

BAB 1 PENDAHULUAN

1.1 Latar Belakang

Tumor ganas pada payudara atau yang dikenal sebagai kanker payudara, merupakan salah satu penyebab kematian pada wanita secara global. Dessie dan Zewotir (2025) telah melakukan meta-analisis komprehensif mengenai mortalitas kanker payudara secara global hingga Juni 2025. Hasil meta-analisis yaitu tercatat total kematian wanita akibat kanker payudara yang dilaporkan sebanyak 160.284 dari 2.447.155 pasien. Oleh karena itu, diperlukan deteksi dini kanker payudara sebagai upaya pencegahan dan pengendalian kanker melalui analisis citra jaringan yang diperoleh dari berbagai modalitas pencitraan seperti mamografi, *Magnetic Resonance Imaging* (MRI), dan Ultrasonografi (USG).

Deteksi dini tumor payudara menggunakan mamografi seperti skrining dan diagnosis berperan penting dalam meningkatkan peluang keberhasilan pengobatan. Meta-analisis yang dilakukan oleh Dibden dkk. (2020) menunjukkan bahwa program skrining mamografi dapat menurunkan mortalitas kanker payudara hingga 33%. Namun, berdasarkan penelitian Boita dkk. (2021), kemampuan dokter radiologi dalam menginterpretasikan citra *mammogram* dipengaruhi oleh berbagai faktor kualitas citra seperti peningkatan *noise*, perubahan kontras, dan penurunan resolusi spasial terutama pada kasus kalsifikasi yang lebih sensitif terhadap perubahan kualitas citra dibandingkan jaringan lunak. Keterbatasan tersebut berpotensi menyebabkan kesalahan dalam proses diagnosis kanker payudara.

Kesalahan interpretasi *mammogram* dapat berupa *False Positive* (FP), yaitu citra terdeteksi kanker pada pasien yang sebenarnya normal, dan *False Negative* (FN), yaitu kanker yang gagal terdeteksi. Penelitian Guertin dkk. (2018) terhadap 1.209 wanita tanpa kanker menunjukkan bahwa kualitas citra *mammogram* yang rendah dapat meningkatkan tingkat FP sebesar 10% hingga 30% sehingga menyebabkan pencitraan dan biopsi jaringan yang tidak perlu. Pada kasus lainnya Hovda dkk. (2023) melakukan penelitian terhadap 1.223 kasus kanker payudara

yang terdeteksi melalui skrining mamografi, menunjukkan tanda FN sebesar 31% menyebabkan keterlambatan diagnosis dan pengobatan.

Permasalahan FP dan FN mendorong pengembangan teknologi pengolahan citra *mammogram* berbasis *Artificial Intelligence* (AI), yang mencakup *machine learning* (ML) dan *deep learning* (DL). Alzubaidi dkk. (2021) melakukan tinjauan komprehensif mengenai perkembangan DL, khususnya arsitektur *Convolutional Neural Network* (CNN), yang memiliki kemampuan dalam pengolahan citra medis melalui pembelajaran pola dan ekstraksi fitur secara otomatis. Wisudawati (2021) melakukan penelitian dengan memanfaatkan kemampuan CNN pada tahap klasifikasi untuk membedakan citra berdasarkan kelas tertentu, seperti normal, tumor jinak, dan tumor ganas. Klasifikasi citra menggunakan metode CNN memiliki beberapa metrik penting meliputi akurasi, presisi, *recall* (sensitivitas), F1-score, dan *Area Under the Receiver Operating Characteristic Curve* (AUC atau AUROC) seperti yang digunakan oleh Talaat dkk. (2024) serta metrik spesifisitas seperti yang digunakan oleh Akbar dkk. (2025).

Pengembangan DL masih memiliki keterbatasan seperti kurangnya standarisasi data, kompleksitas bentuk lesi, kebutuhan data berlabel dalam jumlah besar, serta rendahnya interpretabilitas model (Gao dkk., 2023). Pengembangan pendekatan *Explainable Artificial Intelligence* (XAI) dengan memanfaatkan metode seperti *Shapley Additive Explanation* (SHAP), *3D Gradient-weighted Class Activation Mapping* (Grad-CAM), dan *Contextual Importance and Utility* (CIU) dapat meningkatkan transparansi model (Akbar dkk., 2025).

Beberapa peneliti telah mengembangkan berbagai model dari arsitektur CNN dikombinasikan dengan XAI untuk meningkatkan performa klasifikasi kanker payudara. Qasrawi dkk. (2024) membandingkan performa model VGG16 dan *DenseNet121* pada *dataset* publik dan *dataset* klinis untuk klasifikasi tumor jinak dan ganas, serta menerapkan metode Grad-CAM sebagai interpretasi visual model. Hasil penelitian menunjukkan bahwa *DenseNet121* memiliki performa terbaik dengan akurasi sebesar 98,3%, presisi sebesar 98,4%, *recall* sebesar 98,4%, F1-score sebesar 98,4%, dan AUC sebesar 1. Visualisasi Grad-CAM memperlihatkan bahwa area aktivasi model sesuai dengan lokasi lesi pada citra

mammogram. Selain itu, Chelloug dkk., (2025) menggunakan *modified InceptionNet-v3* untuk klasifikasi kanker payudara ke dalam tiga kelas, yaitu normal, tumor jinak, dan tumor ganas. Model yang dikembangkan menghasilkan akurasi sebesar 99,10%, *recall* sebesar 98,90%, presisi sebesar 99,90%, dan F1-score sebesar 98,80%. Grad-CAM menunjukkan bahwa model mampu memfokuskan perhatian pada area lesi.

Singh dkk. (2022) melakukan analisis grafik pelatihan dan validasi model *Inception-ResNet V2* pada *dataset Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDSM)* untuk klasifikasi tumor jinak dan ganas pada payudara dengan membandingkan beberapa *optimizer*, yaitu ADAM, AdaGrad, AdaDelta, dan RMSProp. Hasil penelitian menunjukkan metrik klasifikasi pada data validasi menggunakan *optimizer* ADAM memiliki akurasi sebesar 93%, AUC sebesar 95%, dan *recall* sebesar 96%. Penggunaan *Inception-ResNet V2* juga dikembangkan oleh Talaat dkk. (2024) yang mengombinasikan arsitektur 3D *Inception-ResNet V2* dengan metode Grad-CAM untuk klasifikasi tiga kelas, yaitu normal, jinak, dan ganas. Penelitian tersebut menghasilkan performa dengan akurasi sebesar 98,53%, *recall* sebesar 98,53%, presisi sebesar 98,40%, F1-score sebesar 98,43%, dan AUROC sebesar 99,33%. Hasil interpretasi visual Grad-CAM menunjukkan bahwa *heatmap* tidak selalu berfokus pada area lesi yang relevan akibat adanya artefak pada citra.

Selain metrik klasifikasi seperti akurasi, presisi, *recall*, F1-score, dan AUROC, spesifisitas juga merupakan parameter penting dalam mengevaluasi kinerja model dalam mengukur kemampuan model dalam mengenali sampel negatif secara benar sehingga dapat mengurangi terjadinya *False Positive*. Penelitian oleh Fardela dkk. (2024) yang membandingkan metode CNN dan model *Roboflow* untuk klasifikasi citra foto toraks COVID-19 menunjukkan bahwa CNN mencapai spesifisitas sebesar 93,33%. Selain itu, analisis *Receiver Operating Characteristic (ROC)* menunjukkan bahwa kedua metode memiliki kemampuan yang baik dalam membedakan citra pasien COVID-19 dan pasien normal.

Berdasarkan kajian tersebut, diperlukan penelitian yang tidak hanya berfokus pada pencapaian performa klasifikasi yang tinggi, tetapi juga pada analisis

interpretabilitas dan kestabilan model *deep learning*. Oleh karena itu, pada penelitian ini dilakukan analisis performa model *Inception-ResNet V2* dengan menganalisis grafik proses pelatihan dan validasi, menganalisis matrik konfusi dan metrik klasifikasi, analisis kurva ROC, serta interpretasi menggunakan Grad-CAM. Kombinasi *dataset* citra *mammogram* publik dari *platform Kaggle* (<https://www.kaggle.com/>) dan dari RSUP Dr. M. Djamil Padang diharapkan dapat meningkatkan kemampuan generalisasi model sekaligus menghasilkan interpretasi visual yang lebih relevan untuk mendukung proses diagnosis kanker payudara.

1.2 Tujuan Penelitian

1. Menganalisis grafik hasil pelatihan dan validasi model *Inception-ResNet V2* dalam mengklasifikasikan citra *mammogram* ke dalam kelas normal dan abnormal (tumor).
2. Menganalisis hasil matriks konfusi dan metrik klasifikasi data uji model *Inception-ResNet V2*.
3. Menganalisis kurva *Receiver Operating Characteristic* (ROC) model *Inception-ResNet V2*.
4. Menginterpretasikan visualisasi *Gradient-weighted Class Activation Mapping* (Grad-CAM) pada model *Inception-ResNet V2*.

1.3 Manfaat Penelitian

1. *Inception-ResNet V2* dapat mengelompokkan citra *mammogram* sekaligus menampilkan bagian citra yang paling memengaruhi keputusan model melalui visualisasi Grad-CAM, sehingga dapat menjadi referensi bagi peneliti selanjutnya dalam mengembangkan model klasifikasi citra *mammogram*.
2. Memberikan gambaran awal bagi dokter radiologi mengenai kemampuan model *Inception-ResNet V2* dalam mengelompokkan citra *mammogram*, sehingga dapat dipahami potensi penggunaannya sebagai bantuan awal untuk mempermudah dan menyamakan proses pembacaan citra, tanpa menggantikan keputusan medis dokter.

1.4 Ruang Lingkup Penelitian

1. Menggunakan data citra *mammogram* dua dimensi (2D) berjumlah 3.000 citra meliputi 2.800 citra dari *dataset* publik (*platform kaggle*) dan 200 citra dari RSUP Dr. M. Djamil Padang sesuai jumlah data yang tersedia rumah sakit tersebut. Penggunaan citra 2D dipilih karena mampu merepresentasikan karakteristik jaringan payudara, memiliki ketersediaan data lebih luas, dan membutuhkan komputasi lebih rendah dibandingkan citra 3D.
2. Klasifikasi citra dibagi menjadi kelas normal (kategori BI-RADS 1) dan abnormal (kategori BI-RADS 2 hingga 6) menggunakan model *Inception-ResNet V2*.
3. Evaluasi performa model dilakukan melalui analisis grafik *loss*, akurasi, presisi, *recall* (sensitivitas), *F1-score*, spesifisitas, dan *Area Under the Receiver Operating Characteristic Curve* (AUROC) terhadap *epoch* selama proses pelatihan (*training*) dan validasi (*validation*).
4. Analisis matriks konfusi dan metrik klasifikasi pada data yang di tes (*tesing*) yang meliputi akurasi, presisi, *recall*, *F1-score*, spesifisitas dan AUROC.
5. Analisis kurva *Receiver Operating Characteristic* (ROC) berupa hubungan antara *True Positive Rate* (TPR) dan *False Positive Rate* (TPR) serta kategori AUROC.
6. Interpretasi hasil klasifikasi melalui visualisasi *heatmap* menggunakan metode Grad-CAM untuk mengidentifikasi area citra yang berkontribusi terhadap keputusan model. Visualisasi *heatmap* yaitu teknik visualisasi data yang menggunakan warna untuk merepresentasikan nilai atau intensitas suatu data.