

BAB I

PENDAHULUAN

Bab I menjelaskan mengenai Latar Belakang, Rumusan Masalah, Tujuan dan Manfaat, Ruang Lingkup, dan Sistematika Penulisan pada penelitian yang dilakukan.

1.1 Latar Belakang

Kanker Payudara merupakan salah satu jenis kanker berbahaya dengan jumlah kematian yang tinggi di dunia. Menurut data yang terdapat pada *World Health Organization (WHO)* yang dikutip dari *American Cancer Society (2022)*, tahun 2022 terdapat 2,3 juta wanita yang didiagnosa menderita kanker payudara, dengan angka kematian mencapai 670 ribu di seluruh penjuru dunia. Indonesia sendiri menempati urutan ke-8 sebagai negara dengan tingkat kanker payudara tertinggi dengan jumlah kasus sebanyak 66 ribu dengan *ASR/100.000* sebanyak 41,8. *ASR* merupakan singkatan dari *age-standardised rates* yang merupakan ukuran ringkasan dari tingkat penyakit yang akan dimiliki oleh suatu populasi jika memiliki struktur usia standar. Standarisasi diperlukan ketika membandingkan populasi yang berbeda dalam hal usia karena usia memiliki pengaruh yang kuat terhadap risiko kematian akibat kanker. Sedangkan untuk angka kematian, Indonesia menduduki peringkat ke-4 dengan jumlah 23 ribu kasus kematian dengan *ASR/100.000* sebanyak 14,4.

Tingginya kasus kanker payudara ini menjadi permasalahan penting untuk ditemukan solusinya, salah satu solusinya adalah melakukan deteksi dini agar dapat dilakukan tindakan pertolongan pertama. Deteksi dini pada kanker merupakan kunci untuk penanganan kanker payudara yang berhasil, maka itu ketersediaan teknologi dalam melakukan *screening* yang memadai menjadi krusial untuk mendeteksi tanda awal dari kanker payudara. *Screening* untuk kondisi ini dapat dilakukan menggunakan beberapa macam pemodelan citra, beberapa model yang paling populer adalah *mammography*, *ultrasonography*, *thermography*, dan citra mikroskopis. *Mammography* merupakan pemodelan citra *x-ray* yang digunakan untuk deteksi kanker payudara pada tahap awal dan penyakit pada payudara lainnya (Nicosia dkk., 2023).

Ultrasonography merupakan model pembuatan citra yang menggunakan gelombang suara untuk membuat gambar organ, jaringan, dan struktur lainnya di dalam tubuh, sementara *thermography* menggunakan *infrared* video untuk membuat citra (*thermograms*) yang menunjukkan perbedaan suhu pada permukaan setiap bidang yang dideteksi. Citra

mikroskopis merupakan citra yang didapatkan dari sampel organ yang kemudian diteliti menggunakan mikroskop (Nicosia dkk., 2023). Salah satu contoh citra mikroskopis adalah citra histopatologi yang sering digunakan untuk deteksi apakah terdapat patologi/penyakit pada sel yang berada pada sampel organ yang diteliti. Salah satu model yang sering digunakan dalam membangun model deteksi kanker payudara adalah *Convolutional Neural Network (CNN)*.

Salah satu model pada algoritma *Deep Learning* adalah *Convolutional Neural Network (CNN)*. *CNN* merupakan salah satu kelas pada *deep neural network* dan paling sering digunakan untuk menganalisa citra visual yang pada kasus ini digunakan untuk mengidentifikasi kanker payudara berdasarkan *image segmentation*. Terdapat beberapa metode yang dapat diterapkan untuk melakukan pengklasifikasian citra visual, beberapa antaranya adalah *Elastic SCAD SVM*, tetapi pada metode tersebut memiliki kelemahan yaitu tidak dapat mendukung proses seleksi fitur. Dan metode lainnya adalah menggunakan Algoritma *Gain Ratio* memiliki tingkat akurasi yang tinggi untuk mengidentifikasi kanker payudara, tetapi untuk pemrosesannya membutuhkan waktu yang lebih lama dan perhitungan yang lebih rumit dibandingkan menggunakan *Convolutional Neural Network (CNN)* (Yap dan Chan, 2020). Maka dari itu, tujuan dilakukannya studi lebih lanjut mengenai pengklasifikasian kanker payudara menggunakan *Convolutional Neural Network (CNN)* adalah untuk membuat model berbasis *deep learning* yang lebih efisien khususnya dalam hal klasifikasi citra.

Penelitian yang dilakukan oleh (Aslam dan Cui, 2020), penelitian dilakukan dengan membandingkan 3 algoritma, yaitu *LeNet*, *AlexNet*, dan *DenseNet* menggunakan *BreakHis dataset* dan didapatkan hasil akurasi 95,4% menggunakan algoritma *DenseNet*. Penelitian yang dilakukan oleh Lim dkk. (2018) juga menggunakan *BreakHis dataset*, dengan menggunakan metode *Transfer Learning*, *Data Augmentation*, *Data Pre-Processing*, *Data-Imbalance-Under sampling*, yang diaplikasikan menggunakan algoritma *VGG-16* dan *Inception-V3* mendapatkan hasil akurasi 98% untuk algoritma *VGG-16*. Dalam penelitian ini, akan dilakukan perbandingan pada 3 algoritma *Convolutional Neural Network (CNN)*, yaitu *ResNet-50*, *DenseNet-201*, dan *VGG-16* dengan menggunakan *BreakHis dataset* untuk mencari hasil akurasi tertinggi dari *binary classification (benign dan malignant)*.

Algoritma *ResNet-50*, *DenseNet-201*, dan *VGG-16* dipilih karena memiliki tingkat akurasi yang tinggi dalam beberapa penelitian yang sudah dilakukan dan memiliki

karakteristik yang berbeda dalam hal arsitektur setiap algoritmanya, sehingga dapat ditentukan arsitektur terbaik dari 3 algoritma yang berbasis *Convolutional Neural Network (CNN)* untuk kasus klasifikasi citra histopatologi kanker payudara.

1.2 Rumusan Masalah

Rumusan masalah yang dapat disimpulkan dari permasalahan yang sudah dipaparkan pada bagian latar belakang adalah menentukan dari 3 algoritma yang dipilih (*ResNet-50*, *DenseNet-201*, dan *VGG-16*), manakah yang memiliki akurasi terbaik dan lebih efisien untuk kasus klasifikasi citra histopatologi kanker payudara dengan menggunakan model *Convolutional Neural Network (CNN)*. Citra histopatologi adalah citra mikroskopis jaringan dari suatu organ yang biasanya digunakan menjadi sampel pemeriksaan lebih lanjut, apakah jaringan tersebut normal atau patologis (jaringan dengan penyakit, seperti tumor ganas atau jinak, infeksi dan lainnya).

1.3 Tujuan dan Manfaat

Tujuan dari pembuatan program pada skripsi ini adalah mendapatkan hasil pengolahan citra yang lebih akurat (memiliki tingkat akurasi lebih tinggi) pada kasus pengolahan citra histopatologi kanker payudara (*BreakHis Dataset*) pada 400x perbesaran dengan menentukan *hyper-parameters* terbaik. *Hyper-parameters* merupakan variabel eksternal yang nilainya dapat diatur oleh mesin atau diatur secara manual. Penentuan nilai *hyper-parameters* dapat mempengaruhi nilai akurasi. *Hyper-parameters* yang nilainya diteliti adalah *epoch*, *batch size*, *learning rate*, dan *dropout*.

Epoch dalam pembelajaran mesin berarti data latih yang melewati program yang dibangun sebanyak satu kali, sehingga nilai *epoch* sering diartikan sebagai jumlah iterasi yang dilakukan pada saat mesin dijalankan. *Batch size* merepresentasikan jumlah data latih yang digunakan untuk satu langkah maju atau mundur melalui suatu jaringan. *Learning rate* menentukan seberapa penting informasi terbaru dibandingkan dengan informasi sebelumnya selama proses pembelajaran. *Learning rate* mengontrol ukuran dari parameter yang terbaharui dan mengatur kecepatan dan kestabilan pada proses optimasi. *Dropout* merupakan teknik dimana nilai *neurons* yang dipilih secara acak akan diabaikan atau menjadi 0.

Manfaat yang ingin dicapai dari pembuatan program ini adalah dapat mengolah atau mengkomputasi citra histopatologi yang lebih efektif dan efisien yang meliputi waktu

pemrosesan citra maupun tingkat akurasi yang didapat setelah dilakukan pemrosesan citra dan menentukan model terbaik dari *ResNet-50*, *DenseNet-201*, dan *VGG-16* untuk klasifikasi citra menggunakan *BreakHis dataset*.

1.4 Ruang Lingkup

Ruang lingkup digunakan sebagai acuan mengenai batasan dari pembuatan program pada skripsi ini, sehingga tidak menyimpang maupun melebihi batas pembahasan dari skripsi dan tetap sesuai dengan target yang ingin dicapai. Adapun ruang lingkup dari penulisan skripsi ini adalah sebagai berikut:

1. Melakukan klasifikasi pada citra kanker payudara dengan membangun program menggunakan model *Convolution Neural Network (CNN)*.
2. Mencari model *Convolution Neural Network (CNN)* terbaik, antara arsitektur *ResNet-50*, *DenseNet-201*, dan *VGG-16* untuk klasifikasi citra kanker payudara dengan menggunakan *hyper-parameter tuning*.
3. Menentukan *hyper-parameter tuning* terbaik. *Hyper-parameter* yang diuji adalah *epochs*, *batch size*, *learning rate*, dan *dropout*.
4. Mengolah data yang diambil dari *BreakHis dataset* perbesaran 400x dengan 2 kelas, yaitu *benign* (jinak) dan *malignant* (ganas). Jumlah data pada kelas *benign* adalah 588, sementara data pada kelas *malignant* berjumlah 1232.
5. Membangun program menggunakan model *Convolution Neural Network (CNN)* dengan Bahasa pemrograman *Python 3* dan *library Keras*.
6. Melakukan pelatihan dan pengujian program yang dibangun menggunakan platform *Google Colaboratory* di lingkungan *Jupyter Notebook*.

1.5 Sistematika Penulisan

Bagian sistematika penulisan ini menjelaskan mengenai garis besar dan gambaran isi laporan pada skripsi ini yang dibagi menjadi 5 bab. Berikut ini penjelasan secara umum mengenai setiap bab yang ada pada laporan skripsi ini.

BAB I PENDAHULUAN

Bab pertama membahas mengenai latar belakang dari penyusunan dan pemilihan topik yang akan dikaji dalam pembuatan skripsi ini, kemudian memaparkan rumusan

masalah yang didapatkan dari pemaparan latar belakang. Selanjutnya menjelaskan mengenai tujuan dan manfaat, ruang lingkup, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab kedua berisi penjelasan ilmu-ilmu yang digunakan pada pembuatan skripsi ini. Teori dasar yang berkaitan dengan penelitian yang dilakukan dipaparkan pada bab kedua ini. Beberapa teori yang dipaparkan pada bab kedua ini adalah kanker payudara, pembelajaran mesin (*machine learning*), segmentasi citra, *Convolutional Neural Network (CNN)*, *ResNet-50*, *DenseNet-201*, *VGG-16*, *batch normalization*, fungsi aktivasi, *evaluation metrics*, *loss function*, *backpropagation*, *adam optimizer*, *hyperparameter tuning*, *tools* dan *library*.

BAB III METODOLOGI PENELITIAN

Bab ketiga menjelaskan mengenai tahapan pelaksanaan penelitian yang dilakukan, mulai dari *pre-processing* data yaitu mengubah ukuran citra, menentukan persentase data latih maupun data uji, kemudian membuat model arsitektur *ResNet-50*, *DenseNet-201*, dan *VGG-16* untuk pemrosesan input, mengevaluasi hasil akurasi dari pengujian yang dilakukan, melakukan pengkajian ulang jika hasil yang didapatkan belum optimal, dan terakhir menyimpulkan hasil akhir perhitungan dari model yang dibangun.

BAB IV HASIL DAN ANALISIS

Bab keempat berisi tentang perangkat yang digunakan dalam melakukan pengkomputasian citra dan lingkungan tempat dilakukannya penelitian, kemudian terdapat pemaparan analisa hasil penelitian klasifikasi citra biopsi mikroskopik kanker payudara dan menentukan model terbaik dari program yang dibangun akan dijelaskan pada bab ke-IV ini.

BAB V PENUTUP

Bab kelima berisi kesimpulan dari keseluruhan pengerjaan skripsi berdasarkan penjelasan yang sudah dipaparkan pada bab I hingga bab IV dan juga berisi saran agar kedepannya dapat menjadi acuan dalam pengembangan model klasifikasi citra biopsi mikroskopik kanker payudara yang lebih efektif dan efisien.