

I. PENDAHULUAN

1.1 Latar Belakang

Kanker adalah penyakit di mana beberapa dari sel tumbuh tidak terkendali dan menyebar ke bagian tubuh lainnya (National Cancer Institute, 2021). Kanker dapat bermula hampir di mana saja dalam tubuh manusia, yang terdiri dari triliunan sel. Payudara adalah salah satu organ tubuh yang dapat terserang kanker. Kanker payudara terjadi karena adanya sel-sel yang tumbuh dengan tidak terkendali, hingga mengambil jaringan sehat di payudara dan sekitarnya, paling sering muncul di epitel duktal (karsinoma duktal) tetapi juga dapat berkembang di lobulus payudara (karsinoma lobular) (Kashyap et al., 2022).

Menurut data *Global Burden Cancer* yang dikumpulkan oleh *International Agency for Research on Cancer* (IARC), kanker payudara adalah jenis kanker dengan persentase kasus baru (berdasarkan umur) tertinggi, yaitu 43,3%, dan persentase kematian akibat kanker payudara sebesar 12,9% per 100.000 perempuan di Indonesia. Kanker payudara merupakan kanker yang paling umum didiagnosis pada wanita, menyumbang lebih dari 10% diagnosis kanker setiap tahunnya, dan merupakan penyebab kematian akibat kanker paling umum kedua di seluruh dunia. Pada tahun 2022, terdapat 2,3 juta wanita yang didiagnosis menderita kanker payudara dan 670.000 kematian di seluruh dunia. (WHO, 2024).

Tumor pada payudara dapat digunakan untuk mendeteksi kanker payudara secara dini. Tumor ini biasanya diklasifikasikan sebagai *benign* (jinak) dan *malignant* (ganas) (Ravly et al., 2022). Tenaga medis perlu menggunakan pendekatan penentuan aktif untuk mendeteksi kanker ganas, tetapi bahkan untuk profesional, mengidentifikasi keganasan suatu kanker sangatlah sulit. Meskipun tenaga ahli berpengalaman, interpretasi manual citra ini rentan terhadap ketidakkonsistenan dan kesalahan. Oleh karena itu, penerapan sistem berbasis algoritma sangat diperlukan untuk membantu mendiagnosis keganasan secara lebih akurat dan efisien untuk mendeteksi kanker (Pania et al., 2024). Salah satu solusi yang paling efektif adalah menggunakan *machine learning*.

Machine Learning adalah cabang dari kecerdasan buatan yang memungkinkan komputer untuk belajar dari data dan memprediksi pola yang ditemukan. Beberapa model *machine learning* yang populer *Convolutional Neural Network* (CNN) dan *Random Forest*, CNN dikenal dengan sangat baik dalam mengolah data citra berupa gambar 2D. *Random Forest* memiliki sifat *versatile* dan tanpa *tuning* yang dapat memberikan akurasi yang baik dalam klasifikasi (Henderi et al., 2021). Klasifikasi ini sangat penting bagi para dokter untuk

mendeteksi dan mendiagnosis kanker payudara pada tahap awal, sehingga meningkatkan peluang penyembuhan pasien. Algoritma-algoritma ini memungkinkan analisis yang lebih cepat dan akurat terhadap data medis, membantu para profesional kesehatan memanfaatkan data secara lebih efektif dalam mengambil keputusan klinis yang tepat (Goenawan & Hartati, 2024).

Dalam beberapa tahun terakhir, teknologi *machine learning*, khususnya algoritma *Convolutional Neural Network* (CNN) dan *Random Forest*, telah menunjukkan potensi besar dalam pengolahan citra medis. CNN merupakan metode *deep learning* yang dirancang untuk mengekstrak fitur dari citra dengan cara yang efisien dan otomatis. Penelitian oleh (Ulagamuthalvi et al., 2022) menunjukkan bahwa CNN dapat mencapai *accuracy* hingga 99,90% dalam klasifikasi gambar mammografi, menandakan efektivitasnya dalam mendeteksi kanker payudara secara dini. Di sisi lain, *Random Forest* adalah algoritma *ensemble learning* yang dikenal karena kemampuannya dalam menangani data dengan dimensi tinggi dan memberikan hasil yang stabil tanpa memerlukan *tuning* parameter yang rumit (Kathale & Thorat, 2020).

Berdasarkan latar belakang ini, maka peneliti melakukan klasifikasi penyakit kanker payudara dengan model *Convolutional Neural Network* (CNN) dan *Random Forest*. Meskipun telah banyak penelitian yang menggunakan berbagai algoritma *machine learning* untuk klasifikasi penyakit kanker payudara, belum ada studi yang secara khusus membandingkan efektivitas antara algoritma CNN dan *Random Forest* dalam konteks analisis citra mammogram. Sebagian besar penelitian sebelumnya lebih fokus pada penggunaan satu jenis algoritma atau tidak menyebutkan penggunaan Python sebagai alat untuk implementasi algoritma ini. Dengan menggunakan Python dan berbagai *library* terkait seperti *TensorFlow* dan *Scikit-Learn*, penelitian ini bertujuan untuk mengembangkan sistem klasifikasi kanker payudara yang lebih efisien.

Pada penelitian ini, peneliti melakukan klasifikasi penyakit kanker payudara menjadi dua jenis yaitu *benign* dan *malignt*. Penelitian diujikan dengan perbandingan klasifikasi *Convolutional Neural Network* (CNN) dan *Random Forest* dengan melihat perbandingan nilai *accuracy*, *recall*, *precision*, dan *F1-score* berdasarkan parameter yang ada pada dataset.

1.2 Identifikasi dan Rumusan Masalah

Identifikasi Masalah

Kanker Payudara merupakan penyakit yang sangat serius dan dapat menyerang dengan cepat sehingga dapat mengakibatkan fatal jika tidak ditangani dengan segera. Salah satu cara untuk melihat algoritma klasifikasi yang lebih baik dan efektif dalam klasifikasi kanker dengan menggunakan *Convolutional*

Neural Network (CNN) dan *Random Forest* dengan bahasa Python. Tujuannya adalah menentukan model mana yang paling efektif dalam mengklasifikasikan kasus kanker payudara, serta mengidentifikasi faktor-faktor kunci yang berkontribusi pada diagnosis penyakit kanker payudara.

Rumusan Masalah

Berdasarkan latar belakang dan penjelasan identifikasi masalah diatas, maka rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana kinerja algoritma *Convolutional Neural Network* (CNN) dan *Random Forest* dalam mengklasifikasikan kanker payudara berdasarkan dataset citra mammogram?
2. Algoritma mana yang memberikan *accuracy*, *precision*, *recall*, dan *F1-score* tertinggi berdasarkan dataset mammogram yang digunakan?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah :

- a. Menerapkan algoritma *Convolutional Neural Network* (CNN) dan *Random Forest* untuk klasifikasi kanker payudara berdasarkan dataset citra mammogram.
- b. Membandingkan kinerja algoritma dalam hal *accuracy*, *precision*, *recall*, dan *F1-score* dalam klasifikasi penyakit kanker payudara berdasarkan dataset citra mammogram.

1.4 Manfaat Penelitian

Adapun manfaat penelitian ini adalah sebagai berikut :

Manfaat Bagi Mahasiswa

1. Mahasiswa dapat mengembangkan keterampilan dalam memproses penyakit kanker payudara, penerapan algoritma *Convolutional Neural Network* (CNN) dan *Random Forest* dan pemrograman python, menjadi keterampilan yang sangat berharga di dalam dunia teknologi.
2. Mahasiswa dapat mengklasifikasikan penyakit kanker payudara berdasarkan dataset dengan model *Convolutional Neural Network* (CNN) dan *Random Forest*.
3. Mahasiswa dapat menambah wawasan dan pengetahuan dalam bidang fisika medis dan komputasi khususnya mengenai klasifikasi penyakit kanker payudara.

Manfaat Bagi Masyarakat

1. Penelitian ini dapat membantu khususnya pekerja di bidang tenaga medis dan memberikan informasi dalam mendiagnosakan penyakit kanker

payudara dengan lebih akurat. Memungkinkan skrining awal dan penanganan yang lebih cepat.

2. Memungkinkan pendekatan yang lebih dalam perawatan pasien dengan mengidentifikasi pola-pola yang mungkin tidak terlihat oleh manusia.