

# Deteksi Kanker Kulit Melanoma dengan *Linear Discriminant Analysis-Fuzzy k-Nearest Neighbour Lp-Norm*

Mustika Mentari<sup>1</sup>, Yuita Arum Sari<sup>2</sup>, Ratih Kartika Dewi<sup>3</sup>

<sup>1</sup>Teknologi Informasi, Politeknik Negeri Malang, Malang

<sup>2,3</sup>Teknik Informatika, Universitas Brawijaya, Malang

E-mail: <sup>1</sup>must.mentari@gmail.com, <sup>2</sup>y.arumsari@gmail.com, <sup>3</sup>ratihkartikadewi@yahoo.com

## Abstrak

Seiring perkembangan teknologi dilakukan otomatisasi deteksi kanker kulit melalui citra *dermoscopy*. Pengambilan informasi fitur citra *dermoscopy* terganggu dengan outlier dan *overfitting*, karena faktor jenis kulit, penyebaran kanker yang tidak merata atau kesalahan sampling. Penelitian ini mengusulkan deteksi kanker kulit melanoma dengan mengintegrasikan metode fuzzy K-Nearest Neighbour (FuzzykNN), Lp-norm dan *Linear Discriminant Analysis* (LDA) untuk mengurangi outlier dan *overfitting*. Masukan berupa citra warna RGB yang dinormalisasi menjadi RGBr. Reduksi dimensi dengan LDA menghasilkan fitur dengan nilai *eigen* paling menonjol. LDA pada penelitian ini menghasilkan dua fitur paling menonjol dari 141 jenis fitur, yaitu wilayah tumor dan minimum wilayah tumor channel R. Kemudian dilakukan klasifikasi FuzzykNN dan metode pengukur jarak Lp-norm. Penggunaan metode LDA dan Lp-norm dalam proses klasifikasi ini mengatasi terjadinya *overfitting*. Akurasi yang dihasilkan metode LDA-fuzzykNN Lp Norm, yaitu 72% saat masing-masing nilai p dan k = 25. Metode gabungan ini terbukti cukup baik dari pada metode yang dijalankan terpisah.

**Kata kunci:** melanoma, fuzzy, KNN, Lp-norm, LDA.

## Abstract

As the advancement of technology skin cancer detection need to be automated with the use of *dermoscopy* image. Outlier and *overfitting* are the problem in feature extraction of *dermoscopy* image, this can be caused by skin type, uneven cancer distribution or sampling error. This study proposed melanoma skin cancer detection by fuzzy K-Nearest Neighbour (FuzzykNN) with Lp-norm integrated with *Linear Discriminant Analysis* (LDA) to reduce the problem of outlier and *overfitting*. Input used in this study are images with RGB channel, then it adapted to RGBr. Dimensional reduction with LDA result in features with highest eigen value. LDA in this research select 2 discriminant, they are tumor area and minimum tumor area in R channel. This features then classified by fuzzykNN with Lp-Norm. Integration of LDA and Lp-norm in classification can reduce the problem of *overfitting*. This study results in 72% accuracy when the value of p and k are 25. Integration of LDA and fuzzykNN with Lp-norm has better result than unintegrated method.

**Key word:** melanoma, fuzzy, KNN, Lp-norm, LDA.

## 1. Pendahuluan

Diagnosis kanker kulit merupakan salah satu cabang dari bidang kesehatan yang digunakan untuk mendeteksi adanya kelainan pada kulit. Pada citra *dermoscopy* (kanker kulit), warna merupakan fitur dengan banyak informasi untuk mengenali kanker kulit yang diderita oleh pasien. Seiring perkembangan teknologi penelitian menunjukkan perlu adanya otomatisasi untuk deteksi kanker kulit melalui citra *dermoscopy* (Iyatomi, Celebi, Schaefer, & Tanaka, 2011).

Beberapa penelitian terkait penggunaan fitur warna citra *dermoscopy* yang menggunakan sistem kalibrasi warna secara otomatis untuk mengenali jenis penyakit kanker kulit melanoma (Iyatomi, Celebi, Schaefer, & Tanaka, 2011). Penelitian mengenai ekstraksi fitur dengan menggunakan warna dengan metode histogram, diantaranya adalah *Pixel based Color Moments Descriptor* (PCMD), *Color Histogram Moments Descriptor* (CHMD), *Single Channel Histogram Moments Descriptor* (SCHMD), *Maximum Frequency Symmetrical Color Spatial Feature* (MFSCSF), *Symmetrical Color Spatial Histogram* (SCSH), dan *Binnary Haar Color Descriptor* (BHCD) (Banu & Nallaperumal, 2010). Fungsi ekstraksi fitur diimplementasikan ke dalam teknik *Content Based Image Retrieval* (CBIR). Hasil dalam penelitian tersebut yang paling optimal menggunakan BHCD dengan model warna HSV.

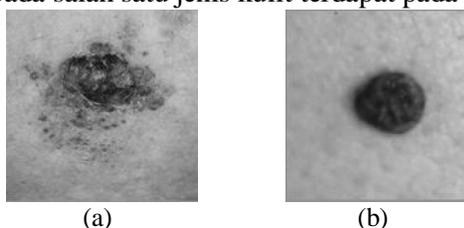
Penelitian citra *dermoscopy* dengan KNN pada salah satunya mengenai deteksi melanoma secara otomatis dengan *K-Nearest Neighbor* (Ganster, et al., 2001). Penelitian (Ganster, et al., 2001) menggunakan teknik seleksi fitur, kemudian melakukan klasifikasi dan mendapatkan performa terbaik saat parameter masukan  $k$  bernilai 24. Penelitian (Burroni, et al., 2004) mengembangkan aplikasi dengan menggunakan bantuan komputer untuk mendeteksi adanya melanoma dengan menggunakan algoritma KNN. Evaluasi akurasi dari 4 pengelompokan berbeda untuk diterapkan pada citra *dermoscopy*, 3 pengelompokan yang bersifat linear dan menunjukkan bahwa dengan menggunakan metode KNN menghasilkan akurasi tertinggi. Deteksi melanoma otomatis dapat menggunakan teknik klasifikasi sebagaimana penelitian yang dilakukan oleh (Maglogiannis & Kosmopoulos, 2006) dan (Ruiz, Berenguer, Soriano, & Sánchez, 2011). Penelitian (Ruiz, Berenguer, Soriano, & Sánchez, 2011) membangun sistem penunjang keputusan untuk diagnosis melanoma dengan pendekatan yang membandingkan metode klasifikasi, yaitu KNN, *perceptron* dan bayes untuk kasus melanoma. Penelitian (Ruiz, Berenguer, Soriano, & Sánchez, 2011) melakukan klasifikasi citra *dermoscopy* dengan analisis diskriminan, jaringan saraf tiruan dan *support vector machine* (SVM). Fuzzy KNN pada penelitian (Farag & Elhabian, 2008) menunjukkan hasil yang signifikan, di mana metode tersebut peka terhadap derau.

Penggunaan LDA rentan terhadap adanya *outlier* dan *overfitting*, di mana keadaan tersebut sering terjadi saat pengambilan informasi fitur. Untuk menghindari adanya penyebaran informasi yang tidak merata, maka pada penelitian ini diusulkan metode untuk deteksi kanker kulit melanoma dari citra *dermoscopy* menggunakan LDA sebagai metode ekstraksi fitur yang diintegrasikan dengan fuzzykNN dan metode pengukur jarak Lp-Norm (perbaikan metode *euclidean*) sebagai metode klasifikasi dengan tujuan mengurangi *outlier* dan *overfitting*.

## 2. Metode Penelitian

### 2.1. Analisis Data

Data yang digunakan adalah citra *dermoscopy* dengan format model warna RGB. Terdapat 200 data citra *dermoscopy* yang diambil dari beberapa penderita kanker kulit dengan bentuk yang beragam untuk dua jenis kanker kulit yang diangkat pada penelitian ini, masing-masing jenis kanker tersebut terdiri dari 100 jenis data melanoma dan data 100 jenis data non melanoma. Data dibagi ke dalam 75% data yang digunakan untuk proses *training* dan 25% data untuk proses *testing*, sehingga terdapat data 150 untuk data *training* dan 50 untuk data *testing*. Contoh gambar melanoma dan non-melanoma dengan pencahayaan yang cukup baik pada salah satu jenis kulit terdapat pada Gambar 1.



Gambar 1 (a) Melanoma, (b) Non-melanoma

### 2.2. Metode LDA-FuzzykNN-Lp-Norm

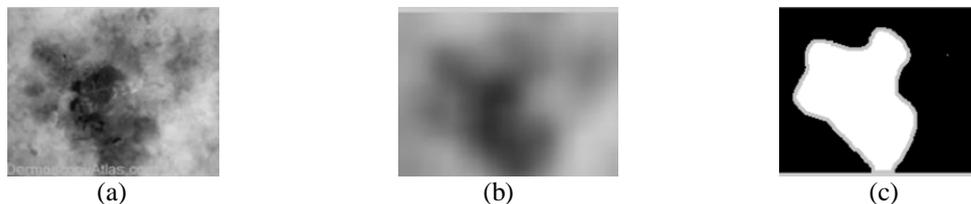
Sistem dibangun secara *offline* ini menggunakan data berupa citra dengan format RGB yang ditransformasikan dengan RGB modifikasi (Liu, 2011), menjadi model warna  $RGBr$  dimana proses menghitung nilai  $RGBr$  dijelaskan pada persamaan (1).

$$\begin{aligned}
 \hat{R}_r &= \frac{R(x,y) - G(x,y)}{2} \\
 \hat{G}_r &= \frac{R(x,y) - B(x,y)}{2}
 \end{aligned} \tag{1}$$

Di mana ukuran piksel citra berwarna dimisalkan sebagai  $M \times N$ , di mana masing-masing piksel memiliki nilai RGB pada posisi  $x$  dan  $y$ . Nilai  $R_r$ ,  $G_r$ , dan  $B_r$  adalah hasil modifikasi dari warna utama RGB. Normalisasi merupakan pemetaan dari warna asli ke dalam prosentase warna dari sebuah piksel

pada citra agar dikenali lebih baik. Secara umum metodologi pada penelitian ini dapat digambarkan pada Gambar 2.

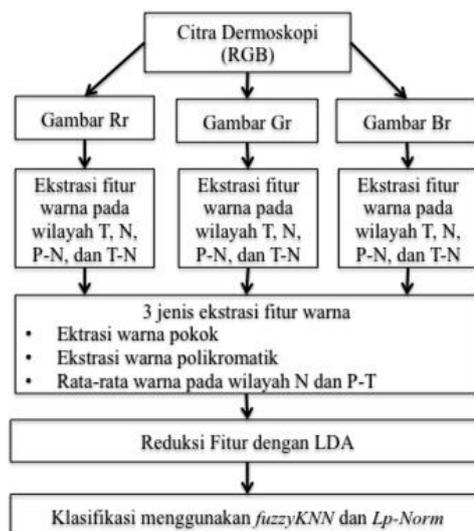
Masing-masing dari hasil normalisasi gambar dengan warna R, G, dan B pada masing-masing wilayah dengan warna kulit yang terdeteksi tumor, wilayah kulit normal, kulit antara tepi dengan kulit normal, dan wilayah antara kulit yang terdeteksi tumor dengan kulit normal. Gambar 2 menunjukkan cara menentukan wilayah kulit yang dideteksi sebagai melanoma, kulit normal, serta tepi.



Gambar 2 Penentuan wilayah-wilayah untuk deteksi melanoma (a) Gambar asli, (b) gambar filterisasi, (c) pengenalan wilayah terdeteksi kanker (warna putih), tepi (warna hijau), dan kulit normal (warna hitam)

Penentuan wilayah tersebut dilakukan dengan ekstraksi fitur dari warna pokok, warna polikromatik, dan warna lain. Ekstraksi fitur warna terdiri dari *mean*, minimum, maksimum, standar deviasi, dan *skewness*. Ekstraksi warna polikromatik dengan menggunakan 8 dan 16 kuantisasi. Warna lain meliputi rata-rata warna di sekitar kulit normal, rata-rata warna tepi dengan warna kulit yang terdeteksi tumor. Secara keseluruhan sistem deteksi kanker melanoma terdapat pada Gambar 3.

Masing-masing dari *channel* model warna RGB dilakukan reduksi fitur dengan menggunakan metode LDA dan *Lp-norm*. Metode LDA dan digabung dengan *Lp-norm* untuk proses klasifikasi berfungsi mengatasi permasalahan *overfitting* dan *outlier*. *Overfitting* adalah suatu keadaan di mana tidak terdapat keseimbangan informasi yang terdapat dalam data sampel, sedangkan *outlier* adalah data yang terlalu jauh dengan anggota lain pada suatu kelas.



Gambar 3 Diagram blok metodologi sistem

Metode LDA metransformasikan secara optimal dengan meminimalkan perbedaan rasio *within-class* dan memaksimalkan perbedaan rasio *between-class* (Liu, 2011), diawali dengan perhitungan *mean* tiap kelas dan perhitungan  $S_i$  yaitu *covariance* masing-masing kelas tersebut (*within covariance*) dengan persamaan (2). *Covariance* seluruh kelas yang telah didapat dijumlahkan pada  $S_w$  (*Within-class scatter matrix*) sesuai persamaan (3).

Perbedaan rasio antar kelas dihitung dengan  $S_B$  (*Between-class scatter matrix*) sesuai persamaan (4). Kemudian, perhitungan *eigen value*  $\lambda$  dan *eigen vector*  $w$  melalui perkalian antara pada persamaan (5). Melalui *eigen value* ini dapat diketahui fitur yang paling berpengaruh. *Eigen value* merepresentasikan fitur bagian kolom *eigen vector* yang akan digunakan dalam perhitungan proyeksi  $y$ , yaitu dengan mengalikan data *training* atau data *testing* dengan *transpose* dari *eigen vector* tersebut sesuai persamaan (6).

$$S_i = \sum_{x \in \omega_i} (x - \mu_i)(x - \mu_i)^T \tag{2}$$

$$S_w = S_1 + S_2 \tag{3}$$

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T \tag{4}$$

$$S_w^{-1} S_B w = \lambda w \tag{5}$$

$$y = w^T x \tag{6}$$

Hasil LDA berupa proyeksi data digunakan sebagai masukan pada proses klasifikasi menggunakan *FuzzykNN* yang meningkatkan distribusi informasi pada masing-masing anggota data sampel (Song, Zheng, Wu, Yang, & Yanga, 2010). Persamaan dari Lp-norm dapat dijelaskan pada persamaan (7).

$$d_{st} = \sqrt[p]{\sum_{j=1}^n |x_{sj} - x_{tj}|^p} \tag{7}$$

Di mana  $x_s$  dan  $x_t$  adalah dua data yang dihitung jarak. Parameter masukan yang berupa bilangan bulat positif adalah  $p$ , dan  $n$  adalah seluruh data (Oh & Kwak, 2013). Sedangkan nilai keanggotaan pada *Fuzzy* dihitung dengan persamaan (8).

$$u_i(x) = \frac{\sum_{j=1}^k u_{ij} (1/\|x - x_j\|^{2/(m-1)})}{\sum_{j=1}^k (1/\|x - x_j\|^{2/(m-1)})} \tag{8}$$

Di mana  $u_i(x)$  merupakan nilai keanggotaan *inverse* jarak atau kedekatan antar data data *training*  $x_j$  ke data *testing*  $x$ . Parameter  $k$  memberikan pilihan jumlah data yang mempunyai kemiripan paling tinggi atau jarak paling dekat dengan data *testing*  $x$ . Untuk bobot pangkat  $m$  harus bernilai lebih besar dari 1 ( $m > 1$ ) (Keller, Gray, & Givens, 1985).

Klasifikasi KNN telah dikembangkan menjadi metode *FuzzykNN* dengan teknik perhitungan jarak Lp-norm. Masukan berupa data *training*, label data *training*,  $k$ , dan data *testing*, yang kemudian dihitung jarak setiap data *testing* dengan data *training* menggunakan Lp-norm. Setelah perhitungan jarak ini akan ditambahkan metode fuzzy dengan perhitungan nilai keanggotaan sesuai persamaan. Kemudian diambil nilai yang paling besar sebagai hasil klasifikasi sesuai jumlah parameter  $k$  yang telah ditentukan.

### 3. Hasil Penelitian dan Pembahasan

LDA-fuzzykNN-LpNorm membantu dalam melakukan pemilihan fitur yang memiliki informasi paling dominan. Fitur yang terbentuk pada awalnya terdiri dari 141 fitur. Reduksi dimensi dengan menggunakan LDA menghasilkan 2 fitur dari 141 fitur awal. Dua fitur dengan ciri paling kuat yaitu nilai dari rata-rata dan minimum dari *channel* warna R. Hasil percobaan dengan dan tanpa menggunakan LDA dapat ditunjukkan pada Tabel 1.

Tabel 1 Hasil akurasi dengan LDA dan tanpa LDA

Metode	Dengan LDA(%)	Tanpa LDA(%)
f_kNN-L1	52	48
f_kNN-L2	58	48
f_kNN-Lp=21	70	44
f_kNN-Lp=25	72	44

*FuzzykNN* dengan pengukuran jarak, melalui parameter nilai masukan untuk  $p$  yang berbeda, menghasilkan akurasi yang lebih baik jika menggunakan LDA. Fitur yang terlalu banyak menyebabkan salah informasi dalam pembentukan model, sehingga terdapat banyak kesalahan pada proses *testing*. Teknik LDA dapat mengetahui hubungan antar fitur yang memiliki korelasi lebih kuat (menonjol). Informasi yang menonjol membuat hasil pembentukan model dengan pengujian data *testing* menjadi lebih baik. *FuzzykNN* dapat melakukan proses

klasifikasi secara teliti, karena perhitungan tiap data dipertimbangkan nilai kekaburannya, sehingga proses klasifikasi dapat mengelompokkan data dengan lebih akurat. Hasil percobaan menunjukkan parameter  $k$  seharusnya tidak terlalu rendah atau tinggi. Nilai yang menunjukkan hasil paling baik pada penelitian ini adalah nilai  $k = 25$ . Metode ini menghitung jarak paling terdekat dengan posisi lain yang berada di sekitarnya dengan metode *minkowski distance* (Lp-norm), terbukti mempunyai ketelitian yang lebih baik dibandingkan L1-norm (*Manhattan Distance*) ataupun L2-norm (*Euclidean Distance*).

Penggabungan metode reduksi dimensi LDA dengan metode klasifikasi fuzzykNN dan metode perhitungan jarak Lp-Norm menunjukkan hasil yang lebih tinggi dibandingkan jika percobaan tidak menggunakan LDA pada proses klasifikasi menggunakan fuzzykNN-Lp-norm. Pada Tabel 1, akurasi tertinggi sebesar 72% dilakukan dengan LDA fuzzykNN Lp-Norm dengan nilai parameter  $k$  dan  $p$  masing-masing sebesar 25.

Kekurangan pada LDA-fuzzykNN-Lp-norm adalah tidak dapat menggunakan parameter masukan yang terlalu tinggi atau terlalu rendah (terlalu tinggi menjauhkan informasi antar fitur, terlalu rendah rentan derau). LDA mempunyai peranan yang cukup baik ketika berkorelasi dengan fuzzykNN dan Lp-norm, dibandingkan klasifikasi fuzzykNN-Lp-norm tanpa menggunakan LDA.

#### 4. Kesimpulan

Reduksi dimensi dengan LDA meningkatkan akurasi pada proses klasifikasi citra *dermoscopy* dibandingkan tanpa menggunakan metode LDA. LDA dapat memisahkan variansi antar kelas dan mengumpulkan variasi dalam kelas dengan cukup baik. Penelitian selanjutnya dapat digunakan metode reduksi yang lebih baik dibandingkan LDA untuk mengetahui keterkaitan antar satu fitur dengan fitur yang lain. Penentuan kedekatan dengan menggunakan Lp-norm pada fuzzykNN sangat dipengaruhi nilai  $p$  dan  $k$  yang tidak bisa memiliki nilai yang terlalu kecil atau terlalu tinggi, sehingga untuk penelitian selanjutnya, dapat digunakan metode untuk menentukan batas ambang dengan lebih baik dan efisien.

#### 5. Ucapan Terima Kasih

Ucapan terima kasih kami sampaikan kepada Bapak Agus Zainal Arifin yang telah membimbing kami dalam pembuatan jurnal ini.

#### 6. Referensi

- Banu, M., & Nallaperumal, K. (2010). Analysis of Color Feature Extraction Techniques for Pathology Image Retrieval System. *Computational Intelligence and Computing Research (ICCIC), 2010 IEEE International Conference on*. Coimbatore.
- Burroni, M., Corona, R., Dell'Eva, G., Sera, F., Bono, R., Puddu, P., . . . Rubegni, P. (2004). Melanoma Computer-Aided Diagnosis Reliability and Feasibility Study. *Clinical Cancer Research, 10*(6), 1881–1886.
- Farag, A. A., & Elhabian, S. Y. (2008). A Tutorial on Data Linear Discriminant Analysis (LDA). Louisville, Kentucky: CVIP Lab University of Louisville.
- Ganster, H., Pinz, A., Röhrer, R., Wildling, E., Binder, M., & Kittler, H. (2001). Automated Melanoma Recognition. *Ieee Transactions On Medical Imaging, 20*(3), 233-239.
- Iyatomi, H., Celebi, M. E., Schaefer, G., & Tanaka, M. (2011). Automated Color Calibration Method for Dermoscopy Images. *Computerized Medical Imaging and Graphics, 35*(2), 89–98.
- Keller, J. M., Gray, M. R., & Givens, J. A. (1985). A Fuzzy K-Nearest Neighbor Algorithm. *Systems, Man and Cybernetics, IEEE Transactions on, SMC-15*(4), 580 - 585.
- Liu, C. (2011). Extracting Discriminative Color Features for Face Recognition. *Pattern Recognition Letters, 32*(14), 1796–1804.
- Maglogiannis, I., & Kosmopoulos, D. I. (2006). Computational Vision Systems for The Detection of Malignant Melanoma. *Oncology Reports, 15*(4), 1027-1032.

- Oh, J. H., & Kwak, N. (2013). Generalization of Linear Discriminant Analysis Using Lp-norm. *Pattern Recognition Letters*, 34(6), 679–685.
- Ruiz, D., Berenguer, V., Soriano, A., & Sánchez, B. (2011). A Decision Support System for The Diagnosis of Melanoma: A Comparative Approach. *Expert Systems with Applications*, 38(12), 15217–15223.
- Song, X. N., Zheng, Y. J., Wu, X. J., Yang, X. B., & Yanga, J. Y. (2010). A complete fuzzy discriminant analysis approach for face recognition. *Applied Soft Computing*, 10(1), 208–214.