



Analisis Komparatif VGG19 pada Data Kanker Payudara Berbasis Augmentasi

Maie Istighosah^{1*}, Yudha Islami Sulistya²

¹Program Studi Teknik Informatika, Universitas Telkom, Indonesia

²Program Studi Rekayasa Perangkat Lunak, Universitas Telkom, Indonesia

*Penulis Korespondensi: maieistigh@telkomuniversity.ac.id¹

Abstract. *Class imbalance in breast cancer imaging often leads to models prioritizing the majority class, reducing sensitivity to actual cancer cases. This study evaluates data augmentation as a class balancing strategy for breast cancer classification using VGG19 with transfer learning. The model was trained and tested in two settings: before and after augmentation, to measure performance improvement. The results show a clear improvement after balancing, with accuracy rising from 94.63% to 97.59%, recall and specificity increasing from about 85.60% to 97.58%, and the F1 score rising from 0.8933 to 0.9759, indicating better balance between precision and recall. Interpretability analysis using Grad-CAM supports this improvement, with activations before augmentation being spread out and sometimes focusing on background artifacts, while the heatmap after augmentation concentrated on the lesion region, indicating that the network learned clinically meaningful features. Overall, the findings demonstrate that targeted augmentation effectively addresses class imbalance, enhances generalization, and improves lesion detection with VGG19. This approach enhances cancer sensitivity while reducing false alarms, supporting its potential for adoption in computer-aided diagnostic pipelines to provide more reliable breast cancer detection in clinical practice.*

Keywords: *Augmentation; Breast Cancer; Class Imbalance; Grad-CAM; VGG19.*

Abstrak. Ketidakseimbangan kelas dalam pencitraan kanker payudara seringkali menyebabkan model cenderung mengutamakan kelas mayoritas dan menurunkan sensitivitas terhadap kasus kanker yang sebenarnya. Studi ini mengevaluasi augmentasi data sebagai strategi penyeimbangan kelas untuk klasifikasi kanker payudara menggunakan VGG19 dengan transfer learning. Model dilatih dan diuji dalam dua pengaturan sebelum dan setelah augmentasi untuk mengukur peningkatan kinerja. Hasil menunjukkan perbaikan yang jelas setelah penyeimbangan, akurasi meningkat dari 94,63% menjadi 97,59%, recall dan spesifisitas meningkat dari sekitar 85,60% menjadi 97,58%, dan skor F1 meningkat dari 0,8933 menjadi 0,9759, menunjukkan keseimbangan yang lebih baik antara presisi dan recall. Analisis interpretabilitas menggunakan Grad-CAM mendukung peningkatan ini, aktivasi sebelum augmentasi tersebar dan terkadang terfokus pada artefak latar belakang, sedangkan heatmap setelah augmentasi terkonsentrasi pada wilayah lesi menunjukkan bahwa jaringan belajar fitur yang secara klinis bermakna. Secara keseluruhan, temuan ini menunjukkan bahwa augmentasi yang ditargetkan secara efektif mengatasi ketidakseimbangan kelas, meningkatkan generalisasi, dan memperbaiki deteksi lesi dengan VGG19. Pendekatan ini meningkatkan sensitivitas kanker sambil membatasi *false alarm*, mendukung potensinya untuk diadopsi dalam pipeline diagnosis *computer-aided* guna menyediakan deteksi kanker payudara yang lebih andal dalam praktik klinis.

Kata kunci: Augmentasi; Grad-CAM; Kanker Payudara; Ketidakseimbangan; VGG19.

1. LATAR BELAKANG

Kanker payudara menjadi isu kesehatan utama secara global terutama bagi perempuan, sehingga kebutuhan akan deteksi dini yang andal sangat dibutuhkan. Kemajuan *deep learning*, mendorong penggunaan Convolutional Neural Network (CNN) untuk meningkatkan analisis citra medis sehingga dapat mendukung radiolog dalam pengambilan keputusan. Berbagai pendekatan seperti *transfer learning*, fusi radiomik (Maruf dkk., 2025), hingga integrasi data klinis telah terbukti dapat di implementasikan pada data mamografi, USG, dan histopatologi.

Meskipun demikian, karakteristik data tetap menantang, seperti variasi intensitas, kontras, artefak pencitraan serta heterogenitas biologis yang menuntut kurasi dan prapemrosesan yang cermat.

Tantangan lainnya yang paling krusial dalam praktik di dunia nyata ialah ketidakseimbangan kelas hal ini disebabkan jumlah citra kanker jauh lebih sedikit dibandingkan citra non kanker. Kondisi ini menyebabkan model cenderung mempelajari pola kelas mayoritas dan mengabaikan kelas minoritas, sehingga menurunkan sensitivitas terhadap kasus kanker. Sejumlah penelitian menegaskan bahwa pentingnya augmentasi data untuk menekan *overfitting*, menambah variasi sampel minoritas, dan meningkatkan generalisasi, termasuk opsi lanjutan berbasis generatif ketika transformasi sederhana tidak memadai (Ahirwar & Agrawal, 2023; Alshamrani, 2025; Blahová dkk., 2025; Dhomane & Shinde, 2023). Disisi lain, arsitektur VGG terutama VGG19 tetap relevan sebagai *feature extractor* yang baik pada skema *transfer learning*, walaupun masih diperlukan kontrol regularisasi dan strategi penyeimbangan yang tepat pada data kecil dan tidak seimbang (Abdikenov dkk., 2025; Maruf dkk., 2025; Wang dkk., 2024). Dengan demikian, evaluasi komparatif yang ketat atas dampak augmentasi terhadap performa arsitektur pada dataset menjadi kebutuhan metodologis.

Sejalan dengan itu, berbagai studi mencantumkan bahwa metode augmentasi seperti rotasi, flip, zoom, penyesuaian kecerahan (*contrast*), penambahan *noise* ringan pada citra dapat meningkatkan kemampuan model mendeteksi kanker dan membuat model lebih tahan ketika kondisi gambar berubah-ubah (*robustness*). Penerapan transfer learning membantu model cepat konvergen saat datanya sedikit. Namun, kesalahan *false negative* atau kasus kanker yang luput terdeteksi masih sering terjadi, terutama jika lesinya kecil atau kontrasnya rendah. Pada citra non-mammografi, gabungan fitur radiomik dan arsitektur *deep learning* lainnya bisa mencapai akurasi sangat tinggi, tetapi dengan konsekuensi kompleksitas lebih besar dan kebutuhan data yang meningkat. Oleh karena itu, dibutuhkan protokol yang terarah dan efisien cukup sederhana untuk diterapkan, namun efektif dalam menaikkan kemampuan deteksi kelas minoritas.

Urgensi penelitian bersifat klinis dan metodologis. Secara klinis, *false negative* pada mammografi dengan distribusi kelas timpang perlu di turunkan. Secara metodologis, diperlukan protokol augmentasi spesifik mammografi yang menjaga validitas anatomi serta meningkatkan *recall* tanpa menaikkan *false positive* secara berlebihan. Berdasarkan celah tersebut, studi ini mengevaluasi dampak penyeimbangan kelas berbasis augmentasi terhadap performa VGG19 pada mammografi, melalui perbandingan dua skenario sebelum dan sesudah augmentasi dengan metrik supaya peka terhadap ketidakseimbangan serta analisis Grad-CAM untuk memastikan validitas klinis perhatian model.

2. KAJIAN TEORITIS

Kemajuan *deep learning* untuk deteksi kanker payudara berkembang pesat di berbagai jenis citra medis dan situasi klinis. *Transfer learning* yang digabung dengan radiomik (fitur terukur dari citra) terbukti membantu model mengenali pola penting untuk klasifikasi maupun deteksi lesi, terutama ketika informasi klinis juga ikut digabungkan (*multimodal feature fusion*) (Maruf dkk., 2025). Pada mamografi, banyak laporan menunjukkan kenaikan kinerja deteksi lesi setelah model pra-latih di *fine tune* secara tepat (Abdikenov dkk., 2025). Di ultrasonografi (USG), kolaborasi antara *deep learning*, fitur radiomik, dan data klinis mampu meningkatkan prediksi biomarker seperti HER-2 (Hu dkk., 2025). Dari sisi perangkat, rancangan CNN multitugas untuk pencitraan *near-infrared* menjanjikan pemrosesan *real-time* yang dapat mendukung alur kerja klinis (Momtahan & Golnaraghi, 2025). Sejumlah studi juga menekankan pentingnya *pipeline* yang rapi mulai dari pra-proses, ekstraksi fitur, hingga pemodelan karena keteraturan alur ini konsisten terkait dengan peningkatan akurasi dan efisiensi (Conte dkk., 2025; Desai & Mahto, 2025; Fan dkk., 2025; Rai dkk., 2025; Sappa & Lingam, 2025). Secara umum, praktik *transfer learning* lintas domain (dari data nonmedis ke medis) tetap relevan untuk mengatasi keterbatasan data beranotasi di bidang kesehatan (Wang et al., 2024).

Salah satu tantangan utama dalam praktik nyata adalah ketidakseimbangan kelas, di mana jumlah citra kanker umumnya jauh lebih sedikit daripada non-kanker. Kondisi ini memicu bias pembelajaran ke kelas mayoritas, menurunkan sensitivitas pada kasus minoritas (kanker). Tinjauan dan studi eksperimental menegaskan pentingnya augmentasi data untuk memperkaya variasi sampel minoritas, menekan *overfitting*, dan meningkatkan generalisasi ketika transformasi dasar tidak memadai (Ahirwar & Agrawal, 2023; Alshamrani, 2025; Blahová dkk., 2025; Dhomane & Shinde, 2023). Pendekatan-pendekatan ini sejalan dengan praktik penyeimbangan kelas modern seperti *oversampling* atau augmentasi terarah, yang dimaksudkan menjaga distribusi data latih tanpa mengorbankan validitas label.

Di luar MRI misalnya mamografi, USG, dan histopatologi, literatur menunjukkan bahwa strategi augmentasi dan pilihan arsitektur yang tepat meningkatkan kinerja model. Penelitian di bidang histologi dan skenario tak berlabel menunjukkan bahwa teknik segmentasi dan deteksi berbasis *active contours*, *blob detection*, serta *pipeline* tanpa pengawasan mampu menyoroti area mencurigakan secara efektif (Becirovic dkk., 2024; Majanga dkk., 2025b, 2025a). Pada *ultrasound* payudara, penggabungan augmentasi dan arsitektur *deep learning* meningkatkan keberhasilan prediksi secara keseluruhan, menunjukkan manfaat langsung dari penambahan keragaman data (Aygün & Kaya, 2024). Di sisi lain, validasi metrik dan cara evaluasi

menekankan pentingnya pelaporan metrik yang peka terhadap ketidakseimbangan (misalnya PR-AUC) serta uji signifikansi berpasangan untuk memastikan perbedaan kinerja benar-benar bermakna secara statistik (Alekseev dkk., 2025; Cheruvathoor & Babu, 2025; Długosz-Pokorska dkk., 2025). Bukti dari berbagai modalitas ini menunjukkan bahwa prinsip penyeimbangan kelas dan evaluasi yang tepat bersifat umum dan dapat diterapkan, termasuk dalam domain mammografi payudara.

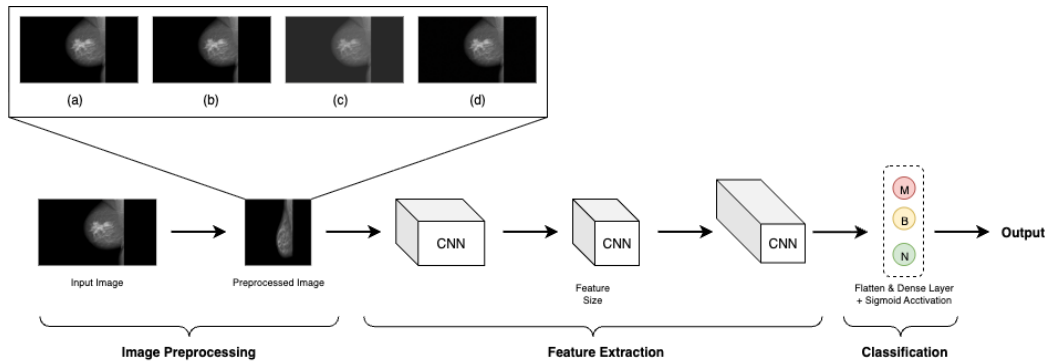
Pada tingkat desain arsitektur CNN dan *transfer learning*, arsitektur VGG, terutama VGG19 masih digunakan sebagai pengambil fitur yang efektif. Kelebihannya berasal dari struktur bertingkat yang dalam dengan ukuran kernel 3×3 yang sederhana serta teknik *fine-tuning* yang telah dikembangkan dengan baik. Dalam pengolahan citra histopatologi payudara, pendekatan berbasis VGG dengan *transfer learning* bisa memberikan hasil akurasi dan nilai AUC yang relatif baik. Akan tetapi, masih sering terjadi kesalahan negatif palsu pada lesi yang ukurannya kecil atau memiliki kontras rendah, sehingga diperlukan metode penyeimbangan kelas serta teknik regularisasi yang lebih ketat (Jain & Singh, 2024; Kaushik & Choudhary, 2024; Kaushik & Sharma, 2024). Di sisi lain, beberapa studi membandingkan bahwa arsitektur seperti ResNet dan jenisnya lainnya mampu meningkatkan akurasi lebih tinggi, tetapi biasanya memerlukan volume data yang lebih besar serta manajemen kompleksitas yang lebih rumit (Wang dkk., 2024). Dengan mempertimbangkan hal tersebut, VGG19 sering dipilih ketika dibutuhkan keseimbangan antara performa, stabilitas proses *fine-tuning*, dan kemudahan dalam penerapan.

Dengan demikian, masih ada kesenjangan yang memerlukan evaluasi secara komparatif yang ketat antara pelatihan menggunakan data tidak seimbang dan data yang sudah diseimbangkan melalui augmentasi pada citra mammografi payudara, dengan fokus pada metrik seperti akurasi, recall, presisi, F1-Score dan spesifitas kelas kanker serta interpretabilitas menggunakan *Grad-CAM*, supaya dapat memastikan validitas klinis (Abdikenov dkk., 2025; Alekseev dkk., 2025; Blahová dkk., 2025; Cheruvathoor & Babu, 2025; Długosz-Pokorska dkk., 2025; Jain & Singh, 2024; Kaushik & Choudhary, 2024).

3. METODE PENELITIAN

Penelitian yang dilakukan merupakan studi eksperimental yang bertujuan menganalisis performa model VGG19 untuk deteksi kanker payudara dengan membandingkan data tidak seimbang versus seimbang. Serangkaian pengujian dilakukan dengan menerapkan augmentasi data. Dalam penelitian ini digunakan arsitektur VGG19, model tersebut diuji dan dibandingkan guna menilai sejauh mana teknik-teknik tersebut meningkatkan akurasi dan keandalan dalam

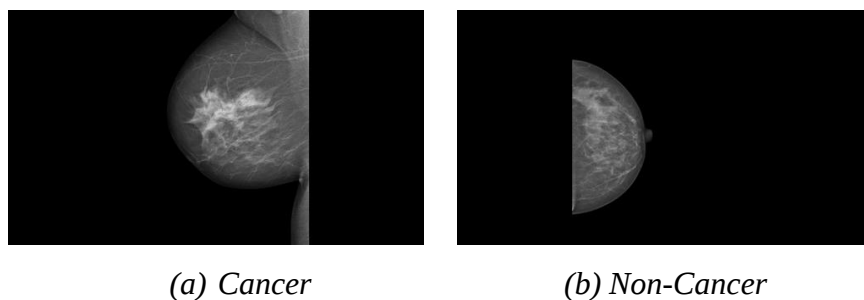
diagnosis kanker payudara. Performa model dinyatakan dalam metrik terhitung, seperti akurasi, presisi, recall, dan F1-score. Pendekatan ini memungkinkan evaluasi yang objektif sehingga hasil yang diperoleh dapat diandalkan serta memberikan informasi yang relevan bagi pengembangan sistem diagnosis kanker payudara. Ilustrasi model berbasis CNN untuk diagnosis kanker payudara dapat dilihat pada gambar 1 berikut.



Gambar 1. Ilustrasi Model Berbasis CNN untuk Diagnosis Kanker Payudara.

Dataset

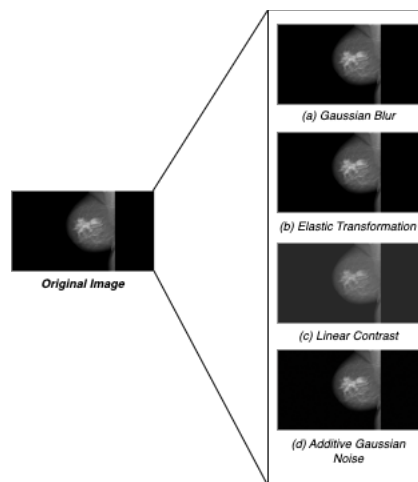
Untuk mengembangkan metode yang andal untuk identifikasi kanker payudara, dataset yang terannotasi dengan baik, merupakan kebutuhan yang krusial. Dataset yang digunakan pada penelitian ini berupa citra mammografi kanker payudara (Aqdar dkk., 2024) yang dikumpulkan dari pasien yang mencakup dua kategori utama yaitu terdiagnosis kanker payudara dan tanpa kondisi tersebut. Seluruh data yang ada telah dikurasi, diverifikasi, dan diklasifikasikan secara cermat oleh tenaga medis ahli, sehingga menjamin keandalan serta ketepatan untuk kepentingan penelitian. Dataset ini berasal dari wilayah Sulaymaniyah, Irak yang memberikan perspektif unik mengenai prevalensi dan karakteristik kanker payudara di daerah tersebut. Dataset terdiri dari 745 citra, dengan 125 data kanker, dan 620 data non kanker. Serta terdapat data yang telah di augmentasi sejumlah 9685 data. Dataset ini kemudian dibagi ke data training, testing, dan validasi dengan porsi 80:10:10 guna mencegah kebocoran data. Pada citra dilakukan resize menjadi 224x224. Dataset kanker payudara dapat dilihat pada Gambar 2.



Gambar 2. Dataset Citra Mammografi (a) Cancer dan (b) Non-Cancer.

Data Augmentation

Citra mamografi kanker payudara perlu melalui preprocessing untuk mengatasi masalah seperti kontras rendah, noise, pencahayaan tidak seragam, serta artefak akuisisi seperti pada label, marker, dan otot pektoralis. Faktor ini menurunkan kinerja klasifikasi maupun deteksi, sehingga langkah pra-pemrosesan yang terarah menjadi penting. Pada dataset Mammografi yang digunakan telah dilakukan augmentasi, namun teknik augmentasi yang dilakukan terlalu banyak seperti *crop*, *additive gaussian noise*, *dropout*, *elastic transformation*, *flip*, *gaussian blur*, *linear contrast*, *multiply*, *perspective transform*, *piecewise affine*, *resize*, dan *rotate*. Namun teknik augmentasi yang dipilih dalam penelitian ini yaitu *gaussian blur*, *elastic transformation*, *linear contrast*, dan *additive gaussian noise*. Keempat teknik pada Gambar 3 tersebut dianggap konservatif karena hanya mengenalkan perubahan minor yang masih dalam ruang lingkup variabilitas citra medis asli. Teknik tersebut membantu meningkatkan *robustness* model tanpa menciptakan pola lesi yang benar benar baru atau ada artefak yang aneh. Dengan demikian, deteksi lesi tidak dirugikan, justru model menjadi lebih andal terhadap variasi realistis dan validitas klinis citra hasil augmentasi tetap tinggi. Seperti disimpulkan dalam tinjauan (Oza dkk., 2022), augmentasi berbasis *noise injection*, *blur*, dan filter intensitas dapat menghasilkan dataset yang lebih *robust*, asalkan diterapkan secara hati-hati agar tidak menghilangkan informasi kritis.



Gambar 3. Augmentasi Data dengan Teknik (a) *Gaussian Blur*, (b) *Elastic Transformation*, (c) *Linear Contrast*, (d) *Additive Gaussian Noise*.

Berbeda dengan teknik yang dipilih, beberapa metode augmentasi umum sengaja tidak digunakan dalam penelitian ini karena berisiko memotong atau mengubah bentuk lesi, membalik *laterality*, atau menghasilkan sampel di luar distribusi klinis.

Feature Extraction

Membangun model dari awal bukan pilihan yang paling praktis ketika data terbatas, karena biaya komputasi tinggi dan resiko konvergensi lambat. Melatih jaringan besar pada korpus kecil juga rawan *overfitting* serta membutuhkan waktu yang lama. Untuk mengatasinya, digunakan *transfer learning* dimana bobot pra-latih dimanfaatkan sebagai titik awal, lalu diadaptasi oleh *fine tuning* ke tugas target. Pendekatan ini mempercepat pelatihan, meningkatkan stabilitas optimisasi, dan umumnya menghasilkan kinerja lebih baik pada dataset kecil.

Dalam konteks deteksi kanker payudara berbasis citra, VGG19 sering digunakan sebagai *feature extractor* yang kuat. Arsitektur VGG19 mulanya akan membekukan lapisan awal hingga menengah agar fitur umum tepi dan tekstur tetap terjaga, mengganti *head* klasifikasi dengan lapisan padat untuk tugas biner, lalu melakukan *fine tuning* bertahap dengan *learning rate* rendah pada blok atas untuk menangkap pola spesifik domain. *Fine tuning* sendiri merupakan proses mengadaptasi model CNN pra-latih ke tugas baru dengan menyesuaikan sebagian parameternya. Untuk menjaga fitur yang kuat yang telah dipelajari tidak rusak. Prosedur ini umumnya dilakukan secara multi langkah, pada unit *fully connected* akhir diganti dengan lapisan baru yang diinisialisasi ulang, lapisan konvolusional awal dibekukan, dan hanya lapisan atas yang dilatih lebih dulu sebelum secara selektif membuka (*unfreeze*) lapisan tambahan. Dalam klasifikasi citra, pendekatan yang dilakukan, *transfer learning* pada arsitektur VGG dapat mempercepat pelatihan, meningkatkan stabilitas optimisasi, dan menaikkan akurasi pada dataset terbatas. Efektivitasnya didukung oleh *preprocessing* yang konsisten, augmentasi terarah untuk menyeimbangkan kelas serta teknik regularisasi seperti *dropout*, *early stopping*, dan *reduce-on-plateau*, serta menggunakan optimizer Adam. Dengan kombinasi tersebut, pengembangan model menjadi lebih cepat dan andal, sekaligus menekan kebutuhan komputasi dan ukuran data dibanding melatih jaringan mendalam dari awal.

Confusion Matrix

Confusion Matrix merupakan pendekatan yang memberikan pernyataan secara singkat hasil prediksi pada masalah klasifikasi biner maupun multikelas. Matriks ini menampilkan jumlah prediksi benar dan prediksi salah untuk setiap kelas, misalnya *True Positive*, *False Positive*, *True Negative*, dan *False Negative*. Hal ini berguna untuk mengevaluasi sifat diskriminatif dan kinerja model. Berbasis perhitungan statistik sederhana, adanya *confusion matrix* memudahkan interpretasi melalui visualisasi, serta menjadi dasar perhitungan matrik turunan seperti akurasi, presisi, *recall*, dan F1-Score. Dengan demikian, *confusion matrix* membantu mengidentifikasi pola kesalahan model dan area yang perlu ditingkatkan.

a. **Akurasi**

Akurasi menunjukkan seberapa akurat model prediktif yang dibuat.

$$Akurasi = \frac{TP + TN}{(TP + TN + FP + FN)} \times 100\%$$

b. **Presisi**

Presisi merupakan persentase dari total kasus yang diprediksi. Sederhananya, Presisi menunjukkan persentase prediksi model benar, jika dikatakan benar.

$$Presisi = \frac{TP}{(TP + FP)} \times 100\%$$

c. **Recall**

Recall memberitahukan seberapa sering model memprediksi benar. Recall dikenal juga sebagai sensitivity. Recall menunjukkan presentase kasus positif dari seluruh kasus aktual dalam dataset.

$$Recall = \frac{TP}{(TP + FN)} \times 100\%$$

d. **F1-Score**

F1-Score digunakan untuk menilai model. F1-Score merupakan rata-rata dari presisi dan recall. Semakin tinggi F1-Score, semakin baik model tersebut. F1-Score memberikan bobot sama pada presisi dan recall, hal ini menunjukkan keseimbangan diantara keduanya.

$$F1 - Score = 2 \frac{recall \times presisi}{(recall + presisi)} \times 100\%$$

Dalam konteks klasifikasi *cancer* dan *non-cancer*, *True Positive* (TP) terjadi ketika kelas sebenarnya *cancer* dan model memprediksi *cancer*; *False Positive* (FP) ketika kelas sebenarnya *non-cancer* tetapi model memprediksi *cancer*; *True Negative* (TN) ketika kelas sebenarnya *non-cancer* dan model memprediksi *non-cancer*; serta *False Negative* (FN) ketika kelas sebenarnya *cancer* namun model memprediksi *non-cancer*.

Grad-CAM

Pada penelitian ini Grad-CAM digunakan untuk menjelaskan Keputusan model pada data medis kanker payudara dengan memetakan kontribusi fitur dari lapisan konvolusi terakhir ke ruang citra, menghasilkan heatmap kemudian di overlay pada mammogram. Dengan perlakuan ini, dapat dilihat apakah model memperhatikan area massa yang relevan saat memberi prediksi kanker. Grad-CAM seringkali digunakan pada kasus TP/FP/FN untuk analisis kesalahan. FP sering menunjukkan aktivasi pada jaringan tepi, sedangkan FN mengungkap lesi sangat kecil

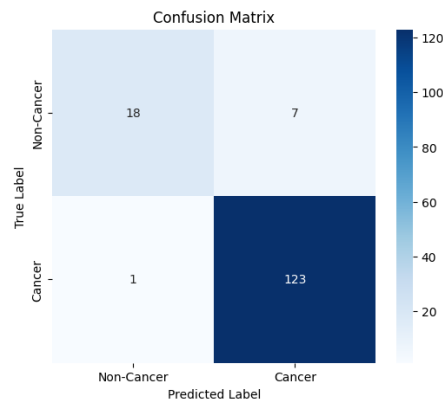
atau kontras yang rendah yang kurang tertangkap. Hal ini membantu memperbaiki pipeline dari *preprocessing*, skema augmentasi dan penyesuaian *threshold*, dan mempermudah komunikasi dengan klinisi, karena alasan prediksi dapat ditelusuri ke lokasi anatomi yang jelas.

4. HASIL DAN PEMBAHASAN

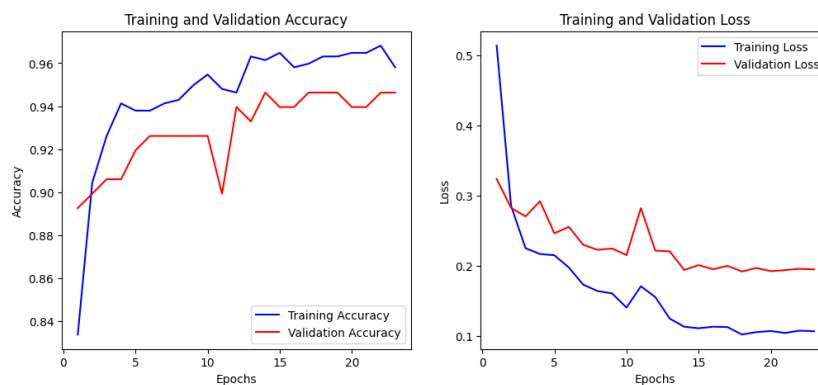
Pada bagian ini menyajikan temuan kuantitatif dan kualitatif dari dua skenario sebelum dan sesudah augmentasi. Dinamika pembelajaran diringkas melalui kurva akurasi dan loss, confusion matrix menunjukkan distribusi kesalahan per kelas; dan Grad-CAM memperlihatkan area citra yang mendorong keputusan model. Seluruh metrik kuantitatif akurasi, presisi, recall, F1-Score, dan spesifitas yang juga menjadi dasar perbandingan antar skenario. Sehingga bagian ini akan menyajikan temuan utama, analisis kritis atas perilaku model, serta implikasi metodologis dan klinis dari penerapan augmentasi pada klasifikasi citra kanker payudara.

Pada skenario data tanpa augmentasi, model mencapai akurasi 94,63% meskipun akurasi tampak tinggi analisis mendalam terhadap metrik lain mengungkap adanya bias performa antar kelas akibat ketimpangan data. Nilai presisi sebesar 94,68% menunjukkan proporsi prediksi positif yang benar cukup tinggi, namun recall dan spesifitasnya masing-masing hanya mencapai 85,60%. Recall yang relatif rendah ini menandakan model gagal menangkap banyak kasus pada kelas minoritas. Hal tersebut konsisten dengan fenomena pada data tidak seimbang dimana model cenderung mempelajari pola kelas mayoritas dan mengabaikan kelas minoritas, sehingga menyebabkan bias.

Konsekuensi dari bias ini dapat dilihat confusion matrix pada Gambar 4, model hanya berhasil mengenali 18 sampel kelas minoritas dengan benar, sementara 7 sampel minoritas lainnya salah diklasifikasikan hal ini ditunjukkan dengan FN yang tinggi. Sebaliknya, untuk kelas mayoritas model mendeteksi hampir seluruhnya dengan TP yaitu 123, dengan sedikit kesalahan prediksi sebagai positif FP yaitu 7. Hal ini menunjukkan sensitivitas model terhadap kelas mayoritas mendekati 99%, sedangkan sensitivitas terhadap kelas minoritas hanya sekitar 72%. Perbedaan drastis ini mengakibatkan balanced accuracy turun hingga 85%, mencerminkan performa yang timpang. Secara praktis, model tanpa augmentasi sering luput mendeteksi kasus kanker (minoritas) karena dominasi pola kelas non-kanker dalam pelatihan. Temuan ini sejalan bahwa ketidakseimbangan kelas menurunkan kemampuan deteksi lesi pada kasus minor.

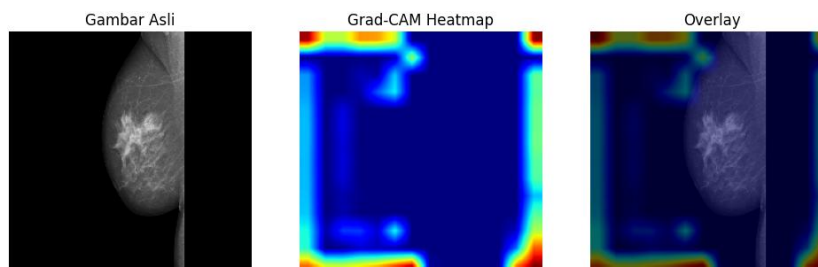


Gambar 4. *Confusion Matrix* pra-augmentasi.



Gambar 5. Grafik *Training* dan *Validation* pra-augmentasi.

Dilihat dari kurva pembelajaran pada Gambar 5, tampak gap signifikan antara loss dan akurasi pada data training vs. validation. Kurva training loss turun lebih cepat sementara validation loss stagnan lebih tinggi, yang mengindikasikan potensi overfitting. Model cenderung menghafal data latih mayoritas tanpa mampu generalisasi terhadap data validasi yang mengandung proporsi kelas minoritas lebih besar. Pola ini lumrah terjadi ketika dataset tidak diimbangi, karena model memperoleh variansi data yang terbatas untuk kelas minoritas. Sebagai akibatnya, stabilitas dan keandalan prediksi model pada data baru menurun.

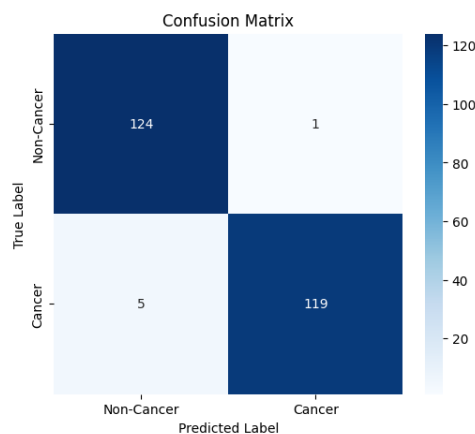


Gambar 6. Visualisasi Grad-CAM pra-augmentasi.

Pada Gambar 6 yang di uji cobakan menggunakan citra non kanker menunjukkan kondisi di atas turut berdampak pada interpretabilitas model. Visualisasi Grad-CAM pra-augmentasi menunjukkan fokus aktivasi model yang menyebar dan kurang spesifik pada area lesi. Alih-alih

menyoroti tumor atau kelainan pada citra mammografi, heatmap Grad-CAM tersebar di area background atau jaringan normal. Pola aktivasi yang tidak terarah ini mengindikasikan bahwa model mungkin memanfaatkan fitur non-diagnostik atau artefak pencitraan dalam mengambil keputusan. Dengan kata lain, tanpa penyeimbangan data, model kesulitan mengenali pola radiologis spesifik kanker dan cenderung tertarik pada karakteristik umum yang tidak terkait langsung dengan lesi. Hal ini tentu kurang ideal dari sudut pandang klinis, karena model yang tidak melihat area lesi akan sulit dipercaya sebagai sistem bantu diagnosa.

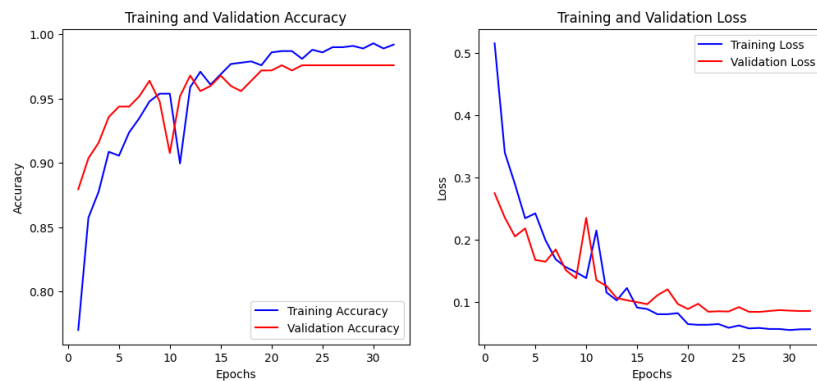
Sedangkan pada data setelah dilakukan augmentasi data untuk menyeimbangkan kelas, kinerja model mengalami perbaikan menyeluruh. Model VGG19 yang dilatih dengan data seimbang mencapai akurasi 97,59%, meningkat dibanding sebelumnya. Lebih penting lagi, model kini menunjukkan performa yang konsisten tinggi pada kedua kelas. Baik precision (97,65%) maupun recall meningkat, menghasilkan F1-score sebesar 0,9759. Peningkatan paling signifikan tampak pada recall (sensitivitas) dan spesifisitas model yang mencapai 97,58% jauh lebih tinggi dibanding 85,60% sebelum augmentasi. Ini berarti model pasca-augmentasi mampu mendeteksi hampir semua kasus kanker sekaligus mengenali kasus non-kanker dengan sangat baik. Peningkatan recall dan spesifisitas secara simultan menandakan tercapainya keseimbangan sensitivitas-spesifisitas yang optimal, alias berkurangnya bias terhadap salah satu kelas.



Gambar 7. *Confusion Matrix* setelah Augmentasi.

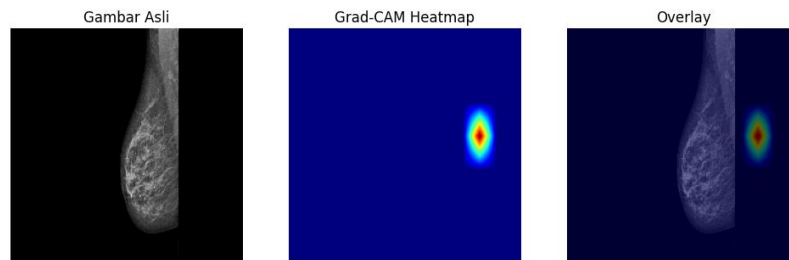
Pada Gambar 7 *confusion matrix* memperjelas dampak augmentasi dalam menekan kesalahan. Model hasil augmentasi hanya menghasilkan 1 *false positive* (salah deteksi non-kanker menjadi kanker), turun drastis dari 7 kasus pada model awal. Penurunan FP ini mencerminkan lonjakan spesifisitas, artinya sistem jarang memberi alarm palsu pada sampel non-kanker. Dari sisi lain, *false negative* sedikit meningkat (5 kasus kanker luput terdeteksi, dibanding 1 kasus sebelumnya). Namun demikian, jumlah FN yang masih rendah ini bersifat

kompromi minimal jika dibanding manfaat augmentasi dalam meningkatkan deteksi kelas minoritas. Pasca augmentasi, model berhasil mengenali 119 dari 124 kasus kanker (sensitivitas 96%) dan 124 dari 125 kasus non-kanker (spesifisitas 99,2%). Dengan kata lain, augmentasi membuat performa model lebih seimbang: kedua kelas memiliki tingkat deteksi mendekati sempurna. Balanced accuracy model meningkat hingga 97,58%, mendekati akurasi keseluruhan. Temuan ini sejalan dengan studi sebelumnya bahwa menambah variasi data kelas minoritas melalui augmentasi efektif meningkatkan generalisasi dan recall model. Hasil empiris ini membuktikan bahwa strategi augmentasi terarah berhasil mengurangi bias model dan meningkatkan *confidence* prediksi pada kedua kelas.



Gambar 8. Grafik *Training* dan *Validation* setelah Augmentasi.

Dari perspektif proses pelatihan, augmentasi data berperan sebagai regularisasi yang kuat. Grafik akurasi atau *loss training* dan *validation* setelah augmentasi pada Gambar 8 menunjukkan pola yang lebih stabil dan konvergen. Tidak ada lagi gap mencolok, kurva *loss validasi* menurun seiring *loss training*, menandakan *overfitting* berhasil ditekan. Tambahan variasi sampel dari teknik augmentasi membantu model menggeneralisasi pola alih-alih menghafal detail spesifik dataset. Hal ini selaras dengan literatur yang menyatakan augmentasi mampu menambah keragaman sampel minoritas, mengurangi *overfitting*, dan meningkatkan ketangguhan model terhadap perubahan kondisi citra. Dengan demikian, VGG19 pasca augmentasi tidak hanya unggul pada metrik akurasi, tetapi juga lebih *robust* terhadap variabilitas data nyata.



Gambar 9. Visualisasi Grad-CAM setelah Augmentasi.

Peningkatan kinerja kuantitatif tersebut diperkuat oleh peningkatan kualitas interpretabilitas model. Visualisasi Grad-CAM setelah augmentasi pada Gambar 4.6 menunjukkan fokus aktivasi yang jelas pada area lesi kanker, yang di uji cobakan pada citra non kanker. Berbeda dengan model awal, model teraugmentasi menyoroti wilayah tumor atau massa mencurigakan pada mammogram secara konsisten. Contohnya, bila terdapat nodul pada kuadran tertentu, heatmap Grad-CAM pasca augmentasi akan memusat di sekitar nodul tersebut, menandakan jaringan convolutional menangkap fitur-fitur penting yang relevan secara klinis. Peningkatan fokus spasial ini menunjukkan model belajar perhatian yang lebih tepat. Hal ini sangat krusial untuk aplikasi medis, karena model yang benar-benar "melihat" lesi cenderung mengambil keputusan berdasarkan karakteristik penyakit, bukan noise atau artefak. Beberapa penelitian juga menekankan pentingnya evaluasi interpretabilitas dengan Grad-CAM untuk memastikan validitas klinis model deep learning. Dalam studi ini, hasil Grad-CAM mengonfirmasi bahwa augmentasi data membantu VGG19 memahami fitur radiologis kanker dengan lebih akurat. Secara tidak langsung, model yang terlatih dengan data seimbang ini menjadi lebih terpercaya untuk digunakan sebagai alat bantu diagnosis karena keputusan yang diambil dapat ditelusuri ke area patologis sebenarnya.

Secara keseluruhan, Hasil dan Pembahasan ini menunjukkan bahwa augmentasi data efektif meningkatkan performa deteksi kanker payudara pada VGG19. Setiap temuan mulai dari metrik akurasi hingga visualisasi Grad-CAM mengarah pada kesimpulan yang konsisten: penyeimbangan data melalui augmentasi berhasil mengatasi kendala ketidakseimbangan kelas, meningkatkan recall kelas minoritas dan spesifisitas kelas mayoritas, menekan *overfitting*, serta mengarahkan model untuk belajar fitur yang lebih bermakna. Dampak positif ini mendukung implementasi metode augmentasi sebagai komponen standar dalam pengembangan model klasifikasi citra medis, terutama ketika data yang tersedia terbatas dan tidak seimbang. Dengan *model yang lebih akurat dan interpretable, diharapkan sistem deteksi computer-aided* untuk kanker payudara dapat memberikan kinerja yang andal dan dapat diandalkan dalam skenario klinis nyata.

5. KESIMPULAN DAN SARAN

Penyeimbangan data melalui augmentasi konservatif pada mamografi terbukti meningkatkan sensitivitas (*recall*) sekaligus menjaga/menaikkan spesifisitas, dengan kenaikan F1 dan akurasi yang konsisten. Kurva validasi yang stabil dan Grad-CAM yang terfokus pada lesi memperkuat validitas klinis model VGG19. Penelitian ini membuktikan secara empiris bahwa augmentasi data berperan krusial dalam meningkatkan kinerja model VGG19 untuk klasifikasi kanker payudara pada dataset yang tidak seimbang. Melalui strategi augmentasi yang tepat, model berhasil dihindarkan dari bias terhadap kelas mayoritas, sehingga sensitivitas (*recall*) terhadap kasus kanker meningkat signifikan disertai spesifisitas tinggi terhadap kasus non-kanker. Hasil perbandingan menunjukkan model dengan data teraugmentasi mencapai keseimbangan performa yang jauh lebih baik dibanding model tanpa augmentasi. Peningkatan *balanced accuracy* dari 85% menjadi 97% setelah augmentasi mencerminkan penurunan kesalahan klasifikasi pada kedua kelas. Selain itu, augmentasi terbukti efektif menekan *overfitting* terbukti dari kurva validasi yang stabil dan mendorong model mempelajari fitur relevan, yang ditunjukkan melalui fokus Grad-CAM pada area lesi sebenarnya.

Secara keseluruhan, augmentasi data memberikan manfaat nyata dalam sistem deteksi kanker payudara berbasis CNN. Dengan meningkatkan *recall* tanpa mengorbankan *precision* secara berlebihan, teknik ini membantu model mencapai kinerja diagnostik yang lebih andal. Potensi penerapannya di sistem diagnosis klinis sangat menjanjikan. Model VGG19 terlatih dengan data seimbang dapat diintegrasikan ke dalam *Computer-Aided Diagnosis* (CAD) untuk membantu radiolog mengidentifikasi lesi pada mamografi secara lebih akurat. Tingginya sensitivitas berarti lebih banyak kasus kanker dapat terdeteksi sejak dini, sementara spesifisitas yang tinggi mengurangi alarm palsu sehingga menekan prosedur biopsi yang tidak perlu. Implikasinya, *workflow* klinis dapat menjadi lebih efisien sehingga radiolog dapat memanfaatkan model ini sebagai alat bantu keputusan kedua yang memberikan indikasi lesi dengan tingkat kepercayaan tinggi. Tentunya, sebelum implementasi luas, diperlukan validasi lebih lanjut di lingkungan klinis sesungguhnya. Meski demikian, temuan penelitian ini telah menegaskan bahwa strategi penyeimbangan data berbasis augmentasi adalah langkah penting untuk meningkatkan kinerja model AI dalam diagnosis kanker payudara, sehingga membawa kita selangkah lebih maju menuju sistem pendukung keputusan klinis yang akurat, andal, dan interpretatif.

DAFTAR REFERENSI

- Abdikenov, B., Rakishev, D., Orazayev, Y., & Zhaksylyk, T. (2025). Enhancing breast lesion detection in mammograms via transfer learning. *Journal of Imaging*, 11(9), 314. <https://doi.org/10.3390/jimaging11090314>
- Ahirwar, M., & Agrawal, A. (2023). Performance analysis of deep learning models over BreakHis dataset using up-sampling and down-sampling techniques for classification of breast cancer. In *Proceedings of the 9th International Conference on Smart Computing and Communications (ICSCC 2023)* (pp. 594–599). <https://doi.org/10.1109/ICSCC59169.2023.10334935>
- Alekseev, A., Shcherbakov, V., Avdieiev, O., Denisov, S. A., Kubytskyi, V., Blinchevsky, B., Murokh, S., Ajeer, A., Adams, L., Greenwood, C., Rogers, K., Jones, L. J., Mourokh, L., & Lazarev, P. (2025). Benign/cancer diagnostics based on X-ray diffraction: Comparison of data analytics approaches. *Cancers*, 17(10). <https://doi.org/10.3390/cancers17101662>
- Alshamrani, S. S. (2025). Machine learning techniques improving the Box–Cox transformation in breast cancer prediction. *Electronics*, 14(16). <https://doi.org/10.3390/electronics14163173>
- Aqdar, K. B., Abdalla, P. A., Mustafa, R. K., Abdulqadir, Z. H., Qadir, A. M., Shali, A. A., & Aziz, N. M. (2024). *Mammogram mastery: A robust dataset for breast cancer detection and medical education* [Dataset]. (Sumber/platform belum dicantumkan)
- Aygün, E. N., & Kaya, M. (2024). Medical image segmentation with U-Net for breast cancer and lump type prediction. In *Proceedings of the 2024 International Conference on Decision Aid Sciences and Applications (DASA 2024)*. <https://doi.org/10.1109/DASA63652.2024.10836584>
- Becirovic, M., Kurtovic, A., Pozderac, D., & Omanovic, S. (2024). Ultrasound breast cancer image classification with GAN-based synthetic data augmentation. In *Proceedings of the 2024 32nd Telecommunications Forum (TELFOR 2024)*. <https://doi.org/10.1109/TELFOR63250.2024.10819074>
- Blahová, L., Kostolný, J., & Cimrák, I. (2025). Neural network-based mammography analysis: Augmentation techniques for enhanced cancer diagnosis—A review. *Bioengineering*, 12(3). <https://doi.org/10.3390/bioengineering12030232>
- Cheruvathoor, J. M., & Babu, N. V. (2025). Advanced ensemble learning and feature enhancement for robust breast cancer classification in histopathological images. In *Proceedings of the 2025 International Conference on Data Science, Agents and Artificial Intelligence (ICDSAAI 2025)*. <https://doi.org/10.1109/ICDSAAI65575.2025.11011593>
- Conte, L., Rizzo, R., Sallustio, A., Maggiulli, E., Capodici, M., Tramacere, F., Castelluccia, A., Raso, G., De Giorgi, U., Massafra, R., Portaluri, M., Cascio, D., & De Nunzio, G. (2025). Radiomics and machine learning approaches for the preoperative classification of in situ vs. invasive breast cancer using dynamic contrast-enhanced magnetic resonance imaging (DCE–MRI). *Applied Sciences*, 15(14). <https://doi.org/10.3390/app15147999>

- Desai, A., & Mahto, R. (2025). Multi-class classification of breast cancer subtypes using ResNet architectures on histopathological images. *Journal of Imaging*, 11(8). <https://doi.org/10.3390/jimaging11080284>
- Dhomane, L., & Shinde, S. (2023). Enhancing breast cancer diagnosis with deep learning in histopathology images. In *Proceedings of the 2023 1st DMIHER International Conference on Artificial Intelligence in Education and Industry 4.0 (IDICAIEI 2023)*. <https://doi.org/10.1109/IDICAIEI58380.2023.10406323>
- Długosz-Pokorska, A., Janecki, T., Janecka, A., & Gach-Janczak, K. (2025). Synergistic effects of oxaliplatin, 5-fluorouracil, and novel synthetic uracil analog U-359 on breast cancer cell carcinogenesis. *International Journal of Molecular Sciences*, 26(7). <https://doi.org/10.3390/ijms26072964>
- Fan, Y., Sun, K., Xiao, Y., Zhong, P., Meng, Y., Yang, Y., Du, Z., & Fang, J. (2025). Deep learning predicts HER2 status in invasive breast cancer from multimodal ultrasound and MRI. *Biomolecules and Biomedicine*. <https://doi.org/10.17305/bb.2025.12475>
- Hu, M., Zhang, L., Wang, X., & Xiao, X. (2025). Enhanced HER-2 prediction in breast cancer through synergistic integration of deep learning, ultrasound radiomics, and clinical data. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-12825-7>
- Jain, E., & Singh, A. (2024). Revolutionizing breast cancer diagnosis: VGG16's breakthrough in histopathological image classification. In *Proceedings of the 2024 4th International Conference on Soft Computing for Security Applications (ICSCSA 2024)* (pp. 386–391). <https://doi.org/10.1109/ICSCSA64454.2024.00068>
- Kaushik, P., & Choudhary, S. (2024). Enhanced breast cancer detection using ResNet50V2-based convolutional neural networks. In *Proceedings of the 2024 4th International Conference on Soft Computing for Security Applications (ICSCSA 2024)* (pp. 374–379). <https://doi.org/10.1109/ICSCSA64454.2024.00066>
- Kaushik, P., & Sharma, P. (2024). A MobileNetV3-based approach to breast cancer detection through transfer learning and histopathological image analysis. In *Proceedings of the 3rd International Conference on Advances in Computing, Communication and Materials (ICACCM 2024)*. <https://doi.org/10.1109/ICACCM61117.2024.11059095>
- Majanga, V., Mnkandla, E., Wang, Z., & Moulla, D. K. (2025a). Active contours connected component analysis segmentation method of cancerous lesions in unsupervised breast histology images. *Bioengineering*, 12(6). <https://doi.org/10.3390/bioengineering12060642>
- Majanga, V., Mnkandla, E., Wang, Z., & Moulla, D. K. (2025b). Automatic blob detection method for cancerous lesions in unsupervised breast histology images. *Bioengineering*, 12(4). <https://doi.org/10.3390/bioengineering12040364>
- Maruf, N. A., Basuhail, A., & Ramzan, M. U. (2025). Enhanced breast cancer diagnosis using multimodal feature fusion with radiomics and transfer learning. *Diagnostics*, 15(17). <https://doi.org/10.3390/diagnostics15172170>
- Momtahn, M., & Golnaraghi, F. (2025). A multitask CNN for near-infrared probe: Enhanced real-time breast cancer imaging. *Sensors*, 25(8). <https://doi.org/10.3390/s25082349>

- Oza, P., Sharma, P., Patel, S., Adedoyin, F., & Bruno, A. (2022). Image augmentation techniques for mammogram analysis. *Journal of Imaging*, 8(5). <https://doi.org/10.3390/jimaging8050141>
- Rai, H. M., Yoo, J., Agarwal, S., & Agarwal, N. (2025). LightweightUNet: Multimodal deep learning with GAN-augmented imaging data for efficient breast cancer detection. *Bioengineering*, 12(1). <https://doi.org/10.3390/bioengineering12010073>
- Sappa, N., & Lingam, G. (2025). An adaptive Cycle-GAN-based augmented LIME-enabled multi-stage transfer learning model for improving breast tumor detection using ultrasound images. *Electronics*, 14(8). <https://doi.org/10.3390/electronics14081571>
- Wang, Y. M., Wang, C. Y., Liu, K. Y., Huang, Y. H., Chen, T. B., Chiu, K. N., Liang, C. Y., & Lu, N. H. (2024). CNN-based cross-modality fusion for enhanced breast cancer detection using mammography and ultrasound. *Tomography*, 10(12), 2038–2057. <https://doi.org/10.3390/tomography10120145>