BAB 1
PENDAHULUAN

1.1 Latar Belakang Masalah

Kanker payudara merupakan salah satu masalah kesehatan utama di dunia. Deteksi dini terbukti meningkatkan peluang keberhasilan pengobatan, sehingga pengembangan sistem berbasis kecerdasan buatan (*Artificial Intelligence*/AI) menjadi penting untuk mendukung diagnosis klinis [1, 2]. Namun, penerapan AI di bidang medis menghadapi tantangan besar terkait privasi dan kerahasiaan data pasien. Data medis bersifat sangat sensitif dan diatur ketat oleh regulasi seperti HIPAA di Amerika Serikat dan GDPR di Uni Eropa [3, 4, 5, 6]. Oleh karena itu, dibutuhkan metode komputasi yang dapat menjaga privasi pasien tanpa mengorbankan kinerja model [5].

Federated Learning (FL) merupakan pendekatan pembelajaran terdistribusi yang memungkinkan pelatihan model lintas institusi tanpa memindahkan data mentah, melainkan hanya membagikan parameter model [5, 6, 7]. Pendekatan ini banyak digunakan di bidang medis, termasuk klasifikasi kanker payudara, karena mampu menjaga desentralisasi data. Jiménez-Sánchez *et al.*[8] menggunakan arsitektur *ResNet-22* dalam sistem *Federated Learning* untuk klasifikasi kanker payudara dan melaporkan bahwa metode *Fed-Align-CL* yang mereka usulkan mencapai AUC sebesar 79% dan PR-AUC sebesar 82%, melampaui beberapa metode federasi lainnya. Hasil ini menunjukkan bahwa kombinasi antara FL dan arsitektur *CNN* yang dalam seperti *ResNet-22* memiliki potensi tinggi untuk meningkatkan performa klasifikasi citra medis. Namun, FL tetap rentan terhadap berbagai serangan privasi, seperti serangan *model inversion* atau *membership inference*, di mana informasi sensitif dari data pelatihan dapat diperkirakan kembali dari parameter model yang dibagikan [3, 9]. Pati *et al.* [3] menegaskan bahwa serangan semacam ini menunjukkan perlunya mekanisme privasi yang lebih kuat untuk melindungi data medis dalam sistem FL.

Salah satu solusi yang digunakan adalah *Homomorphic Encryption* (HE), yang memungkinkan operasi matematika dilakukan langsung pada data terenkripsi tanpa perlu proses dekripsi, sehingga server dapat melakukan agregasi model tanpa mengakses informasi sensitif [10, 11, 12, 13]. Namun, skema HE konvensional masih memiliki keterbatasan efisiensi karena ukuran *ciphertext* yang besar dan

waktu komputasi yang tinggi. Untuk mengatasinya, dikembangkan pendekatan *Packed CKKS Homomorphic Encryption*, yang mengemas sejumlah nilai *plaintext* ke dalam satu *ciphertext*, sehingga memungkinkan komputasi paralel di ranah terenkripsi dan mengurangi waktu komputasi [13, 14].

Packed CKKS Homomorphic Encryption (HE) merupakan skema enkripsi homomorfik modern yang memungkinkan operasi aritmetika aproksimasi pada bilangan real, sehingga sangat sesuai untuk proses agregasi berbobot dalam *Federated Learning* [12, 14]. Penelitian oleh Nan Yan *et al.* [14] menunjukkan bahwa pendekatan *Federated Learning* yang mengombinasikan *Packed CKKS Homomorphic Encryption* (HE) dengan mekanisme *Top-k sparsification* mampu mencapai akurasi yang mendekati model *plain FL*, yakni 95.04% pada MNIST, 88.96% pada Fashion-MNIST, dan 71.37% pada CIFAR-10. Selain itu, metode ini juga berhasil mengurangi beban komputasi dan ukuran data terenkripsi hingga 16,89 kali dibandingkan dengan skema CKKS konvensional. Hasil tersebut membuktikan bahwa integrasi antara mekanisme *packing* dan *sparsification* pada CKKS dapat mempertahankan privasi tanpa mengorbankan efisiensi maupun performa model.

Penelitian ini berfokus pada implementasi *Packed CKKS Homomorphic Encryption* dalam *Federated Learning* dengan model CNN ResNet-22 untuk klasifikasi kanker payudara berbasis citra mamografi. Sistem dirancang agar proses agregasi parameter dilakukan secara aman di ranah terenkripsi dengan mekanisme *top-k sparsification*. Evaluasi dilakukan terhadap performa model (Akurasi, AUC dan PR-AUC), waktu komputasi (enkripsi, dekripsi, agregasi, dan *training time*), serta *data transfer volume* (MB) untuk menilai efektivitas dan efisiensi penerapan *Packed CKKS Homomorphic Encryption* pada sistem *Federated Learning* di lingkungan medis.

1.2 Rumusan Masalah

1. Bagaimana cara implementasi *Packed CKKS Homomorphic Encryption* pada *Federated Learning* diterapkan dalam sistem klasifikasi kanker payudara dengan model CNN ResNet-22?
2. Bagaimana implementasi *Packed CKKS Homomorphic Encryption* memengaruhi performa (Akurasi, AUC dan PR-AUC), efisiensi komputasi, dan biaya komunikasi pada sistem *Federated Learning* berbasis CNN ResNet-22 untuk klasifikasi kanker payudara?

1.3 Batasan Masalah

1. Dataset yang digunakan terbatas pada *CBIS-DDSM Breast Cancer Image Dataset* yang tersedia secara publik di Kaggle.
2. Penelitian difokuskan pada klasifikasi kanker payudara menggunakan model CNN ResNet-22 dalam skema *Federated Learning* tanpa akses langsung terhadap data mentah pada tiap klien.
3. Mekanisme privasi dibatasi pada penerapan *Packed CKKS Homomorphic Encryption* dengan parameter `poly_modulus_degree` sebesar 8192 serta konfigurasi `coeff_modulus_bit_sizes` = [60, 40, 40, 60].
4. Simulasi federasi dijalankan secara lokal pada satu perangkat dan satu proses (*single-process*) melalui satu terminal, dengan tiga klien dan satu server agregasi disimulasikan dalam satu skrip.
5. Komunikasi federasi pada penelitian ini disimulasikan dalam satu proses Python tanpa jaringan. Pertukaran model/*cipher* dilakukan melalui pemanggilan fungsi dan passing variabel antar-role (client–server).
6. Teknik sparsifikasi model menggunakan metode *Top-k* dengan beberapa variasi rasio, yaitu $k = 0.2$, $k = 0.5$, dan $k = 0.8$, untuk mengevaluasi pengaruh tingkat kompresi terhadap performa dan biaya komunikasi.
7. Dalam penelitian ini diasumsikan bahwa pemilik data melakukan pelatihan model secara kolaboratif tanpa melakukan pertukaran data citra medis mentah antar institusi [8].
8. Mekanisme enkripsi mengacu pada asumsi bahwa pasangan kunci *Homomorphic Encryption* (HE key-pair) telah dibagikan kepada seluruh klien melalui saluran komunikasi yang aman (*secure channel*), serta *Parameter Server* (PS) tidak berkolusi dengan klien mana pun [14].
9. Model ancaman (*threat model*) yang digunakan mengikuti asumsi bahwa *Parameter Server* (PS) tidak berkolusi dengan klien mana pun dan bahwa baik PS maupun klien bersifat *honest-but-curious*, yaitu menjalankan protokol *federated learning* yang ditetapkan namun berpotensi mencoba memperoleh informasi dari pembaruan model yang diterima [14].

1.4 Tujuan Penelitian

1. Mengimplementasikan *Packed CKKS Homomorphic Encryption* pada *Federated Learning* untuk sistem klasifikasi kanker payudara berbasis model CNN ResNet-22, dalam pengelolaan dan perlindungan data medis untuk klasifikasi kanker payudara.
2. Menganalisis pengaruh antara tingkat privasi dan performa model (Akurasi, AUC dan PR-AUC), efisiensi komputasi, dan biaya komunikasi yang dihasilkan dari penerapan *Packed CKKS Homomorphic Encryption* dalam *Federated Learning* untuk klasifikasi kanker payudara.

1.5 Manfaat Penelitian

1. Memberikan kontribusi akademis dalam pengembangan sistem klasifikasi kanker payudara berbasis AI yang aman dan sesuai regulasi privasi.
2. Menjadi referensi praktis bagi penerapan *Federated Learning* dengan *Packed CKKS Homomorphic Encryption* pada kolaborasi multi-institusi di bidang kesehatan.
3. Mendemonstrasikan penerapan *Packed CKKS Homomorphic Encryption* dalam *Federated Learning* berbasis model CNN ResNet-22 untuk klasifikasi kanker payudara dengan tetap menjaga kualitas performa model (Akurasi, AUC, dan PR-AUC), serta memberikan analisis efisiensi waktu komputasi dan biaya komunikasi.

1.6 Sistematika Penulisan

Sistematika penulisan laporan penelitian ini disusun untuk memberikan alur pembahasan yang jelas dan terstruktur, mulai dari pendahuluan hingga simpulan. Adapun sistematika penulisannya adalah sebagai berikut:

- **Bab 1 PENDAHULUAN**
  Bab ini memuat latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan penelitian, serta sistematika penulisan.
- **Bab 2 LANDASAN TEORI**
  Bab ini berisi kajian pustaka yang mendasari penelitian, meliputi teori terkait

klasifikasi kanker payudara, *Federated Learning*, *CKKS Homomorphic Encryption* (khususnya *packed CKKS*), serta penelitian terdahulu yang relevan sebagai landasan konseptual perancangan sistem.

- **Bab 3 METODOLOGI PENELITIAN**
  Bab ini menjelaskan metodologi penelitian, meliputi pengumpulan dan pra-pemrosesan data, pembagian dataset, rancangan sistem *Federated Learning*, integrasi *Packed CKKS Homomorphic Encryption*, serta rancangan evaluasi dan pengukuran.

- **Bab 4 HASIL DAN DISKUSI**
  Bab ini menyajikan hasil implementasi dan pengujian sistem. Pembahasan mencakup analisis performa model (Akurasi, AUC dan PR–AUC), analisis efisiensi waktu komputasi (pelatihan lokal, agregasi global, enkripsi, dan dekripsi), serta analisis efisiensi komunikasi (*data transfer volume*). Bab ini juga membandingkan mode *plain* dan mode terenkripsi menggunakan *Packed CKKS* pada variasi *top-k*.

- **Bab 5 KESIMPULAN DAN SARAN**
  Bab ini memuat simpulan berdasarkan hasil penelitian serta saran untuk pengembangan penelitian selanjutnya, baik dari sisi peningkatan performa sistem maupun penguatan mekanisme privasi.