

IDENTIFIKASI KEGANASAN TUMOR KULIT PADA CITRA DERMOSKOPI DENGAN METODE SUPPORT VECTOR MACHINE

Ikram Yunizar¹, Didit Widiyanto², Noor Falih³,
Program Studi Informatika / Fakultas Ilmu Komputer
Universitas Pembangunan Nasional Veteran Jakarta
Jl. Rs. Fatmawati Pondok Labu, Jakarta Selatan 12450
Ikramyunizar@gmail.com

Abstrak. Kanker kulit adalah salah satu kanker yang paling umum ditemukan. Perlu adanya metode diagnosa yang dapat mendiagnosa keganasan tumor kulit dengan cepat. Penelitian ini bertujuan untuk membuat prototipe sistem yang dapat membantu dokter dalam mengklasifikasikan keganasan tumor kulit dari citra dermoskopi dan mengevaluasi kinerja Support Vector Machine dalam mengklasifikasikan citra tersebut. Citra akan diolah melalui praproses, segmentasi, ekstraksi ciri, dan klasifikasi. Praproses pada citra antara lain mengubah resolusi citra, mengubah citra dari channel warna RGB menjadi grayscale dan HSV. Segmentasi dilakukan dengan metode thresholding. Metode ekstraksi ciri yang digunakan dari masing-masing adalah GLCM untuk ciri tekstur dan Color Moments untuk ciri warna. Hasil akurasi terbaik yang didapat pada tahap klasifikasi dan evaluasi dengan K-Fold Cross Validation adalah sebesar 84,8% yang didapatkan dengan model SVM kernel Radial Basis Function dengan parameter cost sebesar 1 dan gamma sebesar 0.125. Nilai akurasi yang didapatkan model dalam mengklasifikasikan citra data testing adalah 76,9%.

Kata Kunci : SVM , Dermoskopi, Tumor Kulit.

1 Pendahuluan

Salah satu bagian tubuh manusia yang dapat terkena tumor adalah kulit. Laporan American Cancer Society (2018) menyebutkan bahwa kanker kulit adalah salah satu kanker yang paling umum teridentifikasi di Amerika Serikat. Ada beberapa tindakan yang dapat dilakukan untuk mengidentifikasi sebuah tanda pada kulit tergolong kanker atau tumor, seperti pemeriksaan fisik secara visual dan biopsi. Pemeriksaan fisik dapat dilakukan dengan mata telanjang maupun dengan bantuan alat, yaitu *dermoscope*. Hasil pemeriksaan fisik sangat dipengaruhi oleh keahlian dan pengalaman dokter yang menangani, sehingga menghasilkan diagnosa yang subjektif. Biopsi dilakukan dengan cara mengambil sampel sel tumor melalui pembedahan pada daerah yang terjangkau dan pemeriksaan laboratorium. Dalam pelaksanaannya, biopsi memiliki tingkat akurasi untuk mendiagnosa sel kanker yang lebih tinggi dibandingkan dengan pemeriksaan visual. Namun, biopsi memerlukan banyak sumber daya yang diperlukan untuk mengidentifikasi suatu sel tergolong sel tumor atau kanker. Oleh karena itu, diperlukan instrument diagnosa yang dapat meminimalisir subjektifitas dari diagnose tumor kulit yang dapat dijadikan faktor pertimbangan dalam diagnose akhir keganasan tumor kulit yang diderita.

Eltayel (2017) melakukan penelitian dalam mendeteksi kanker kulit. Penelitian dilakukan menggunakan *Artificial Neural Network* untuk klasifikasi dan *Particle Swarm Optimizarion* untuk segmentasinya. Evaluasi dilakukan dengan K-Fold validation dan menghasilkan akurasi sebesar 95,97%.

Penelitian ini akan mencoba untuk menggunakan salah satu metode *supervised learning*, yaitu *Support Vector Machine* untuk mengklasifikasikan jenis tumor kulit yang diderita.

2 Tinjauan Pustaka

Adapun pengantar materi yang diperlukan untuk penelitian ini, yaitu :

4.1 Tumor Kulit

Sel-sel yang ada pada tubuh manusia umumnya memiliki siklus hidup. Sel-sel akan tumbuh dan muncul ketika sel-sel yang ada menua dan mati. Dalam beberapa kesempatan, sel-sel akan terus tumbuh dan tidak mati sehingga sel akan terus hidup dan tidak digantikan oleh sel yang baru. Sel-sel lama akan terus hidup bersamaan dengan sel baru sehingga membentuk gumpalan pada tempat dimana proses ini terjadi. Gumpalan dari sel-sel yang tidak dapat mati ini dinamakan tumor (National Cancer Institute, 2012). Tumor dapat tumbuh pada seluruh organ tubuh yang dimiliki manusia (Saleh, 2016). Penyebab tumor dapat diklasifikasikan menjadi tiga faktor utama yaitu faktor genetik, karsinogenik, dan co-karsinogen (Saleh, 2016). Faktor genetik didasari pada faktor keturunan, faktor karsinogenik didasari pada eksternal seperti radiasi, dan faktor co-karsinogen disebabkan oleh pola hidup yang salah. Tumor dapat diklasifikasikan menjadi dua jenis berdasarkan keganasan yang dimilikinya yaitu tumor jinak dan tumor ganas (Sinha, 2018) . Adapun perbedaan dari tumor jinak dan tumor ganas yang dideskripsikan pada tabel 1

Tabel 15 : Perbedaan tumor jinak dan tumor ganas

Karakteristik	Tumor Jinak	Tumor Ganas
Pertumbuhan	Tidak Tumbuh	Tumbuh Tak Teratur
Pendarahan	Tidak Ada	Ada
Lokasi Kemunculan	Dimana saja	Umumnya kulit yang terkena sinar matahari
Bentuk	Simetris	Asimetris
Warna	Warna Seragam	Beragam warna
Lama Kemunculan	Ada selama bertahun-tahun	Baru

4.2 Gray-level Co-occurrence Matrix

GLCM adalah metode yang dapat memberikan ekstraksi fitur berupa tekstur. GLCM merupakan suatu matriks yang menggambarkan frekuensi munculnya pasangan dua pixel dengan intensitas tertentu dalam jarak dan arah tertentu pada citra (Prasetyo, 2011). Pikel-pikel yang ada pada citra memiliki jarak dan orientasi sudut yang direpresentasikan dengan tingkat derajat. Orientasi sudut GLCM didasari pada empat arah sudut yaitu 0° , 45° , 90° , dan 135° (Surya, et al., 2017). GLCM dapat menghasilkan beberapa nilai yang dapat menggambarkan nilai tekstur dari citra yang diolah, nilai nilai itu antara lain:

1. Contrast

Contrast adalah nilai atau ukuran perbedaan antar derajat keabuan suatu daerah pada citra. Rumus yang digunakan untuk menghitung nilai Contrast adalah

$$Contrast = \sum_{i,j=0} (i - j)^2 p_{i,j} \quad (2.1)$$

2. Homogeneity

Nilai Homogenitas menunjukkan keketatan distribusi dari elemen yang ada pada GLCM. Rumus untuk menghitung GLCM adalah sebagai berikut:

$$Homogeneity: \sum_{i1} \frac{p_{i,j}}{1 + |i - j|^2} \quad (2.2)$$

3. Energy

Energy adalah nilai yang merepresentasikan ukuran keseragaman pada citra. Semakin tinggi kemiripan antar elemen citra maka akan semakin tinggi nilai *Energy*

$$Energy: \sum_{i1} \sum_{i2} p^2(i_1 i_2) \quad (2.3)$$

4. Correlation

Correlation adalah nilai yang merepresentasikan keterkaitan Linear dari derajat citra keabuan. Nilai dari correlation ada pada *range* -1 hingga 1. Adapun rumus yang digunakan untuk menghitung nilai *correlation* yaitu

$$Correlation = \sum_{i1} \sum_{i2} \frac{(i - \mu_i)(j - \mu_j) p_{(i,j)}}{\sigma_i \sigma_j} \quad (2.4)$$

4.3 Color Moments

Color moments adalah suatu metode yang digunakan untuk membedakan citra berdasarkan fitur warnanya (Sukafona , 2018). Color moments cukup efektif untuk ekstraksi ciri warna karena metode ini didasarkan pada fitur distribusi probabilitas warna (Kumara, 2013).Color moments disebut compact karena dapat memampatkan informasi warna menjadi beberapa nilai (Sukafona , 2018) Adapun rumus untuk menghitung Mean dan Standar Deviasi dari warna citra dengan rumus berikut (Eltayel, et al., 2017).

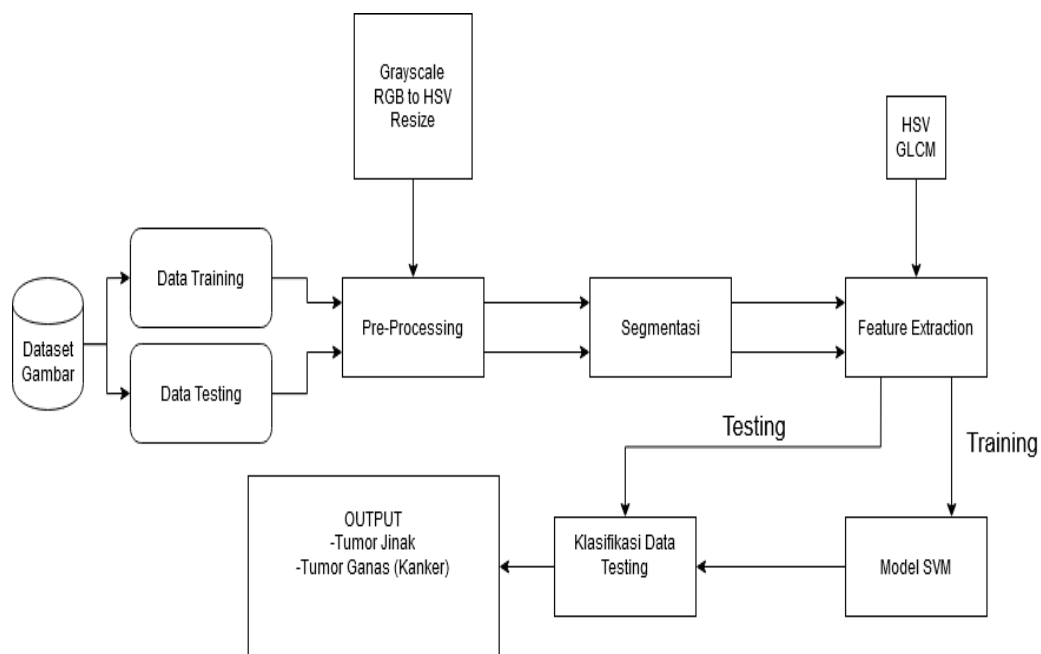
$$\mu = \frac{1}{N} \sum_{i=1}^N P_i \tag{2.5}$$

Dan berikut adalah rumus untuk menghitung standar deviasi

$$\sigma = \sqrt{\left(\frac{1}{N} \sum_{i=1}^N (P_i - \mu_i)^2\right)} \tag{2.6}$$

2 Alur Penelitian

Dalam penelitian ini, ada empat tahapan utama yang akan dijalani, yaitu tahap praproses, segmentasi, ekstraksi ciri dan klasifikasi. Keempat alur ini akan dijalankan sesuai urutannya masing-masing dan dijalankan guna mendapatkan hasil terbaik dari model yang akan digunakan, yaitu *Support Vector Machine*. Adapun gambar 1 dibawah ini akan menggambarkan alur penelitian yang akan dijalani.



Gambar 108 : Diagram alur penelitian.

4.1 Dataset

Dataset yang digunakan pada penelitian ini adalah data sekunder yang didapatkan melalui beberapa sumber. Sumber data yang digunakan adalah dataset *International Skin Imaging Collaboration* (ISIC) yang terdiri dari beberapa database milik rumah sakit yang ada di Italia, Amerika Serikat, dan tim pengelola ISIC archive yang bersifat *open-access*. Data dikumpulkan selama 20 tahun dari tahun 1997 hingga 2017. Data yang didapatkan diolah lebih lanjut secara manual untuk menghilangkan informasi yang dapat membocorkan informasi sensitif dari pasien yang didiagnosa. Selanjutnya, histogram dari masing-masing citra diolah secara manual guna memberikan pencahayaan yang seragam dan menghilangkan noise-noise yang ada pada citra. Kelas dari masing-masing citra ditentukan dengan pemeriksaan rutin selama 1,5 tahun dan *histopathology* atau pembedahan untuk selanjutnya diproses dalam laboratorium untuk mengkonfirmasi keganasan tumor yang diderita oleh pasien. Setelah data diolah, data kemudian diperiksa oleh komite etik dari universitas Vienna di Italia dan Universitas Queensland di Australia agar data yang diajukan sesuai dengan kode-kode etik dalam bidang kedokteran. Jumlah data yang diambil berjumlah 1472 citra.

4.2 Praproses

Format awal dari channel warna citra adalah RGB, sehingga perlu dilakukan praproses terlebih dahulu. Tahapan praproses dilakukan agar citra yang diolah dapat diambil ekstraksi ciri dalam bentuk tekstur dan ekstraksi ciri warnanya. Citra yang telah didapat akan menjalani beberapa proses, yaitu *Grayscale*, *RGB to HSV* dan *Resize*.

4.3 Segmentasi

Dalam tahap ini, Citra yang sudah dilakukan proses grayscale akan diolah dengan metode Otsu Thresholding dan berbagai operasi morfologi untuk mendapatkan *Region of Interest (ROI)* pada citra yaitu tumor kulit yang ada pada citra. Otsu Thresholding akan digunakan untuk segmentasi dan operasi morfologi *opening*, *closing*, *clear border*, *region filling*, dan *area filter* untuk penyempurnaan mask.

4.4 Ekstraksi Ciri

Metode ekstraksi fitur yang digunakan pada penelitian ini adalah GLCM dan HSV. GLCM digunakan untuk mencari nilai *Energy*, *Homogeneity*, *Correlation* dan *Contrast* dari masing-masing citra *training* yang akan digunakan pada proses klasifikasi. HSV digunakan untuk memberikan nilai dan ciri warna masing-masing citra sesuai dengan kelasnya. Model warna HSV dipilih karena HSV dapat memberikan hasil lebih baik dalam mengolah citra dermoskopi. Citra yang diolah untuk mendapatkan ciri warna adalah citra dermoskopi yang memiliki model warna HSV, sedangkan citra yang diolah untuk mendapatkan ciri tekstur dengan GLCM adalah citra *Grayscale*. Nilai-nilai yang didapat dari ekstraksi ciri akan digunakan sebagai input atau masukan ke klasifikasi *Support Vector Machine*.

4.5 Klasifikasi

Proses ekstraksi fitur akan menghasilkan beberapa nilai yang dapat digunakan untuk mengklasifikasikan citra. Klasifikasi akan dilakukan dengan metode *Support Vector Machine*. *Tuning parameter* akan dilakukan pada metode SVM guna mendapatkan *Hyperparameter* yang dapat memberikan performa terbaik pada masing masing *kernel* yang ada pada *Support Vector Machine* Perbandingan ini dilakukan untuk mengetahui apakah *Hyperplane* yang ada dalam metode *Support Vector Machine* akan berpengaruh secara signifikan pada hasil akurasi, spesifisitas, dan sensitivitas. Adapun langkah-langkah dari klasifikasi yang dilakukan yaitu:

2.5.1 Parameter Tuning

Parameter tuning dilakukan untuk mendapatkan nilai *hyperparameter* yang optimal pada tiap-tiap *kernel* yang ada pada SVM. Nilai yang didapatkan dari parameter tuning adalah nilai *Cost* dan nilai *Gamma* yang optimal untuk masing-masing *kernel* SVM. Nilai nilai ini akan digunakan untuk mengklasifikasikan *data training*.

2.5.2 K-Fold Cross Validation

K-Fold Cross Validation digunakan untuk membagi *data training* menjadi beberapa bagian yang akan digunakan untuk mencari model terbaik untuk mengklasifikasikan *data training*. Model dengan hasil terbaik akan digunakan untuk mengklasifikasikan *data testing* yang telah dibuat pada proses awal

4.6 Evaluasi

Pada tahap ini, model akan dievaluasi untuk menentukan tingkat kebenaran dari kelas yang telah dihasilkan pada *data testing*. Evaluasi dilakukan dengan menghitung nilai dari spesifisitas, sensitivitas, dan akurasi. *Confusion matrix* digunakan untuk menghitung seberapa baik sistem dalam mengidentifikasi tingkat keganasan tumor kulit yang diderita pasien. Adapun *Confusion matrix* dideskripsikan pada tabel 3.1 :

Tabel 16 : Tabel *Confusion matrix*

		Nilai Sebenarnya	
		True	False
Nilai Prediksi	True	TP	FP
	False	FN	TN

TP : *True Positive*

FP : *False Positive*

FN : *False Negative*

TN : *True Negative*

Adapun rumus dan fungsi dari akurasi, spesifisitas, dan sensitivitas antara lain:

1. Akurasi : nilai keberhasilan sistem dalam mengidentifikasi tumor jinak dan tumor ganas pada kulit sesuai dengan kelas masing-masing.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.1)$$

2. Spesifisitas : Nilai keberhasilan sistem dalam mengidentifikasi tumor jinak dari semua klasifikasi.

$$Spesifisitas = \frac{TN}{TN + FP} \quad (3.2)$$

3. Sensitivitas : Nilai keberhasilan sistem dalam mengidentifikasi tumor ganas dari semua klasifikasi.

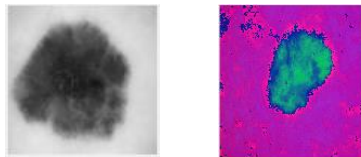
$$Sensitivitas = \frac{TP}{TP + FN} \quad (3.3)$$

4 Hasil dan Pembahasan

Berikut merupakan hasil dan pembahasan dari tiap proses yang dilakukan, antara lain :

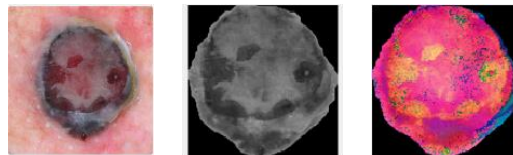
4.1 Praproses dan segmentasi

Praproses menghasilkan citra dengan resolusi 400 x 400 dan citra dalam bentuk *grayscale* dan HSV dari masing-masing citra. Selanjutnya, citra akan disegmentasi untuk dicari objeknya. Objek pada citra memiliki ukuran yang berbeda-beda sehingga perlu dilakukan *resize* ulang guna memberikan hasil ekstraksi ciri yang seragam. Gambar 2 menggambarkan hasil akhir dari praproses.



Gambar 109 : Contoh hasil praproses

Setelah praproses selesai dilakukan, maka citra dapat segmentasi untuk dicari objek tumor kulit pada citranya. Segmentasi dilakukan menggunakan *otsu thresholding*, dan disempurnakan dengan beberapa operasi morfologi seperti *Closing*, *Opening*, *Clear Border*, *Imfill*, dan *bwareafilt*



Gambar 110 : contoh hasil segmentasi dan citra yang akan diolah lebih lanjut

Setelah citra selesai disegmentasi, maka tahapan selanjutnya adalah tahapan pembagian data.

4.2 Pembagian Data.

Dataset dibagi menjadi data *training* dan data *testing*. Dari 1472 data yang diolah, proses praproses dan segmentasi menyisihkan 89 citra dengan hasil segmentasi yang buruk. Kemudian dataset dibagi menjadi data *training* dan data *testing*. Data *testing* berjumlah 130 citra dengan 65 citra tergolong dalam kelas tumor jinak dan tumor ganas, sedangkan 1253 citra dengan jumlah kelas masing-masing 653 citra kelas jinak dan 618 citra kelas ganas. Maka pada tahapan selanjutnya akan dilakukan *balancing* data dengan *Random Undersampling*

4.3 Random Undersampling

Random Undersampling adalah salah satu metode untuk melakukan *balancing data*. *Balancing Data* dilakukan untuk menghasilkan model yang tidak terpengaruh pada *balance* pada masing-masing kelas data *training*. Cara kerja *random undersampling* adalah dengan cara mengambil beberapa citra pada kelas yang memiliki data yang lebih banyak dan tidak menyertakannya pada proses *training*. Metode *Random Undersampling* dipilih karena selisih jumlah data kelas ganas dan kelas jinak tidak terlalu besar, yaitu 35 citra, sehingga efek dari *undersampling* tidak begitu besar dan menghasilkan dataset dengan data-data yang unik, tidak seperti *oversampling* yang memberikan data yang identik dalam

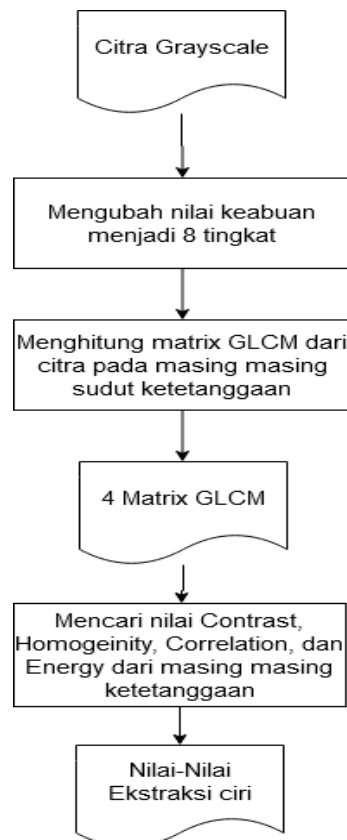
beberapa citra. Dari 1253 citra training yang ada, ada 609 citra yang tergolong ke dalam tumor ganas dan 644 citra yang tergolong ke dalam kelas tumor jinak. *Random Undersampling* dilakukan pada citra yang tergolong pada kelas jinak dengan menyisihkan secara acak 35 citra dari 644 citra yang ada, sehingga menghasilkan 609 citra tumor jinak, jumlah citra yang sama dengan tumor ganas. Jumlah citra yang termasuk ke dalam data training menjadi berjumlah 1218 citra. Setelah *random undersampling* dilakukan, maka tahapan selanjutnya adalah mengekstraksi ciri dari data yang telah dipraproses dan disegmentasi.

4.4 Ekstraksi Ciri

Citra yang telah dipraproses dan disegmentasi selanjutnya akan diolah dengan GLCM untuk mendapatkan ciri tekstur citra dan mengubah *colorspace* citra menjadi HSV untuk mendapatkan ciri warna dari citra.

4.5 Ciri Tekstur

Operasi GLCM dilakukan untuk menghitung nilai ciri tekstur dari citra. Fungsi *graycomatrix()* MATLAB digunakan untuk mendapatkan matriks GLCM, dan fungsi *graycoprops()* digunakan untuk menghitung nilai ciri yang didapatkan dari matriks GLCM. Adapun tahapan tahapan yang akan dilakukan sebagai berikut.



Gambar 111 : Proses-proses GLCM

Setelah dijalankan keempat tahapan diatas, akan didapatkan nilai *Energy*, *Correlation*, *Contrast* dan *Homogeneity* dari citra. Berikut adalah nilai ekstraksi ciri tekstur pada citra Ganas001.

Tabel 17 : Hasil Ekstraksi Ciri Tekstur Ganas001

Fitur	Nilai
Contrast 0°	0.2308
Contrast 45°	0.3734
Contrast 90°	0.2549
Contrast 135°	0.3478
Correlation 0°	0.9460
Correlation 45°	0.9122
Correlation 90°	0.9404
Correlation 135°	0.9182
Energy 0°	0.2356
Energy 45°	0.2221
Energy 90°	0.2353
Energy 135°	0.2236
Homogeneity 0°	0.9367
Homogeneity 45°	0.9170
Homogeneity 90°	0.9356
Homogeneity 135°	0.9185

4.6 Ciri warna

Ciri warna didapatkan dengan cara menghitung nilai dari mean dan standar deviasi dari masing-masing channel pada HSV yaitu *Hue*, *Saturation*, dan *Value*. Gambar HSV yang telah tersegmentasi akan dipisahkan masing-masing channelnya dan dihitung nilai mean dan standar deviasinya pada masing-masing channel warna. Citra yang diolah adalah citra HSV yang telah tersegmentasi. Nilai piksel yang dihitung adalah nilai piksel yang memiliki nilai lebih dari 0, sehingga jumlah piksel yang diolah akan sedikit berbeda-beda tergantung dari hasil *cropping* yang dilakukan. Berikut adalah contoh hasil ekstraksi ciri warna dari citra Ganas001.

Tabel 18 : Hasil Ekstraksi warna Ganas001

Fitur	Nilai
Hue Mean	0.4042
Saturation Mean	0.2505
Value Mean	0.4922
Hue Standard Deviation	0.4131
Saturation Standard Deviation	0.1185
Value Standard Deviation	0.0065

Setelah ekstraksi ciri masing-masing citra sudah didapatkan, maka tahapan selanjutnya adalah Klasifikasi.

4.7 Klasifikasi

Tahapan-tahapan yang ada pada tahapan klasifikasi antara lain parameter tuning, Evaluasi model, dan testing

4.8 Parameter Tuning

Kernel-kernel pada SVM memiliki parameter parameter yang dapat diatur untuk mendapatkan hasil yang lebih baik dan lebih sesuai dengan data yang digunakan. Parameter-parameter yang dapat diatur antara lain adalah nilai *Box Constraint* atau nilai *Cost* dan nilai *Kernel Scale* yang umum disebut nilai *gamma*. Semakin besar nilai *Cost*, maka semakin rentan hasil klasifikasi SVM terhadap outlier, sedangkan semakin kecil nilai maka semakin kecil pengaruh outlier dan data yang terklasifikasi dengan salah. Nilai *gamma* memiliki fungsi dalam seberapa besar pengaruh *kernel* pada model yang dibuat, semakin besar nilai *gamma* maka semakin besar pengaruh *kernel* pada *hyperplane* yang dibuat. *Cost* pada umumnya memiliki nilai dalam kelipatan sepuluh seperti 10^{-2} hingga 10^2 , Sedangkan nilai *gamma* adalah nilai bilangan positif. Parameter Tuning dilakukan dengan cara melakukan *Grid Search* pada data ekstraksi ciri yang dimasukkan. *Grid search* bekerja dengan cara melakukan pemodelan berulang pada data dengan nilai *Cost* dan *Gamma* yang ditentukan melalui *range*.. Dalam penelitian ini, *Grid search* dilakukan dalam range 10^{-2} hingga 10^2 dan dalam range 2^{-3} dan 2^0 untuk nilai *gamma*. Tabel menggambarkan parameter terbaik dari masing-masing kernel.

Tabel 19 : Parameter Terbaik tiap Kernel

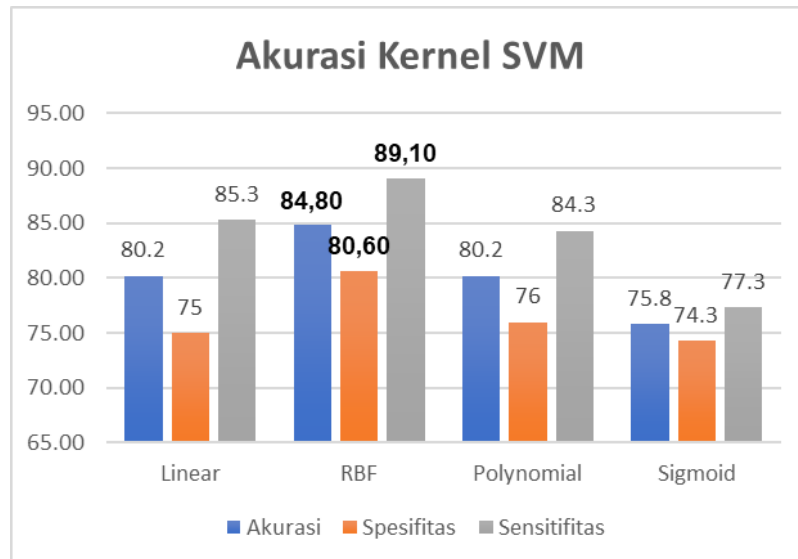
<i>Kernel</i>	<i>Box Constraint (Cost)</i>	<i>Kernel Scale (Gamma)</i>
Linear	100	-*
RBF	1	0.125
Polynomial	0.01	0.5
Sigmoid	0.01	0.125

Setelah nilai-nilai parameter telah didapatkan, maka tahapan selanjutnya adalah menggunakan parameter-parameter di atas untuk mengklasifikasikan *data training* dan menggunakan *K-Fold Cross Validation* untuk mengevaluasinya.

4.9 Evaluasi Kernel SVM

Setelah parameter *kernel* telah didapatkan, maka langkah selanjutnya adalah menguji performa SVM untuk mengklasifikasi citra yang ada pada *data training*. SVM memiliki beberapa *kernel* yang dapat digunakan untuk mengklasifikasikan data. *Kernel* SVM yang diuji antara lain *Linear*, *Radial Basis Function (RBF)*, *Polynomial*, dan *Sigmoid*. Keempat *kernel* ini akan diuji coba pada *data training* dengan metode evaluasi *K-Fold Cross Validation* dan dihitung nilai akurasi guna mendapatkan model terbaik dalam mengklasifikasikan data yang ada pada *data training*. *K-Fold Cross Validation* akan membagi data menjadi data training dan data *Testing* pada tiap fold yang berbeda. Nilai K yang dipilih pada *K-Fold Cross Validation* adalah 10, sehingga data training dibagi menjadi 10 bagian berbeda yang dijadikan sebagai *data Testing* pada tiap iterasinya. Gambar 5 menggambarkan hasil yang

didapat masing-masing *kernel* SVM dalam mengklasifikasikan data training dengan *K-Fold Cross Validation* model



Gambar 112 : Hasil evaluasi kernel

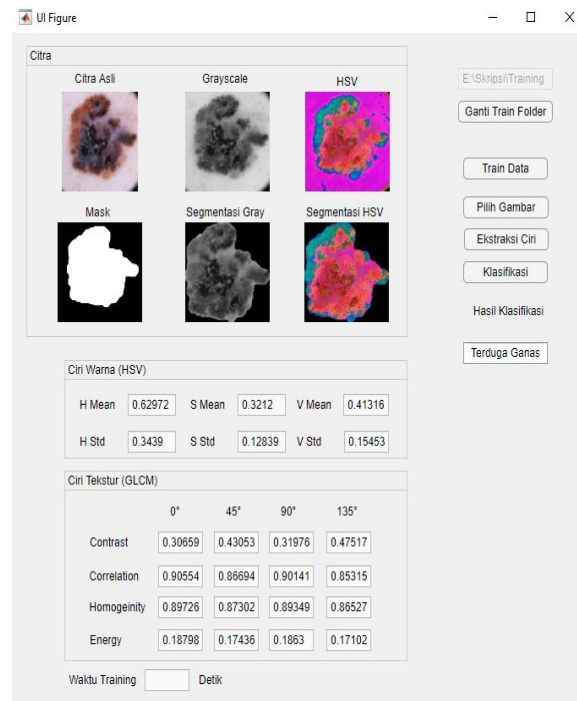
Dari gambar di atas, dapat dilihat bahwa *kernel* SVM RBF mendapatkan hasil terbaik dengan akurasi sebesar 84,8%. *Kernel* Linear mendapatkan akurasi sebesar 80,20%, Polynomial sebesar 80,20% dan Sigmoid sebesar 75,8%. Nilai sensitivitas tertinggi didapatkan dengan *kernel* RBF dengan nilai sebesar 89,1%, dan nilai spesifisitas didapatkan oleh *kernel* RBF dengan nilai sebesar 80,6%. Dari pengujian ini, dapat disimpulkan bahwa *Support Vector Machine* dengan *kernel* RBF memiliki akurasi tertinggi dibandingkan dengan *kernel* lain dan akan digunakan pada aplikasi yang akan dibuat.

4.10 Evaluasi Aplikasi

Data Testing akan diproses melalui aplikasi untuk mengetahui kelas dari citra yang dimasukkan. Data Testing terdiri dari 65 citra yang tergolong tumor ganas dan 65 citra yang tergolong dalam tumor jinak. Adapun tahapan-tahapan yang dilakukan untuk menguji data Testing, antara lain:

- 1) Membuat model dari data training sesuai folder yang sudah ditentukan.
- 2) Memasukan gambar citra Testing melalui tombol pilih gambar
- 3) Mengekstraksi ciri citra Testing yang dimasukkan
- 4) Mengklasifikasikan citra dengan model SVM kernel RBF yang telah dibuat.

Tahapan di atas, kecuali tahapan 1., dilakukan berulang kali untuk tiap citra yang tergolong pada data Testing. Hasil klasifikasi yang dihasilkan akan dievaluasi menggunakan confusion matrix untuk menghitung akurasi, spesifisitas dan sensitivitas dari data yang diuji RBF. Gambar 6 mengilustrasikan proses dari tahapan ini.



Gambar 6 : Proses Aplikasi

Setiap data *Testing* akan dicatat hasil klasifikasinya untuk dapat menghasilkan *confusion matrix* yang digunakan untuk menghitung akurasi, spesifisitas, dan sensitivitas. Adapun *confusion matrix* yang didapatkan dari tahapan di atas dideskripsikan pada **Error! Reference source not found.** dimana G adalah kelas Ganas dan J menggambarkan kelas Jinak.

Tabel 20 : Confusion Matrix Data Testing

Observasi	Prediksi	
	G	J
G	50	15
J	15	50

Dari table diatas, maka dapat dihitung akurasi, spesifisitas, dan sensitifitas dari aplikasi ini sebagai berikut.

1. Akurasi

$$Akurasi = \frac{50 + 50}{50 + 15 + 50 + 15} = 76.9\% \tag{4.1}$$

2. Sensitivitas

$$Sensitivitas = \frac{50}{50 + 15} = 0.769 = 76.9\% \tag{4.2}$$

3. Spesifisitas

$$\text{Spesifisitas} = \frac{50}{50 + 15} = 0.769 = 76.9\% \quad (4.3)$$

5 Penutup

5.1 Kesimpulan

Penelitian ini berhasil mengidentifikasi keganasan tumor kulit dengan Support Vector Machine dengan akurasi sebesar 76.9%. Hasil ini lebih buruk dibandingkan dengan penelitian yang dilakukan Eltayel (2017). Hal ini dipengaruhi oleh resolusi data dan pencahayaan yang tidak seragam dan juga hasil segmentasi dari thresholding yang masih belum sempurna. Penelitian dimulai dari tahapan Pra-proses dan segmentasi yang menghasilkan 1253 citra dengan hasil segmentasi yang baik, dilanjutkan dengan tahapan pembagian data dan Random Undersampling yang memberikan data yang seimbang untuk masing-masing kelas dengan jumlah data 1218 citra. Setelah itu, tahapan ekstraksi ciri yang memungkinkan pengambilan fitur warna dan tekstur dari citra. Tahapan parameter tuning memberikan parameter-parameter yang digunakan untuk memberikan hasil klasifikasi yang optimal. Data training diklasifikasikan menggunakan Support Vector Machine dan dievaluasi dengan K-Fold Cross Validation. Hasil terbaik didapatkan kernel SVM Radial Basis Function dengan nilai akurasi yang didapat sebesar 84,8% dengan evaluasi K-Fold-Validation dengan nilai K sebesar 10. Support Vector Machine dengan kernel RBF selanjutnya dipasang pada aplikasi dan digunakan untuk mengklasifikasikan data testing yang telah dibuat pada tahap pembagian data. SVM kernel RBF mendapatkan nilai akurasi, sensitivitas, dan spesifisitas sebesar 76.9%, 76.9% dan 76.9% dalam mengklasifikasikan citra data testing .

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, ada beberapa ruang peningkatan untuk mengoptimalkan hasil yang didapatkan dari klasifikasi. Resolusi dan pencahayaan pada masing-masing citra yang didapatkan belum seragam sehingga memberikan beberapa limitasi terhadap proses cropping dan ekstraksi fitur ciri warna dari citra. Hal ini dapat diatasi dengan pengumpulan data yang lebih terkontrol atau praproses citra yang berfokus pada pencahayaan . Data yang seragam juga dapat memberikan fitur baru yang dapat diolah seperti ukuran dan bentuk tumor. Pendekatan pada proses segmentasi juga dapat dilakukan dengan menggunakan basis lain seperti segmentasi berbasis tepi dan berbagai metode segmentasi lain yang mungkin memiliki hasil yang lebih baik dalam mensegmentasi citra.

Referensi

- [1] Andono, P. N., Sutojo, T. & Muljono, 2017. *Pengolahan Citra Digital*. 1 ed. Yogyakarta: Andi.
- [2] BT Series, 2018. *Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios*. 7th ed. Geneva: International Telecommunication Union.
- [3] Codella , N. C. et al., 2017. *Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI)*, Melbourne: ISBI.
- [4] Eltayel, K., Li, Y., Dodo, B. & Liu, X., 2017. Skin Cancer Detection in Dermoscopy Images Using Sub-region Features. *Springer International Publishing*, pp. 75-86.
- [5] Hall-Beyer, M., 2017 . *GLCM Texture: A Tutorial 3.0*. 3rd ed. Calgary: PRISM : University of Calgary.

- [6] Khan, A. M. & Ravi, S., 2013. Image Segmentation : A Comparative Study. *International Journal of Soft Computing and Engineering*, 3(4), pp. 84 - 92.
- [7] Kumara, I. N. S., 2013. Identification of Nitrogen Status in Brassica Juncea Using Color Moment, GLCM, And Backpropagation Neural Network. *ISABE*, 1(1), pp. D121 - D126.
- [8] National Cancer Institute, 2012. *What You Need to Know About : Brain Tumor*. 1st ed. Washington: U.S Department of Health and Human Services.
- [9] National Cancer Institute, 2012. *What You Need to Know About Melanoma and Other Skin Cancers*. 1st ed. s.l.:National Institutes of Health.
- [10] Ravi, S. & Khan, A. M., 2013. Morphological Operations For Image Processing. *NCVSComs*, Volume 13, pp. 17 - 19.
- [11] Saleh, E., 2016. *Neoplasma*. Yogyakarta: Universitas Muhammadiyah Yogyakarta.
- [12] Sinha, T., 2018. Tumors : Benign and Malignant. *Cancer Therapy & Ontology*, 10(3), pp. 1 - 3.
- [13] Sukafona , I. M., 2018. Content Based Image Retrieval dengan Metode Color Moment dan K-Means. *RESISTOR*, 1(2), pp. 73 - 78.
- [14] Tschandl, P., Rosendahl, C. & Kittler, H., 2018. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 5(180161).
- [15] Wahyuni, E. S., 2016. PENERAPAN METODE SELEKSI FITUR UNTUK MENINGKATKAN HASIL DIAGNOSIS KANKER PAYUDARA. *SIMETRIS*, 7(1), pp. 283 - 293.
- [16] Yousefi, J., 2011. *Image Binarization using Otsu Method Algorithm*, Ontario: University Of Guelph.