

Bab I

Pendahuluan

1.1 Latar Belakang

Penyakit kanker merupakan penyakit yang diakibatkan tak terkendalinya pertumbuhan sel dalam tubuh dan menyebar ke bagian tubuh yang lainnya. Secara normal, sel tubuh manusia tumbuh dan berkembang (melalui proses yang dinamakan pembelahan sel) untuk membentuk sel baru yang dibutuhkan oleh tubuh. Sel tubuh yang menua akan rusak dan digantikan oleh sel baru. Siklus sel tersebut suatu ketika dapat rusak, dan sel abnormal atau sel rusak dapat berkembang tidak semestinya. Sel-sel ini dapat berkembang menjadi tumor. Tumor inilah yang dapat berkembang menjadi kanker. Kanker dapat menyerang ke jaringan tubuh terdekatnya dan menyebar melalui darah dan membentuk tumor baru di dalam tubuh (proses yang disebut metastasis). Ketika kanker mulai tumbuh dan menyebar dari paru-paru, maka kanker tersebut dinamakan kanker paru-paru [1]. Penyakit kanker paru-paru merupakan salah satu jenis kanker dengan angka kematian tertinggi di dunia [2]. Di Indonesia sendiri, kasus penyakit kanker paru-paru menempati peringkat ke-3 tertinggi setelah kanker payudara dan kanker serviks [3]. Serta laki-laki sebagai pengidap tertinggi kanker paru-paru di Indonesia dengan presentase 14,1% [3]. Beberapa faktor resiko penyakit kanker paru-paru diantaranya perokok aktif maupun pasif, pajanan di lingkungan sekitar, bawaan genetik, dan penderita penyakit paru kronik. Dibandingkan dengan penyakit kanker lainnya, kanker paru-paru memiliki angka tahan hidup yang rendah [3].

Berdasarkan fakta-fakta yang telah diuraikan tersebut, tentunya pengidentifikasian dini terhadap kanker paru-paru sangat dibutuhkan. Terdapat berbagai cara pengidentifikasian tanda-tanda kanker paru-paru yang telah dilakukan oleh para peneliti. Salah satunya menggunakan *deep learning* (DL) dan *machine learning* (ML) untuk menganalisis data-data biomedikal seperti yang dilakukan oleh [4][5][6][7][8][9][10]. Dari beberapa penelitian tentang pendeteksian kanker paru-paru yang telah dilakukan, sebagian besar hanya berfokus dengan pendekatan *deep learning* dengan menggunakan *convolutional neural networks* (CNN), seperti yang dilakukan oleh

[11][12][13][14][15][16][17]. Namun dalam penelitian [16][17], performansi yang dihasilkan dari model yang dikembangkan masih di bawah 90%.

Beberapa penelitian sudah melakukan beberapa pendekatan untuk mengatasi kelemahan dari metode metode CNN, salah satunya menggunakan teknik *ensemble learning*. Salah satu penerapan *ensemble learning* dengan menggunakan algoritma *Best Fitness-based Squirrel Search Algorithm* (BF-SSA) dengan dataset yang telah diekstraksi menggunakan CNN dilakukan oleh [18] pada tahun 2022. Dalam penelitiannya, teknik *ensemble learning* yang digunakan dapat memprediksi kanker paru-paru lebih baik dibandingkan dengan kelima algoritma lainnya, namun model yang digunakan juga memiliki kelemahan, yaitu gagal dalam menyeimbangkan kemampuan algoritma untuk melakukan eksploitasi dan eksplorasi data yang berkaitan dengan optimasi kombinatorial dengan benar.

Penerapan *ensemble learning* juga dilakukan oleh [19] pada tahun 2022. Dalam penelitiannya, teknik *ensemble learning* yang diterapkan menggunakan metode *Hybrid Genetic and Support Vector Machine* (GA-SVM). Dibandingkan dengan teknik *machine learning* lain yang digunakan dalam penelitiannya, model yang diusulkan hanya memiliki akurasi sebesar 85%, dan *F1-score* sebesar 0.92.

Oleh karenanya, untuk mengatasi kelemahan dari *ensemble learning* dari penelitian sebelumnya, dalam tugas akhir ini akan dikembangkan sebuah model untuk mendeteksi kanker paru-paru menggunakan data gambar histopatologi yang menerapkan *feature extraction* dan *ensemble learning*. Untuk modelnya sendiri akan menggunakan model MobileNet. Berdasarkan penelitian sebelumnya [20], model MobileNet menghasilkan rata-rata performansi paling tinggi di antara model lain yang digunakan, yaitu VGG16, VGG19, DenseNet169, dan DenseNet201. Tujuan akhir dari tugas akhir ini adalah mengidentifikasi model dan jenis *ensemble learning* yang paling efektif dan efisien untuk mendeteksi kanker paru-paru berdasarkan data gambar histopatologi yang digunakan.

1.2 Perumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam tugas akhir ini adalah sebagai berikut:

1. Bagaimana melakukan studi algoritma untuk mendeteksi kanker paru-paru berdasarkan gambar histopatologi menggunakan *ensemble learning*?

2. Bagaimana mengembangkan model *ensemble learning* yang lebih akurat untuk mendeteksi kanker paru-paru?
3. Bagaimana menganalisis hasil performansi model yang dikembangkan?

1.3 Pernyataan Masalah

Berdasarkan latar belakang di atas, dapat disimpulkan terdapat beberapa permasalahan sebagai berikut:

1. Penelitian deteksi kanker paru-paru menggunakan *ensemble learning* masih jarang dilakukan
2. Model yang dihasilkan *ensemble learning* menghasilkan akurasi deteksi yang rendah.
3. Jenis *ensemble learning* yang digunakan menghasilkan akurasi deteksi yang rendah

1.4 Tujuan

Tujuan dari penulisan tugas akhir ini adalah sebagai berikut:

1. Membuat taksonomi studi literatur mengenai deteksi kanker paru-paru menggunakan metode *ensemble learning*
2. Mengembangkan model *ensemble learning* yang lebih akurat untuk mendeteksi kanker paru-paru.
3. Menganalisis hasil performansi metode *ensemble learning* yang telah dikembangkan untuk mendeteksi kanker paru-paru

1.5 Batasan Masalah

Berikut merupakan ruang lingkup dalam tugas akhir ini:

1. Pengujian dan pengembangan model *ensemble learning* hanya menggunakan dataset gambar histopatologi yang diambil dari [21] yang berisi 25000 gambar dengan 5 kelas. Namun hanya 3 kelas yang bersesuaian dengan topik yang diteliti, yaitu kelas *Lung benign tissue*, *Lung adenocarcinoma*, dan *Lung squamous cell carcinoma*. Sehingga total data yang diteliti menjadi 15000 gambar.
2. Teknik yang akan dibahas dalam penelitian ini hanya teknik *ensemble learning* yang secara khusus yaitu *hard voting*, *soft voting*, dan *stacking*.

1.6 Hipotesis

1. Model *ensemble learning* yang diusulkan dalam tugas akhir menghasilkan akurasi lebih dari 90% dari proses deteksi kanker paru-paru
2. Performansi model *ensemble learning* yang dikembangkan lebih akurat dibandingkan model yang telah dikembangkan sebelumnya dalam penelitian lain

1.7 Sistematika Penulisan

Tugas Akhir ini disusun dengan sistematika penulisan sebagai berikut:

- **BAB I Pendahuluan.** Bab ini membahas mengenai latar belakang, rumusan masalah, dan tujuan pengerjaan Tugas Akhir ini.
- **BAB II Kajian Pustaka.** Bab ini membahas fakta dan teori yang berkaitan dengan perancangan sistem untuk mendirikan landasan berfikir. Dengan menggunakan fakta dan teori yang dikemukakan pada bab ini penulis menganalisis kebutuhan akan rancangan arsitektur sistem yang dibangun.
- **BAB III Metodologi dan Desain Sistem.** Bab ini menjelaskan metode penelitian, rancangan sistem dan metode pengujian yang dilakukan dalam penelitian.
- **BAB IV Hasil dan Pembahasan.** Bab ini menampilkan hasil yang didapat dari penelitian ini dan melakukan analisis terhadap hasil yang didapatkan.
- **BAB V Kesimpulan dan Saran.** Bab ini membahas tentang kesimpulan yang dapat diambil dari penelitian yang dilakukan dan menyampaikan saran terkait penelitian lanjutan yang dapat dilakukan.