BAB 1
PENDAHULUAN

1.1 Latar Belakang Masalah

Kanker payudara merupakan salah satu penyebab utama kematian pada perempuan di seluruh dunia, dengan angka kejadian yang terus meningkat setiap tahunnya. Oleh karena itu, pengembangan teknologi yang mampu mendukung deteksi secara efektif dan efisien menjadi kebutuhan yang mendesak [1]. Namun, keterbatasan data medis yang tersebar di berbagai institusi serta bersifat rahasia sering menjadi hambatan dalam pengembangan model kecerdasan buatan untuk deteksi kanker payudara [2]. Data medis pasien, termasuk citra medis, dikategorikan sebagai informasi yang sangat sensitif dan harus dijaga kerahasiaannya karena kebocoran data dapat menyebabkan pelanggaran privasi dan konsekuensi hukum. Regulasi perlindungan data seperti *Health Insurance Portability and Accountability Act* (HIPAA) dan *General Data Protection Regulation* (GDPR) secara tegas mengatur bahwa data kesehatan tidak boleh diproses atau dibagikan tanpa mekanisme perlindungan privasi yang memadai [3].

Federated Learning (FL) hadir sebagai solusi dengan menawarkan pelatihan model terdistribusi tanpa perlu memindahkan data mentah ke server pusat. Setiap institusi melatih model lokalnya sendiri, lalu hanya parameter model yang dikirim untuk diagregasi, sehingga kolaborasi dapat dilakukan tanpa melanggar privasi data pasien [4]. Jiménez-Sánchez et al. [5] mengembangkan metode *Fed-Align-CL* berbasis ResNet22 untuk klasifikasi kanker payudara dan memperoleh nilai AUC 79% serta PR-AUC 82%. Meskipun menyinggung pentingnya privasi, penelitian tersebut belum sepenuhnya menerapkan mekanisme perlindungan data. Sejalan dengan itu, Pati et al. [1] menegaskan bahwa penerapan *Federated Learning* di bidang kesehatan memerlukan perhatian khusus terhadap aspek privasi karena tingginya risiko kebocoran data medis.

Federated Learning menjaga data tetap berada di institusi masing-masing, namun proses pertukaran parameter model masih memiliki risiko kebocoran informasi sensitif selama tahap agregasi di server [6]. Oleh karena itu, diperlukan mekanisme tambahan yang dapat menjamin keamanan data dalam proses tersebut. *Homomorphic Encryption* (HE) menjadi salah satu solusi karena memungkinkan perhitungan dilakukan langsung pada parameter terenkripsi tanpa perlu dekripsi

terlebih dahulu. Salah satu skema yang menjanjikan adalah Cheon-Kim-Kim-Song (CKKS), yang mampu melakukan operasi langsung pada *ciphertext* sehingga menjaga kerahasiaan parameter model selama proses pelatihan dan agregasi [7, 8].

Skema CKKS memiliki keterbatasan dalam menyeimbangkan performa, efisiensi, dan tingkat keamanan, serta dibatasi oleh panjang vektor maksimum 16.384 sesuai standar *Homomorphic Encryption*, yang menyulitkan enkripsi parameter model yang berukuran besar. Selain itu, peningkatan tingkat keamanan melalui penggunaan panjang kunci yang lebih besar berpotensi meningkatkan beban komputasi dan biaya komunikasi. Penelitian oleh Pan et al. [6] mengusulkan solusi berupa *Segmented CKKS Homomorphic Encryption*. Metode ini menggunakan enkripsi tersegmentasi untuk mengatasi keterbatasan ukuran vektor dan diharapkan mampu menjaga keseimbangan antara efisiensi komputasi, keamanan, dan performa model.

Penelitian ini berfokus pada implementasi *Segmented CKKS Homomorphic Encryption* pada *Federated Learning* dengan model *ResNet22* untuk klasifikasi kanker payudara. Tujuan penelitian ini adalah mencapai keseimbangan antara performa model dan perlindungan privasi data. Penelitian ini juga melakukan evaluasi untuk melihat sejauh mana penerapan protokol privasi tersebut memengaruhi kinerja model, khususnya terhadap nilai *Accuracy*, AUC, PR-AUC, waktu komputasi, dan *communication cost* yang menjadi indikator penting dalam menilai performa dan privasi klasifikasi.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana implementasi dan performa *Segmented CKKS Homomorphic Encryption* pada *Federated Learning* dengan model ResNet22 untuk klasifikasi kanker payudara?
2. Seberapa besar pengaruh penerapan *Segmented CKKS Homomorphic Encryption* terhadap performa model dan overhead sistem pada *Federated Learning* dengan model ResNet22, yang diukur menggunakan metrik *Accuracy*, AUC, PR-AUC, waktu komputasi, dan *communication cost* pada berbagai konfigurasi panjang kunci dan *security level*?

1.3 Batasan Permasalahan

Batasan permasalahan pada penelitian ini ditetapkan untuk memperjelas ruang lingkup dan fokus penelitian agar tidak melebar dari tujuan yang telah dirumuskan. Adapun batasan permasalahan dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini menggunakan dataset publik *Curated Breast Imaging Subset of DDSM (CBIS-DDSM)* yang berasal dari The Cancer Imaging Archive (TCIA) dan diakses melalui *platform* Kaggle [9].
2. Implementasi *Federated Learning* dilakukan dalam bentuk *local federated learning simulation* pada satu mesin (*single-machine*), dengan melibatkan tiga klien dan satu server.
3. Penelitian ini menggunakan *Segmented CKKS Homomorphic Encryption* sebagai mekanisme perlindungan privasi dalam Federated Learning, dengan tingkat keamanan 128-bit dan 192-bit, panjang kunci (*key length*) sebesar 8192, 16384, dan 32768, serta kedalaman multiplikasi (*multiplicative depth*) sebesar 3.
4. Dalam penelitian ini diasumsikan bahwa seluruh klien bersifat jujur (*honest*) dan tidak melakukan kolusi maupun penyalahgunaan data antar klien [6].
5. Server agregasi diasumsikan bersifat *semi-honest* (*honest-but-curious*), yaitu server menjalankan protokol Federated Learning sesuai aturan yang ditetapkan, namun berpotensi mencoba memperoleh informasi sensitif dari parameter model yang diterima [6].

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengimplementasikan *Segmented CKKS Homomorphic Encryption* pada *Federated Learning* dengan model ResNet22 untuk klasifikasi kanker payudara.
2. Menganalisis pengaruh penerapan *Segmented CKKS Homomorphic Encryption* terhadap performa model dan overhead sistem pada *Federated*

Learning dengan model *ResNet22*, yang diukur berdasarkan nilai *Accuracy*, AUC, PR-AUC, waktu komputasi, dan *communication cost* pada berbagai konfigurasi panjang kunci dan *security level*.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Memberikan kontribusi dalam pengembangan model *Federated Learning* dengan *Segmented CKKS Homomorphic Encryption* yang aman dan menjaga privasi data di bidang medis.
2. Menjadi referensi bagi penelitian lanjutan mengenai penerapan *Federated Learning* dengan enkripsi untuk kasus medis lain yang membutuhkan perlindungan data sensitif.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun menjadi lima bab utama, yaitu sebagai berikut:

1. **Bab 1 – Pendahuluan**
   Bab ini berisi latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan. Bagian ini memberikan gambaran umum tentang alasan dan arah penelitian yang dilakukan.
2. **Bab 2 – Landasan Teori**
   Bab ini membahas teori-teori yang relevan dengan penelitian, termasuk konsep deteksi kanker payudara, *Federated Learning*, *Homomorphic Encryption*, serta penelitian terkait. Landasan teori ini menjadi dasar konseptual dalam perancangan dan pelaksanaan penelitian.
3. **Bab 3 – Metodologi Penelitian**
   Bab ini menjelaskan metode yang digunakan dalam penelitian, mulai dari pengumpulan data, praproses data, perancangan model, implementasi *Federated Learning* serta *Segmented CKKS Homomorphic Encryption*, dan evaluasi performa model.

4. **Bab 4 – Hasil dan Diskusi**
   Membahas hasil eksperimen dari model dalam bentuk grafik dan juga tabel

5. **Bab 5 – Kesimpulan dan Saran**
   Bab ini berisi kesimpulan yang diperoleh dari hasil penelitian serta saran yang dapat digunakan untuk pengembangan penelitian selanjutnya.