

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Objek penelitian pada penelitian ini adalah jenis-jenis kanker kulit yang sudah ditentukan, yakni *melanoma*, *basal cell carcinoma*, dan *squamous cell carcinoma*. Objek ini dipilih untuk diteliti dikarenakan terkenal ketiga jenis kanker kulit ini sebagai jenis paling umum yang ada di dunia. Dari ketiga jenis kanker kulit tersebut, akan dikelompokannya menjadi *class melanoma* dan *non-melanoma* yang mana *melanoma* masuk ke dalam *class melanoma*, sedangkan *basal cell carcinoma* dan *squamous cell carcinoma* masuk ke dalam *class non-melanoma*. Dikelompokannya menjadi *class melanoma* dan *non-melanoma* atas dasar tingkat berbahayanya masing-masing jenis kanker kulit tersebut yang mana *melanoma* merupakan jenis yang paling berbahaya [1], sedangkan *class non-melanoma* yang mencakup *basal cell carcinoma* dan *squamous cell carcinoma* secara relatif tidak mematikan [2].

3.2 Metode Penelitian

3.2.1 Pengumpulan Data

Sumber data yang digunakan merupakan data sekunder di mana penelitian menggunakan data yang dikumpulkan oleh peneliti lain. Data yang digunakan berisikan 25.331 gambar dermoskopik lesi kulit, yang berasal dari beberapa penelitian sebelumnya, meliputi:

1. HAM10000 [35], yang menyediakan gambar dermoskopik lesi kulit yang dikumpulkan selama 20 tahun dari dua tempat berbeda yakni

Departemen Dermatologi di Universitas Kedokteran Vienna dan praktek kanker kulit Cliff Rosendahl di Queensland.

2. BCN20000 [36], yang menyediakan gambar dermoskopik lesi kulit yang dikumpulkan dari tahun 2010 sampai 2016 di Hospital Clínic de Barcelona.
3. MSK [37], yang menyediakan gambar dermoskopik lesi kulit yang dikumpulkan di Memorial Sloan-Kettering Cancer Center.

Berikut Tabel 3.1 merupakan penjelasan dari masing-masing atribut di dataset.

Tabel 3.1 Tabel Dataset

No	Atribut	Representasi	Keterangan	Tipe Data	Deskripsi
1	<i>Image</i>	<i>Image</i>	Gambar dermoskopik lesi kulit	nominal	teks
2	<i>Class</i>	<i>Class</i>	Terkena <i>melanoma</i> atau <i>non-melanoma</i>	nominal	(<i>melanoma/non-melanoma</i>)

3.2.2 Variabel Independen

Variabel Independen adalah variabel yang mempengaruhi variabel lain. Pada penelitian ini, yang merupakan variabel independen adalah gambar dermoskopik lesi kulit.

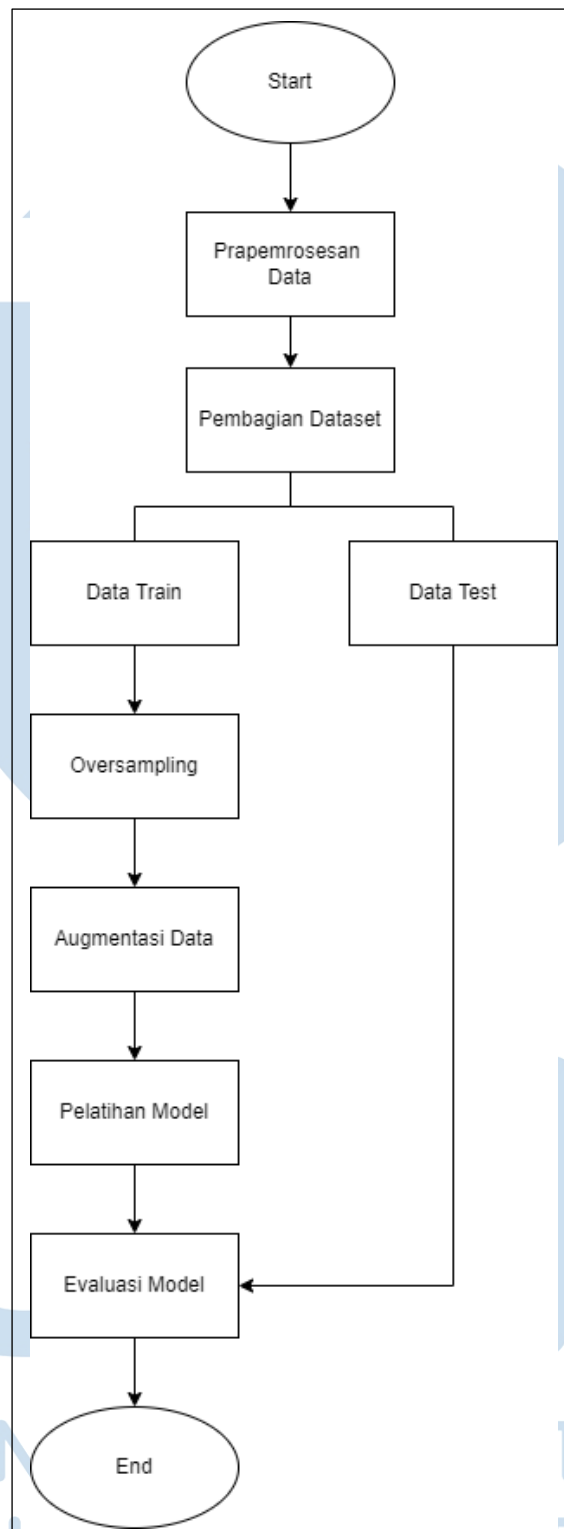
3.2.3 Variabel Dependen

Variabel Dependen adalah variabel yang dipengaruhi oleh variabel independen. Pada penelitian ini, yang merupakan variabel dependen adalah *class*, yaitu apakah pasien terkena *melanoma* atau *non-melanoma*.

3.3 Kerangka Pikir

Dalam penelitian ini, kerangka pikir yang digunakan didasarkan dari penelitian sebelumnya [38]. Gambar 3.1 merupakan *flowchart* yang akan menggambarkan alur dari penelitian.





Gambar 3.1 Diagram Kerangka Pikir

3.3.1 Prapemrosesan Data

Data yang digunakan terbagi ke dalam 8 jenis kanker kulit, yakni *melanoma*, *melanocytic nevus*, *basal cell carcinoma*, *actinic keratosis*, *benign keratosis*, *dermatofibroma*, *vascular lesion*, dan *squamous cell carcinoma*. Dikarenakan penelitian ini hanya akan melakukan klasifikasi terhadap jenis lesi *melanoma*, *basal cell carcinoma*, dan *squamous cell carcinoma* saja, maka jenis lesi yang lain tidak akan digunakan. Lalu, ketiga jenis lesi tersebut dikelompokkan menjadi *class melanoma* dan *non-melanoma* yang mana *melanoma* masuk ke dalam *class melanoma*, sedangkan *basal cell carcinoma* dan *squamous cell carcinoma* masuk ke dalam *class non-melanoma*.

Prapemrosesan gambar memiliki tujuan untuk meningkatkan parameter gambar seperti kualitas dan kejelasan. Dengan menghilangkan atau mengurangi bagian yang tidak diinginkan dari gambar. Prapemrosesan yang dilakukan pada penelitian ini terbagi menjadi empat proses sebagai berikut:

Resize merupakan salah satu teknik *downsampling*. *Downsampling* ini merupakan teknik yang digunakan untuk mengurangi ukuran citra dalam data, yang ditujukan untuk mengurangi proses komputasi [25]. Selain untuk mengurangi proses komputasi, *resize* ini digunakan untuk menyamakan seluruh data gambar, yang mana tiap gambar memiliki ukuran yang berbeda.

Pada umumnya, sebagian besar lesi kulit memiliki partikel rambut tubuh, yang dapat menjadi penghambat dalam klasifikasi jenis kanker kulit tersebut. Maka, untuk menghilangkan partikel rambut yang tidak diinginkan dari gambar, *dull razor method* digunakan. *Dull razor method* melakukan operasinya dengan menggunakan operasi morfologi *grayscale* untuk

mengenali posisi rambut pada lesi kulit [29]. Setelah posisi pixel rambut ditemukan, akan diverifikasinya bentuk atau struktur dari partikel-partikel rambut tersebut, yang lalu akan digantikannya partikel tersebut dengan menggunakan interpolasi bilinear [29]. Lalu, akan diterapkannya median filter untuk menghaluskan pixel rambut yang digantikan [29].

Saat menggunakan gambar apa adanya dalam melakukan proses pelatihan *Deep Learning*, penghitungan nilai numerik yang tinggi dapat menjadi lebih kompleks. Untuk mengurangi ini, dapat dinormalisasikannya nilai ke dalam kisaran 0 hingga 1. Dengan cara ini, angka akan menjadi kecil dan perhitungan menjadi lebih mudah dan lebih cepat. Nilai piksel dalam gambar sendiri memiliki rentang dari 0 hingga 255, sehingga dengan membagi semua nilai dengan 255 akan mengubahnya menjadi rentang dari 0 hingga 1.

3.3.2 Pembagian Dataset

Dataset akan dibagi menjadi *training*, dan *testing* data dengan proporsi masing-masing 80% dan 20% dari keseluruhan jumlah data. Setelah itu, data *training* dibagi lagi menjadi *training*, dan *validation* data dengan proporsi masing-masing 80% dan 20% dari jumlah data *training*. Pembagian sebesar 80% dan 20% ini digunakan berdasarkan apa yang telah dilakukan oleh penelitian sebelumnya [15].

3.3.3 Oversampling

Teknik *Oversampling* merupakan salah satu jenis teknik *resampling* untuk menyeimbangkan rasio antara data yang ada di kelas mayoritas dan minoritas pada *training* data. Teknik ini memilih sampel secara acak dan

menghasilkan sampel baru di kelas minoritas [39]. Akan dilakukannya *oversampling* pada *training* data sehingga data yang akan dilatih ke dalam model seimbang.

3.3.4 Augmentasi Data

Augmentasi data merupakan proses untuk meningkatkan keberagaman data yang tersedia untuk pelatihan model. Akan digunakannya *library* Keras dengan fungsi *ImageDataGenerator* untuk melakukan augmentasi data, sehingga keberagaman data yang dihasilkan dapat mengurangi kemungkinan untuk terjadinya *overfitting* pada model yang dibangun.

3.3.5 Pelatihan Model

Digunakannya *library* Keras untuk membangun model *convolutional neural network* yang akan dilatih dengan *training* data yang telah dibagi pada tahap sebelumnya. Akan dibangunnya dua macam model yakni arsitektur AlexNet dan arsitektur LeNet. Model ini dibangun untuk melakukan klasifikasi atas dua *class*. Lalu, digunakannya *library* Keras dengan fungsi *fit_generator* untuk melatih model dengan *training* data. Kemudian, untuk menjalankan method pembangunan model ini, dibutuhkan beberapa parameter seperti jumlah epoch. Jumlah *epoch* yang digunakan untuk melatih model ini adalah 50 *epochs*.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

3.3.6 Evaluasi Model

Selama proses pelatihan, akan dilakukan pemantauan terhadap akurasi dan *loss* dari data *training* dan validasi pada setiap *epoch*. *Epoch* dengan nilai akurasi validasi terbaik akan digunakan sebagai model yang akan memprediksi data *testing*. Alasan dilakukannya pemilihan atas *epoch* dengan nilai *validation* terbaik dan bukan *epoch* terakhir dikarenakan dengan membiarkan model untuk melanjutkan pelatihan yakni dengan memilih *epoch* terakhir, model akan berkinerja buruk pada data yang tidak terlihat yakni *validation* dan *test* data dikarenakan model hanya menghafal *training* data saja. *Epoch* yang berlanjut dapat menyebabkan peningkatan pada akurasi pelatihan, namun tidak berarti prediksi model pada data baru akan akurat, yang mana seringkali menjadi lebih buruk. Maka dari itu, untuk mencegah hal ini, digunakannya *validation* data untuk memantau akurasi pengujian selama pelatihan. Hal ini ditujukan untuk membuat keputusan yang lebih tepat atas model akan seberapa akuratnya pada data baru.

Evaluasi model klasifikasi gambar akan dilakukan pada data *testing* dengan metode akurasi, presisi, *recall*, dan *F1-score*. Pengukuran akurasi digunakan untuk mengetahui jumlah data yang diprediksi dengan benar dari seluruh data. Pengukuran presisi digunakan untuk mengetahui jumlah prediksi kelas positif yang benar dari seluruh prediksi kelas positif. Pengukuran *recall* digunakan untuk mengetahui jumlah prediksi kelas positif yang benar dari seluruh contoh positif. Pengukuran *F1-score* digunakan untuk mengetahui skor penyeimbang atau *harmonic mean* dari presisi dan *recall*. Rumus untuk menghitung akurasi dapat dilihat pada Rumus 3.1, untuk menghitung presisi

dapat dilihat pada Rumus 3.2, untuk menghitung *recall* dapat dilihat pada Rumus 3.3, dan untuk menghitung *F1-Score* dapat dilihat pada Rumus 3.4.

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + False\ Positive + True\ Negative + False\ Negative}$$

Rumus 3.1 Rumus Perhitungan *Accuracy*

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Rumus 3.2 Rumus Perhitungan *Precision*

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

Rumus 3.3 Rumus Perhitungan *Recall*

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Rumus 3.4 Rumus Perhitungan *F1-Score*

UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA