

BAB II

TINJAUAN PUSTAKA

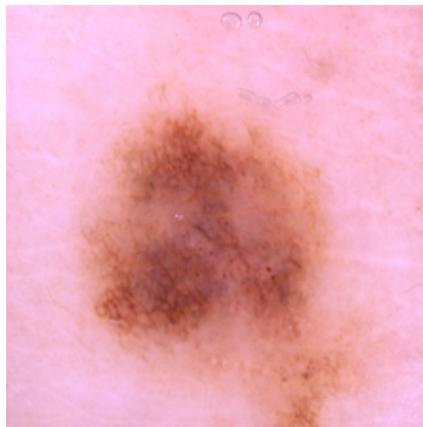
2.1 Penelitian Sebelumnya

Pada penelitian sebelumnya yang dilakukan pada tahun 2021 oleh F. Lia Dwi Cahyanti dkk., terkait Implementasi Algoritma Naïve Bayes dan K-Nearest Neighbor dalam menentukan tingkat keberhasilan *Immunotherapy*. Pengujiannya menggunakan aplikasi weka untuk mengolah data dan melakukan pengujian. Hasil pengujian menunjukkan model KNN mendapat nilai akurasi terbaik yaitu 91%. Sedangkan nilai akurasi Naïve Bayes sebesar 82% [4]. Penelitian selanjutnya dilakukan pada tahun 2020 oleh Teresia R. Savera dkk., terkait deteksi dini kanker kulit menggunakan metode KNN dan *Convolutional Neural Network* dengan menggunakan fitur ketidaksimetrisan dan ketidakteraturan tepian. Menghasilkan akurasi sebesar 75% untuk klasifikasi menggunakan KNN dan 76% menggunakan CNN [3]. Penelitian yang dilakukan pada tahun 2021 oleh Nurul Khasanah dkk., menggunakan metode Algoritma Random Forest mengatikan bahwa hasil perhitungan menggunakan ekstraksi fitur histogram menghasilkan akurasi 83%, sedangkan menggunakan ekstraksi fitur haralick menghasilkan akurasi 61%, sedangkan menggunakan ekstraksi fitur hue moments menghasilkan akurasi 84%. Penelitian tersebut menggunakan ekstraksi fitur yang lebih beragam [1]. Penelitian yang dilakukan pada tahun 2020 oleh Muhammad Faruk, Nur Nafi'iyah menggunakan metode algoritma SVM dan KNN, hasil dari klasifikasi menggunakan algoritma SVM, mendapatkan nilai akurasinya 69,85%, dan akurasi menggunakan algoritma KNN menghasilkan akurasi sebesar 67% sampai 70% [2].

2.2 Kanker Kulit

Kanker merupakan penyakit tidak menular yang ditandai dengan adanya sel atau jaringan tidak normal yang bersifat ganas. Tumbuh cepat tidak terkendali dan dapat menyebar ke tempat lain dalam tubuh penderita. Sel kanker ganas dapat menginvasi dan merusak jaringan. Penyebaran sel kanker dapat melalui pembuluh darah maupun pembuluh getah bening. Sel kanker berasal dari semua unsur yang membentuk organ, selanjutnya sel akan tumbuh dan menggandakan diri sehingga membentuk massa tumor[5]. Kulit manusia terdiri dari epidermis, dermins, jaringan subkutan, dan lapisan lainnya. Epidermis merupakan lapisan yang terbentuk

dari sel skuamosa. Pada epidermis, sel basal yang aktif akan membelah menjadi kreatinosit. Kemudian, keratinosit yang mati menjadi keratin. Diantara sel basal terdapat sel penghasil pigmen yang disebut melanosit. Kanker kulit merupakan kanker yang terjadi akibat keganasan sel basal, keratinosit dan melanosit[6]. Kanker kulit merupakan salah satu kanker yang umum terjadi. Dua tipe kanker kulit yang paling umum terjadi adalah karsinoma sel basal dan karsinoma sel skuamosa [7]. Kedua tipe ini dapat disembuhkan, tetapi akan meninggalkan bekas. Melanoma merupakan tipe kanker kulit yang ketiga dimana hal tersebut berbahaya dan banyak menyebabkan kematian[7]. Kebanyakan dari tiga tipe kanker kulit ini disebabkan oleh paparan sinar UV.



Gambar 2.1. Gambar kanker jinak

Gambar 2.1 adalah gambar kanker kulit dalam kategori jinak yang diambil dari data set *Kaggle*. Kanker kulit yang letaknya di permukaan kulit dapat didiagnosis dengan mudah karena dapat dilihat langsung dengan mata, sehingga deteksi dini menjadi lebih memungkinkan. Namun, ketidaksadaran penderita terhadap kanker kulit yang dialaminya seringkali mengakibatkan keterlambatan dalam pengobatan. Kondisi ini sangat berbahaya, terutama jika penderita terjangkit kanker kulit yang ganas. Risiko dari kondisi tersebut dapat berakibat fatal, bahkan bisa menyebabkan kematian.

2.3 Pengolahan Citra

Citra adalah representasi, kemiripan, atau imitasi dari suatu objek. Keluaran dari sebuah citra dapat berupa sistem perekaman data optik, seperti foto, data analog berupa video yang ditampilkan pada monitor televisi, atau berupa data digital yang dapat langsung disimpan pada media penyimpanan[8]. Pengolahan Citra adalah kegiatan memperbaiki kualitas citra agar mudah diinterpretasi oleh manusia

atau mesin komputer [9]. Dimana jika inputannya citra maka keluarannya juga citra akan tetapi dengan kualitas yang lebih baik daripada citra masukan [10]. Citra yang memiliki masalah seperti warna yang kurang tajam, keaburan (*blurring*), dan keberadaan *noise* memerlukan perbaikan agar informasi yang disampaikan tidak berkurang dan tetap mudah diinterpretasikan.

Citra digital umumnya berbentuk persegi panjang dengan dimensi panjang \times lebar ($M \times N$). Dengan M menyatakan jumlah baris sedangkan N menyatakan jumlah kolom pada matriks citra. Secara matematis citra digital dapat diungkapkan dalam bentuk matriks berikut pada persamaan 2.1. Masing-masing elemen dari larik matriks ini disebut elemen citra, elemen gambar atau piksel [11].

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1) & \cdots & f(0,N-1) \\ f(1,0) & f(1,1) & \cdots & f(1,N-1) \\ \vdots & \vdots & \ddots & \vdots \\ f(M-1,0) & f(M-1,1) & \cdots & f(M-1,N-1) \end{bmatrix} \quad (2.1)$$

Berikut macam-macam tipe citra :

1. Citra Berwarna

Citra berwarna merupakan citra RGB yang disimpan dalam matriks array berukuran $m \times n \times 3$ yang masing-masing menunjukkan warna *red, green, blue* pada setiap pikselnya [12].

2. Citra *Grayscale*

Citra yang nilai intensitas pikselnya berdasarkan derajat keabuan, Intensitas dari citra *grayscale* disimpan dalam 8 bit integer yang memberikan 256 kemungkinan yang mana dimulai dari level 0 sampai dengan 255 (0 untuk hitam dan 255 untuk putih dan nilai diantaranya adalah derajat keabuan) [12].

3. Citra Hitam Putih

Setiap piksel dari citra hitam putih hanya memiliki dua kemungkinan nilai yaitu 0 dan 1. Sebuah citra hitam putih dapat dianggap sebagai tipe citra khusus yang intensitas warnanya hanya berisi hitam dan putih [12].

2.4 Fitur Ekstraksi

Fitur Ekstraksi adalah proses pengambilan ciri atau fitur penting dari suatu objek dalam gambar. Proses ini bertujuan untuk mengidentifikasi daerah atau fitur signifikan pada gambar berdasarkan karakteristik intrinsiknya dan aplikasi yang relevan. Wilayah fitur yang ditargetkan bisa didefinisikan dalam lingkup global atau lokal

dan dapat dibedakan berdasarkan berbagai atribut seperti bentuk, tekstur, ukuran, intensitas, dan sifat statistik [13].

Dalam praktiknya, ekstraksi fitur dilakukan dengan menghitung jumlah titik atau piksel yang ditemukan dalam setiap pengecekan. Pengecekan ini bisa dilakukan dalam berbagai arah, seperti vertikal, horizontal, diagonal kanan, dan diagonal kiri, pada koordinat kartesian dari citra digital yang sedang dianalisis. Metode ini memungkinkan penangkapan informasi rinci dan relevan yang memfasilitasi analisis dan interpretasi citra lebih lanjut.

Berikut beberapa contoh metode fitur ekstraksi:

1. Ekstraksi Fitur Bentuk

Bentuk dari suatu objek merupakan karakter konfigurasi permukaan yang diwakili oleh garis dan kontur. Fitur bentuk bergantung pada teknik yang digunakan. Kategori tersebut berdasarkan batas (*boundary-based*) dan berdasarkan daerah (*region-based*). Teknik berdasarkan batas menggambarkan bentuk daerah dengan menggunakan karakteristik eksternal [14]. Contohnya, piksel sepanjang batas objek. Sedangkan teknik berdasarkan daerah menggambarkan bentuk wilayah dengan menggunakan karakteristik internal, contohnya adalah piksel yang berada dalam suatu wilayah.

2. Ekstraksi Fitur Tekstur

Fitur tekstur dapat digunakan untuk membedakan tekstur satu objek dengan objek lainnya. Menggunakan ciri statistik orde pertama atau ciri statistik orde dua, karakteristik orde pertama didasarkan pada karakteristik histogram citra dan umumnya digunakan untuk membedakan tekstur makrostruktur (perulangan pola lokal secara periodik) Ciri orde pertama antara lain: *mean*, *variance*, *skewness*, *kurtosis*, dan *entropy*. Untuk karakteristik orde kedua didasarkan pada probabilitas hubungan ketetanggaan antara dua piksel pada jarak dan orientasi sudut tertentu. Ciri orde dua umumnya digunakan untuk membedakan tekstur mikrostruktur, diantaranya : *Angular Second Moment*, *Contrast*, *Correlation*, *Variance*, *Inverse Different Moment*, dan *Entropy*.

Teknik statistik lain yang dapat digunakan untuk menentukan tekstur adalah matriks *gray level co-occurrence* (GLCM) [15]. Teknik tersebut dilakukan dengan melakukan pemindaian untuk mencari jejak derajat keabuan setiap dua buah piksel yang dipisahkan dengan jarak dan sudut yang tetap.

3. Ekstraksi Fitur Warna

Pada ekstraksi fitur warna yang dijadikan ciri pembeda adalah warna pada setiap gambar menggunakan citra yang memiliki komposisi warna RGB

(*Red, Green, Blue*). Bekerja dengan membandingkan komponen warna setiap piksel citra dengan warna yang spesifik. Hasil perbandingannya, apabila warna sesuai dengan warna spesifik maka komponen warna piksel akan dibiarkan saja, namun apabila warnanya tidak sesuai dengan warna spesifik, maka komponen warna piksel tersebut diubah menjadi warna *background*. Warna yang digunakan dalam segmentasi warna (*color filtering*) dapat direpresentasikan dalam berbagai ruang warna. Ada beberapa ruang warna yang dikenal, antara lain RGB, HSV (*Hue, Saturation, Value*), YCbCr, dan sebagainya [16].

2.5 Metode Klasifikasi

Klasifikasi merupakan proses pengelompokan objek yang memiliki karakteristik atau ciri yang sama ke dalam beberapa kelas [17]. Klasifikasi merupakan suatu pekerjaan yang melakukan penilaian terhadap suatu obyek data untuk masuk dalam suatu kelas tertentu dari sejumlah kelas yang tersedia. Tujuan dari klasifikasi adalah agar praktisi data dapat menentukan kelas atau kategori data baru berdasarkan karakteristik data yang sudah ada.

Berikut ini beberapa contoh metode-metode klasifikasi yang sering digunakan :

1. Naive Bayes

Naive Bayes merupakan metode klasifikasi yang berakar pada teorema Bayes. Yang memiliki konsep dasar peluang bersyarat, memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya [18].

2. *K-Nearest Neighbor* (KNN)

KNN adalah algoritma yang melakukan klasifikasi suatu data dengan cara mencari tetangga terdekat dari data tersebut [?].

3. *Support Vector Machine* (SVM)

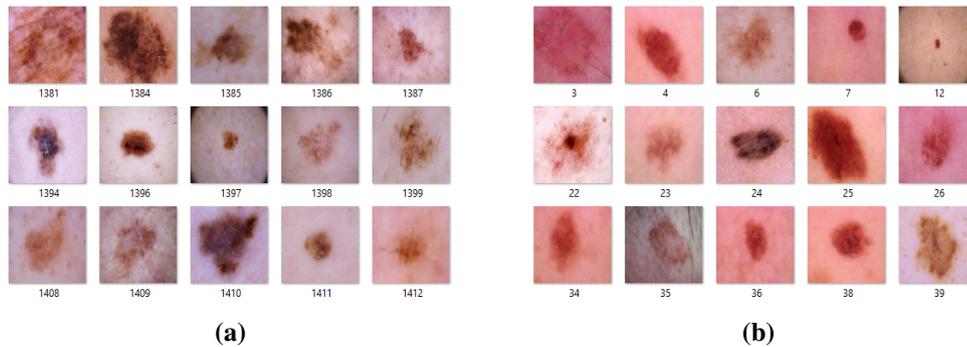
Algoritma ini melakukan klasifikasi dengan mencari hyperplane terbaik yang dapat memisahkan data dalam ruang fitur menjadi dua kelas yang berbeda secara optimal [18].

4. *Random Forest*

klasifikasi dengan menggabungkan beberapa pohon keputusan (*Decision Tree*) yang dibangun pada dataset yang berbeda-beda. Dengan menggabungkan beberapa pohon keputusan, *Random Forest* dapat mengurangi *overfitting* pada model dan meningkatkan kemampuan dalam menggeneralisasi data yang belum pernah dilihat sebelumnya [1].

2.6 Dataset Kanker Kulit

Merupakan dataset yang berisi gambar kanker kulit ganas dan jinak yang akan digunakan pada penelitian ini. Data tersebut diambil dari situs web *Kaggle*. Berikut contoh data set yang akan dipakai.

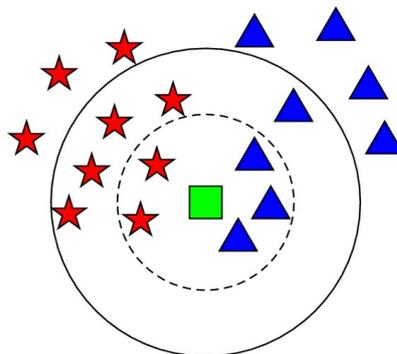


Gambar 2.2. Jenis-jenis kanker (a) Kanker Ganas (b) Kanker Jinak

Pada Gambar 2.2 (a) menunjukkan bahwa gambar tersebut merupakan kumpulan dataset kanker ganas, dan pada Gambar 2.2 (b) menunjukkan bahwa gambar tersebut merupakan kumpulan dataset kanker jinak.

2.7 *K-Nearest Neighbors* (KNN)

Algoritma *K-Nearest Neighbor* (KNN) adalah sebuah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasi sebelumnya. Algoritma ini bekerja dengan berdasarkan pada jarak terdekat dari sampel uji ke sampel latih untuk menentukan nilai KNN [19].



Gambar 2.3. Pemisahan data menggunakan KNN

Pada Gambar 2.3 ditunjukkan proses kerja KNN dalam melakukan klasifikasi berdasarkan tetangga terdekat. Warna merah dan biru menunjukkan dua data

berbeda yang tersebar, dan warna hijau menunjukkan titik data yang ingin diklasifikasi. Lingkaran-lingkaran menunjukkan jarak dari titik yang ingin diklasifikasikan (titik hijau). Pada gambar tersebut diasumsikan bahwa $k=3$, sehingga dari titik hijau akan dihitung tiga tetangga terdekat. Dalam gambar, titik hijau dikelilingi oleh lebih banyak segitiga biru di dalam lingkaran yang lebih dekat dibandingkan dengan bintang merah, yang menunjukkan bahwa titik hijau kemungkinan besar akan diklasifikasikan sebagai kelas biru.

Pada algoritma KNN, data dengan nilai tertentu dapat dihitung jaraknya dari data tersebut ke data yang lain, nilai jarak ini yang akan digunakan sebagai nilai kedekatan atau kemiripan antara data uji dengan data latih. Sedangkan dekat atau jauhnya tetangga dapat dihitung berdasarkan *Euclidean Distance*, berikut persamaannya :

$$D(a, b) = \sqrt{\sum_{k=1}^d (x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (2.2)$$

Dimana nilai $D(a, b)$ adalah jarak skalar dari dua buah vector data a dan b . Berikut langkah-langkah untuk menghitung metode *K-Nearest Neighbor* antara lain:

1. Menentukan parameter k (Jumlah tetangga paling dekat).
2. Menghitung kuadrat jarak *Euclidean* objek terhadap data *training* yang diberikan.
3. Mengurutkan hasil dari langkah dua secara *ascending* (dari yang terkecil).
4. Mengumpulkan kategori Y (Klasifikasi nearest neighbor berdasarkan nilai k).
5. Dengan menggunakan kategori *nearest neighbor* yang paling mayoritas maka dapat diprediksi kategori objek yang paling mendekati.

2.8 Ekstraksi Fitur Tekstur Berbasis Histogram

Menentukan ekstraksi citra menggunakan histogram merupakan bagian dari metode statistik orde pertama yang mana metode pengambilan ciri yang didasarkan pada karakteristik histogram citra. Histogram citra merupakan grafik yang menggambarkan penyebaran nilai-nilai intensitas piksel dari suatu citra. Dari nilai-nilai pada histogram yang dihasilkan dapat dihitung beberapa parameter ciri seperti *mean*, *variance*, *skewness*, *kurtosis*, *energy*, *entropi*, dan *smoothness* untuk menggambarkan tekstur dari citra tersebut [20].

2.8.1 Mean

Merupakan nilai rata-rata dari derajat keabuan seluruh piksel. Dalam citra Fitur pertama yang dihitung secara statis adalah rerata intensitas. Komponen fitur ini dihitung berdasarkan Persamaan 2.3.

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N i \quad (2.3)$$

dimana nilai \bar{x} diartikan sebagai nilai rata-rata, dan i adalah aras keabuan pada citra x dan N menyatakan nilai aras keabuan tertinggi[20].

2.8.2 Variance

Standar deviasi atau disebut juga *Variance*, fitur ini berguna untuk memberikan ukuran kekontrasan pada citra. Komponen fitur ini dihitung berdasarkan Persamaan 2.4.

$$\sigma = \frac{1}{N} \sqrt{\sum_{i=1}^N (i - \bar{x})^2 x(i)} \quad (2.4)$$

dimana σ merupakan simbol dari *Variance*, dan i adalah aras keabuan pada citra x dan $x(i)$ menyatakan probabilitas kemunculan dan N menyatakan nilai aras keabuan tertinggi.

2.8.3 Skewness

Skewness erupakan ukuran ketidaksimetrisan terhadap rerata intensitas atau *mean*. Pada citra dituliskan seperti pada Persamaan 2.5.

$$Skewness = \frac{1}{\sigma^3} \sum_{i=1}^N (i - \bar{x})^3 x(i) \quad (2.5)$$

Dengan demikian i adalah aras keabuan pada citra, \bar{x} merupakan rerata intensitas, $x(i)$ menyatakan probabilitas kemunculan dan N menyatakan nilai aras keabuan tertinggi.

2.8.4 Kurtosis

Kurtosis merupakan ukuran statistik yang menggambarkan seberapa tajam atau landai distribusi intensitas piksel dalam citra. Secara khusus, kurtosis mengukur

seberapa jauh ekor distribusi piksel dari nilai rata-rata (*mean*) distribusi tersebut. Dituliskan pada Persamaan 2.6.

$$\text{Kurtosis} = \frac{1}{\sigma^4} \sum_{i=1}^N (i - \bar{x})^4 \cdot x(i) \quad (2.6)$$

Dengan N merupakan jumlah tingkat intensitas piksel yang berbeda dalam citra, \bar{x} merupakan rerata intensitas, $x(i)$ menyatakan probabilitas kemunculan dan σ merupakan nilai *variance*.

2.8.5 Energy

Energy atau bisa disebut juga keseragaman, untuk menentukan Ukuran yang menyatakan distribusi intensitas piksel terhadap jangkauan aras keabuan. Perhitungannya ditunjukkan pada Persamaan 2.7.

$$\text{Energy} = \sum_{i=1}^N [x(i)]^2 \quad (2.7)$$

Dengan demikian i adalah aras keabuan pada citra, $x(i)$ menyatakan probabilitas kemunculan dan N menyatakan nilai aras keabuan tertinggi.

2.8.6 Entropi

Entropi digunakan untuk mengindikasi kompleksitas citra. Dituliskan seperti pada Persamaan 2.8.

$$\text{Entropi} = - \sum_{i=1}^N x(i) \log_2(x(i)) \quad (2.8)$$

Dengan demikian i adalah aras keabuan pada citra, $x(i)$ menyatakan probabilitas kemunculan dan N menyatakan nilai aras keabuan tertinggi.

2.8.7 Smoothness

Untuk mengukur tingkat kehalusan atau kekerasan intensitas pada citra. Dituliskan seperti pada Persamaan 2.9.

$$R = 1 - \frac{1}{1 + \sigma^2} \quad (2.9)$$

dimana nilai R adalah nilai dari *Smoothness* (tingkat kehalusan atau kekasaran) pada citra, sedangkan σ merupakan nilai *variance*.