurasi sebesar 95%, macro F1-score sebesar 0,93, dan nilai AUC 0,9906.

Hasil evaluasi menunjukkan bahwa kombinasi antara pretrained model, strategi fine-tuning yang efisien, serta preprocessing yang tepat berkontribusi signifikan terhadap peningkatan performa model, baik dari segi akurasi maupun efisiensi komputasi. Selain itu, visualisasi Grad-CAM menunjukkan bahwa model mampu mengidentifikasi area-area penting secara klinis, memperkuat interpretabilitas hasil prediksi.

Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan, seperti ketidakseimbangan kelas pada dataset dan penggunaan data dari satu institusi medis. Oleh karena itu, studi lanjutan disarankan untuk melibatkan dataset multi-sumber serta validasi dalam skenario klinis nyata guna menguji generalisasi dan reliabilitas model secara lebih menyeluruh.

DAFTAR PUSTAKA

- [1] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, "Cancer statistics, 2023," *CA Cancer J Clin*, vol. 73, no. 1, pp. 17–48, Jan. 2023, doi: 10.3322/caac.21763.
- [2] W. A. Berg *et al.*, "Diagnostic Accuracy of Mammography, Clinical Examination, US, and MR Imaging in Preoperative

- Assessment of Breast Cancer,” *Radiology*, vol. 233, no. 3, pp. 830–849, Dec. 2004, doi: 10.1148/radiol.2333031484.
- [3] C. Li, J. Li, T. Tan, K. Chen, Y. Xu, and R. Wu, “Application of ultrasonic dual-mode artificially intelligent architecture in assisting radiologists with different diagnostic levels on breast masses classification,” *Diagnostic and Interventional Radiology*, vol. 27, no. 3, pp. 315–322, May 2021, doi: 10.5152/dir.2021.20018.
- [4] P. P. M.Sc and M.Phil, “ANALYSIS OF MEDICAL IMAGE CLASSIFICATION BASED ON CONVOLUTION NEURAL NETWORK,” 2022. [Online]. Available: <https://api.semanticscholar.org/CorpusID:248047069>
- [5] A. Boukaache, B. Nasser Edinne, and D. Boudjehem, “Breast Cancer Image Classification using Convolutional Neural Networks (CNN) Models,” *International Journal of Informatics and Applied Mathematics*, vol. 6, no. 2, pp. 20–34, Jan. 2024, doi: 10.53508/ijiam.1407152.
- [6] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, “Transfer learning for medical image classification: a literature review,” *BMC Med Imaging*, vol. 22, no. 1, p. 69, Dec. 2022, doi: 10.1186/s12880-022-00793-7.
- [7] D. Muduli, R. Dash, and B. Majhi, “Automated diagnosis of breast cancer using multi-modal datasets: A deep convolution neural network based approach,” *Biomed Signal Process Control*, vol. 71, p. 102825, Jan. 2022, doi: 10.1016/j.bspc.2021.102825.
- [8] Y.-D. Zhang, S. C. Satapathy, D. S. Guttery, J. M. Górriz, and S.-H. Wang, “Improved Breast Cancer Classification Through Combining Graph Convolutional Network and Convolutional Neural Network,” *Inf Process Manag*, vol. 58, no. 2, p. 102439, Mar. 2021, doi: 10.1016/j.ipm.2020.102439.
- [9] S. A. E. ALBAKIA and R. A. Saputra, “Identifikasi Jenis Daun Tanaman Obat Menggunakan Metode Convolutional Neural Network (CNN) Dengan Model VGG16,” *Jurnal Informatika Polinema*, vol. 9, no. 4, pp. 451–460, Aug. 2023, doi: 10.33795/jip.v9i4.1420.
- [10] J. Potsangbam and S. Shuleenda Devi, “Classification of Breast Cancer Histopathological Images Using Transfer Learning with DenseNet121,” *Procedia Comput Sci*, vol. 235, pp. 1990–1997, 2024, doi: 10.1016/j.procs.2024.04.188.
- [11] R. T. S and S. Joshi, “Breast Cancer Image Classification Using Convolutional Neural Networks,” *2024 International Conference on Integrated Intelligence and Communication Systems (ICIICS)*, pp. 1–6, 2024, [Online]. Available: <https://api.semanticscholar.org/CorpusID:276118181>
- [12] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, “Dataset of breast ultrasound images,” *Data Brief*, vol. 28, p. 104863, Feb. 2020, doi: 10.1016/j.dib.2019.104863.
- [13] H. Fathee and S. Sahmoud, “Iris segmentation in uncooperative and unconstrained environments: State-of-the-art, datasets and future research directions,” *Digit Signal Process*, vol. 118, p. 103244, Nov. 2021, doi: 10.1016/j.dsp.2021.103244.
- [14] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.
- [15] “ImageNet. <http://www.image-net.org>.”
- [16] W. Gómez-Flores and W. Coelho de Albuquerque Pereira, “A comparative study of pre-trained convolutional neural networks for semantic segmentation of breast tumors in ultrasound,” *Comput Biol Med*, vol. 126, p. 104036, Nov. 2020, doi: 10.1016/j.compbiomed.2020.104036.
- [17] A. Davila, J. Colan, and Y. Hasegawa, “Comparison of fine-tuning strategies for transfer learning in medical image classification,” *Image Vis Comput*, vol. 146, p. 105012, Jun. 2024, doi: 10.1016/j.imavis.2024.105012.