

# PENDAHULUAN

## 1.1 Latar Belakang

Kanker payudara menjadi salah satu kanker dengan penderita paling banyak di Indonesia, dengan mayoritas penderitanya adalah wanita. Per tahun 2020, kasus baru kanker payudara mencapai 65.858 kasus (Kemenkes RI, 2022). Pencegahan dari kemungkinan kematian yang diakibatkan oleh kanker payudara dapat dikurangi dengan melakukan deteksi dini. Cara paling sederhana adalah dengan pemeriksaan payudara oleh diri sendiri dengan merasakan kondisi payudaranya berdasarkan massa, adanya cairan, pembengkakan, dan kemungkinan adanya kelainan lain (D. Z. Haq et al., n.d.). Namun, cara ini tidak akurat dan dapat memicu diagnosis yang berlebihan.

Alternatif lain untuk deteksi dini adalah dengan memanfaatkan citra ultrasound untuk mengklasifikasikan kondisi payudara menjadi beberapa kelas, yakni normal, tumor jinak, dan kanker. Salah satu hasil perkembangan yang pesat di bidang machine learning adalah adanya teknologi deep learning, yang dapat mempelajari fitur pada citra serta secara otomatis melakukan klasifikasi melalui Transfer Learning Convolutional Neural Network (CNN) (Guan & Loew, n.d.). Pada umumnya, model CNN memerlukan dataset yang besar untuk dapat memberikan performa terbaiknya. Di sisi lain, dataset medis seperti citra ultrasound payudara cenderung terbatas dan memiliki kelas yang tidak seimbang (Rguibi et al., 2023). Salah satu cara umum untuk mengatasi keterbatasan data dan ketidakseimbangan kelas yang menjadi masalah umum dalam proses pelatihan model CNN (Krizhevsky et al., n.d.) adalah dengan melakukan augmentasi data.

GAN merupakan salah satu metode generatif yang dapat menghasilkan citra sintesis dengan mempelajari pola distribusi data citra asli. Meskipun implementasi GAN dalam bidang medis masih tergolong awal, tetapi potensi GAN dalam bidang ini sangat potensial dan signifikan (Septiandi, 2023). Beberapa penelitian terdahulu yang mengimplementasikan GAN dalam augmentasi data untuk model deep learning di antaranya adalah untuk menghasilkan sintesis data citra x-ray COVID-19 (Septiandi, 2023), data rantai

RNA penderita kanker (Xiao et al., 2021), dan untuk data tabular kanker payudara (Inan et al., 2023).

GAN yang diperkenalkan oleh Goodfellow menggunakan divergensi Jensen-Shannon (JS *divergence*) dalam perhitungan fungsi kerugian atau *loss function* (Goodfellow et al., 2014). Implementasi JS *divergence* memiliki kelemahan di mana identik dengan terjadinya gradien hilang (*vanishing gradient*) yang mengakibatkan proses training GAN menjadi tidak stabil. Oleh karenanya, Arjovsky memperkenalkan varian baru GAN yang bernama Wasserstein GAN (Arjovsky et al., 2017). Berbeda dengan GAN original yang menggunakan JS *divergence* untuk perhitungan loss function, WGAN menggunakan perhitungan jarak Wasserstein.

Modifikasi ini membuat WGAN memiliki beberapa keunggulan dibandingkan dengan GAN original sehingga dipilih dalam penelitian ini. WGAN menggunakan jarak Wasserstein untuk menghitung *loss function* sehingga membantu mengatasi masalah *vanishing gradient* yang seringkali terjadi pada GAN original. Implementasi ini membuat proses training WGAN lebih stabil dan terukur. Selain itu, WGAN menyediakan indikator kemajuan yang lebih jelas selama proses training, sehingga peneliti dapat memonitor dan menyesuaikan model dengan lebih efektif dengan tujuan untuk menghasilkan citra sintesis dengan kualitas lebih baik, yang sangat penting dalam aplikasi medis seperti citra ultrasound payudara. WGAN juga telah terbukti efektif dalam berbagai aplikasi medis lainnya, menunjukkan adaptabilitas yang tinggi dan keandalannya dalam berbagai konteks medis seperti yang terlihat dari hasil penelitian oleh (Xiao et al., 2021) dan (Hussain et al., 2022).

Pada penelitian ini, data citra ultrasound payudara akan diaugmentasi dengan menggunakan WGAN untuk menghasilkan data citra sintesis yang dapat mengatasi masalah ketidakseimbangan data pada tiap kelas. Seberapa signifikan efek implementasi WGAN dengan meningkatkan jumlah data training terhadap akurasi prediksi akan diukur dengan mengevaluasi model Transfer Learning CNN. Dengan demikian, penelitian ini berupaya menjembatani kesenjangan antara

kebutuhan yang mendesak akan deteksi dini kanker payudara di Indonesia dan tantangan yang diakibatkan oleh masalah keterbatasan data.

## **1.2 Rumusan Masalah**

Berdasarkan Latar belakang yang telah dijelaskan sebelumnya, didapatkan rumusan masalah sebagai berikut:

1. Bagaimana merancang dan mengembangkan model WGAN untuk augmentasi data citra kanker payudara yang terbatas?
2. Seberapa signifikan proses augmentasi data citra kanker payudara menggunakan WGAN terhadap akurasi model Transfer Learning untuk melakukan prediksi klasifikasi kanker payudara?
3. Bagaimana hasil model Transfer Learning yang telah dilatih diintegrasikan ke sistem diagnosis online?

## **1.3 Batasan Masalah**

Batasan masalah yang ditetapkan pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan pada penelitian ini adalah data sekunder, yang berarti tidak diambil secara langsung.
2. Penelitian ini fokus pada pengaruh data citra sintesis yang dihasilkan WGAN terhadap model klasifikasi. Evaluasi model WGAN tidak menjadi fokus.
3. Luaran dari penelitian ini adalah analisis evaluasi kinerja Transfer Learning CNN untuk melakukan prediksi klasifikasi kanker payudara antara saat menggunakan dataset yang asli dan dataset yang telah ditambah dengan data sintesis.
4. Prototipe aplikasi diagnosis online yang dibuat hanya untuk memberikan gambaran terkait pemanfaatannya. Penekanan pada aspek teknis bukan fokus utamanya.

## **1.4 Tujuan Penelitian**

Berdasarkan rumusan masalah yang ada, tujuan penelitian dirumuskan sebagai berikut:

1. Merancang dan mengembangkan model WGAN untuk augmentasi data citra kanker payudara yang terbatas.
2. Mengetahui signifikansi proses augmentasi data citra kanker payudara menggunakan WGAN terhadap akurasi model Transfer Learning untuk melakukan prediksi klasifikasi kanker payudara.
3. Mengetahui bagaimana hasil model Transfer Learning yang telah dilatih diintegrasikan ke sistem diagnosis online

### **1.5 Manfaat Penelitian**

Manfaat yang diharapkan dapat diberikan oleh adanya penelitian ini di antaranya adalah:

1. Manfaat teoritis
  - a. Memberikan pemahaman mengenai implementasi data augmentasi dengan WGAN
  - b. Memberikan pemahaman terkait bagaimana proses data augmentasi dapat memengaruhi performa Transfer Learning CNN dalam memprediksi klasifikasi kanker payudara untuk deteksi dini kanker.
  - c. Memberikan gambaran potensi pembuatan sistem diagnosis/klasifikasi online yang berbasis Transfer Learning CNN
2. Manfaat praktis
  - a. Membantu melakukan deteksi dini adanya kanker payudara, sehingga diagnosa bisa didapat dengan cepat dan akurat, serta memungkinkan penentuan perawatan yang lebih awal.
  - b. Menjadi bahan referensi untuk penelitian di masa depan, utamanya yang berfokus pada data augmentasi dengan metode generatif GAN dan klasifikasi penyakit medis menggunakan Transfer Learning CNN.