

ANALISIS GANGGUAN POLYCYSTIC OVARY SYNDROME (PCOS) BERDASARKAN CITRA ULTRASONOGRAPHY MENGGUNAKAN METODE LINEAR DISCRIMINANT ANALYSIS (LDA) DAN FUZZY C-MEAN CLUSTERING

Adwin Rahmanto, Dr. Adiwijaya, M.Si, Untari Novia Wisesty, M.T.
Fakultas Informatika Institut Teknologi Telkom
Jl. Telekomunikasi, Bandung
adwinrahmanto@gmail.com

Abstrak

Masalah kesuburan terjadi akibat terganggunya sistem reproduksi pada wanita dan terjadinya penurunan kualitas sperma pada pria. Sebuah penelitian menyatakan bahwa masalah kesuburan terjadi pada 40% akibat perempuan, 40% akibat laki-laki dan 20% akibat keduanya. Oleh karena itu, pemeriksaan dini kesuburan sangatlah perlu khususnya untuk wanita agar dapat dilakukan pencegahan secara dini hal-hal yang menyebabkan kemandulan. Salah satu yang sangat dianjurkan yaitu dengan pemeriksaan USG (*Ultrasonography*). USG (*Ultrasonography*) adalah suatu kaidah pemeriksaan tubuh menggunakan gelombang bunyi pada frekuensi tinggi.

Penelitian ini bertujuan untuk menghasilkan suatu sistem aplikasi yang dapat mendiagnosa citra USG (*Ultrasonography*) dan mengklasifikasikan rahim ke dalam kelas normal atau terdeteksi PCOS (*Polycystic Ovary Syndrome*). Proses pendeteksian diawali dengan pemrosesan awal pada citra rahim, proses ekstraksi ciri menggunakan *Linear Discriminant Analysis* (LDA), dan proses klasifikasi menggunakan *Fuzzy C-Mean Clustering*. Pemrosesan awal dilakukan untuk membuang informasi yang tidak dibutuhkan dalam pengolahan citra.

Data keseluruhan yang digunakan dalam tugas akhir ini berjumlah 167 citra. Pengujian sistem dilakukan dengan penentuan pengambilan nilai w (pembobot), dan jumlah data latih normal maupun terdeteksi PCOS. Dari hasil pengujian diperoleh hasil pengujian terbaik dengan akurasi 94,44% untuk data citra anggota kelas.

Kata Kunci : *Polycystic Ovary Syndrom*, Kesuburan, *Ultrasonography*, *Linear Discriminant Analysis*, *Fuzzy C-Mean Clustering*

Abstract

Fertility problems caused by disruption of the reproductive system in women and decreasing quality of sperm in men. A study stating that fertility problems occurred in 40% as a result of women, 40% as a result of men, and 20% due to both. Therefore, early fertility examination is necessary, especially for women who want to do early prevention of the things that cause fertility. One that is highly recommended is USG (Ultrasonography) examination. USG (ultrasonography) is an examination of the rules of the body using sound waves at high frequencies.

This research aims to produce system application that can diagnose and classify the USG image of the uterus into the normal class or detected PCOS (Polycystic Ovary Syndrome). Detection process begins with the initial processing of the image of uterus, the process of feature extraction using Linear Discriminant Analysis (LDA), and the classification process using Fuzzy C-Mean Clustering. Initial processing is done to get rid of unneeded information in image processing.

Overall the data used in this thesis amounts to 167 images. System testing is done by taking the value of determining w (weighting), and the amount of training data detected normal and PCOS (Polycystic Ovary Syndrome). Of test results obtained by the best test results with 94,44% accuracy for image data members of the class.

Keyword : *Polycystic Ovary Syndrome*, Fertile, *Ultrasonography*, *Linear Discriminant Analysis*, *Fuzzy C-Mean Clustering*

1. Pendahuluan

Masalah kesuburan merupakan hal yang sangat mengganggu bahkan bisa mengancam keutuhan suatu rumah tangga. Masalah kesuburan terjadi akibat terganggunya sistem reproduksi pada wanita dan terjadinya penurunan kualitas sperma pada pria. Sebuah penelitian menyatakan bahwa masalah kesuburan terjadi pada 40% akibat perempuan, 40% akibat laki-laki dan 20% akibat keduanya. Oleh karena itu, pemeriksaan dini kesuburan sangatlah perlu khususnya untuk wanita agar dapat dilakukan pencegahan secara dini hal-hal yang menyebabkan kemandulan. Salah satu yang sangat dianjurkan yaitu dengan pemeriksaan USG (*Ultrasonography*). USG (*Ultrasonography*) adalah suatu kaidah pemeriksaan tubuh menggunakan gelombang bunyi pada frekuensi tinggi. USG ini adalah salah satu aplikasi teknologi radar dan telah ada sejak puluhan tahun lalu. Untuk memeriksa kesuburan wanita dilakukan USG rahim dengan memeriksa keberadaan penyakit di rahim yang menyebabkan kemandulan. Namun, sampai saat ini penentuan hasil USG rahim masih dilakukan secara manual dengan visual manusia. Hal ini memakan waktu yang lama karena harus dilakukan diagnosa terlebih dahulu dari hasil USG tersebut dan menghasilkan diagnosa yang tidak konsisten antar ahli USG karena keterbatasan visual manusia.

Pada penelitian ini bermaksud untuk membantu ahli medis dalam mendiagnosa kesuburan wanita secara terkomputerisasi, sehingga hasil diagnosa dapat dilakukan dengan cepat dan akurat. Simulasi dilakukan dengan mendeteksi citra ultrasonografi rahim berdasarkan hasil ekstraksi ciri ke dalam dua jenis klasifikasi yaitu terdeteksi *PCOS* atau tidak terdeteksi *PCOS*.

Proses penedeteksian gangguan *PCOS* ini diawali dengan pemrosesan awal pada citra USG, proses ekstraksi ciri menggunakan *Linear Discriminant Analysis (LDA)* yang berfungsi untuk memetakan matriks ke dalam dimensi yang lebih rendah. Selain itu LDA juga berfungsi untuk meminimalisasi jarak antar ciri dalam kelas dan memaksimalkan jarak antar ciri antara kelas yang berbeda, dan proses klasifikasi menggunakan *Fuzzy C-Mean Clustering*.

Data citra diperoleh dari Klinik Bersalin Permata Bunda Syariah, yang kemudian dilakukan *preprocessing* sehingga diperoleh citra dengan ukuran 100x100.

2. *Linear Discriminant Analysis (LDA)* : ekstraksi ciri

LDA melakukan analisis secara linier yang memiliki representasi sendiri-sendiri (vektor-vektor basis) dari ruang vektor citra rahim berdimensi tinggi, tergantung dari sudut pandang statistiknya.

Dengan memproyeksikan vektor citra rahim ke vektor basisnya akan didapatkan representasi *feature* dari setiap citra rahim. Pengukuran kemiripan kemudian akan dilakukan antara representasi citra rahim dengan citra uji. Representasi di dalam metode ini dianggap sebagai sebuah transformasi linier dari vektor citra asal ke dalam sebuah ruang proyeksi (vektor-vektor basis).

$$Y = W^T X \quad (1)$$

Dimana Y adalah matriks vektor *feature* berukuran $d \times N$ dengan d adalah dimensi vektor *feature* dan N adalah jumlah citra USG serta $d \ll N$.

Di dalam perhitungannya LDA menggunakan dua buah matriks sebaran, yaitu matriks S_B dan matriks S_W . kedua matriks tersebut didefinisikan sebagai berikut:

$$S_B = \sum_{i=1}^c (\mu_i - \mu)(\mu_i - \mu)^T$$

$$S_W = \sum_{i=1}^c \sum_{y_k \in y_i} (y_k - \mu_i)(y_k - \mu_i)^T$$

dimana :

$$\mu_i = \frac{1}{N_i} \sum_{k=1}^{N_i} Y_k \quad (2)$$

dimana S_B merupakan sebaran antar kelas dan S_W merupakan sebaran dalam kelas. Dari rumus di atas N_i adalah jumlah vektor latih pada kelas i (C_i), μ didefinisikan sebagai mean global, μ_i adalah vektor rata-rata dari kelas i (C_i), dan y_k merupakan vektor-vektor yang berasosiasi dengan kelas i . Dengan menghitung S_B maka akan diketahui jarak atau pemisah antara vektor rata-rata setiap kelas dengan mean global, sedangkan S_W akan mengukur jarak atau pemisah antara vektor-vektor ciri dengan vektor rata-rata dari kelas masing-masing.

Setelah didapatkan matriks SB dan Sw kemudian LDA akan mencari proyeksi yang optimal pada kondisi dimana rasio antara matriks sebaran antar kelas (SB) dari sampel yang diproyeksikan dengan matriks sebaran dalam kelas (Sw) dari sampel yang diproyeksikan, yang dirumuskan sebagai berikut:

$$W_{opt} = \arg \max \frac{W^T S_B W}{W^T S_W W} = [w_1 w_2 \dots \dots \dots w_m] \quad (3)$$

dimana $\{w_i \mid i = 1, 2, \dots, l\}$ merupakan kumpulan vektor eigen dari SB dan Sw yang berkorespondensi dengan eigen value terbesar $\{\tilde{e}_i \mid i = 1, 2, \dots, l\}$. dimana $\{w_i \mid i = 1, 2, \dots, l\}$ merupakan kumpulan vektor eigen dari SB dan Sw yang berkorespondensi dengan eigen value terbesar $\{\tilde{e}_i \mid i = 1, 2, \dots, l\}$.

Dari rumus di atas maka didapatkan hasil yang maksimal yaitu dimana persebaran dalam kelas akan diminimalisasi dan persebaran antar kelas akan dimaksimalkan.

Kemudian representasi akhir LDA adalah zi yang seharusnya dapat menghasilkan separabilitas yang lebih baik dari citra rahim. Dimana zi didefinisikan sebagai berikut :

$$Z_i = W^T y_i \quad (4)$$

Z_i merupakan matriks terproyeksi dengan ukuran $d \times N$. Dengan d merupakan dimensi dari ruang fisher dan N merupakan jumlah citra rahim serta $d \ll N$.

3. Fuzzy C-Mean Clustering : klasifikasi

Fuzzy C-means Clustering (FCM), atau dikenal juga sebagai Fuzzy ISODATA, merupakan salah satu metode clustering yang merupakan bagian dari metode Hard K-Means. FCM menggunakan model pengelompokan fuzzy sehingga data dapat menjadi anggota dari semua kelas atau cluster terbentuk dengan derajat atau tingkat keanggotaan yang berbeda antara 0 hingga 1.

Tingkat keberadaan data dalam suatu kelas atau cluster ditentukan oleh derajat keanggotaannya. Teknik ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981.

Konsep dasar FCM, pertama kali adalah menentukan pusat cluster yang akan menandai lokasi rata-rata untuk tiap-tiap cluster. Pada

kondisi awal, pusat cluster ini masih belum akurat. Tiap-tiap data memiliki derajat keanggotaan untuk tiap-tiap cluster. Dengan cara memperbaiki pusat cluster dan nilai keanggotaan tiap-tiap data secara berulang, maka dapat dilihat bahwa pusat cluster akan menuju lokasi yang tepat. Perulangan ini didasarkan pada minimasi fungsi obyektif (Gelley, 2000).

Berikut beberapa langkah konsep FCM :

- (i) Tentukan beberapa parameter berikut ini :
 - a. Error maksimum
 - b. P awal
 - c. Iterasi awal (t)
 - d. t max
 - e. Pangkat (w)
 - f. Cluster

- (ii) Bentuk matrik U dimana ukuran matrik $[n \times c]$, dengan $n = \sum$ dokumen dan $c = \sum$ cluster. Nilai setiap koordinat matrik U awal adalah nilai random yang didapat dari sistem. Namun jika kita jumlahkan nilai-nilai dari sebuah row yang sama maka kita akan mendapatkan hasil sama dengan 1. Nilai dari setiap koordinat yang awalnya random akan terus di-update sampai mendapat nilai tepatnya, dengan rumus berikut :

$$Q_i = \sum_{k=1}^c \mu_{ik} \quad \text{dan} \quad \mu_{ik} = \frac{\mu_{ik}}{Q_i} \quad (5)$$

- (iii) Bentuk matrik V, yaitu matrik pusat cluster. Matrik V adalah matrik $[k \times j]$, $k = \sum$ cluster dan $j = \sum$ kata. Berikut adalah rumus untuk membangun matrik V :

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \quad (6)$$

- (iv) Fungsi objektif pada iterasi ke-t :

$$P_t = \sum_{i=1}^n \sum_{k=1}^c ([\sum_{j=1}^m (X_{ij} - V_{kj})]^2 (\mu_{ik})^w) \quad (7)$$

- (v) Cek juga perubahan matrik partisi dengan rumus berikut :

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})]^{-\frac{2}{w-1}}}{\sum_{k=1}^c [\sum_{j=1}^m (X_{ij} - V_{kj})]^{-\frac{2}{w-1}}} \quad (8)$$

- (vi) Cek kondisi berhenti, jika $(| \mu_{ik} - \xi | < \epsilon)$ atau $(t > t_{max})$ maka berhenti. Jika tidak : $t=t+1$ dan ulangi dari langkah 4.

4. Hasil Penelitian

Jumlah data citra yang digunakan dalam penelitian ini berjumlah 167 citra *Ultrasonography* (USG). Ukuran citra yang digunakan 100x100 pixel, yang kemudian di *grayscale* dan menjadi *input* dari sistem.

Dalam pengujian sistem, 167 data citra dibagi menjadi 2 bagian yaitu citra latih dan citra uji.

Pada saat pengukuran performansi sistem dalam pengenalan citra USG, pembobot (w) yang digunakan adalah : 2, 3, 4, 5, dan 6, dan nilai iterasi adalah 100.

Tabel 1. Hasil pengujian untuk citra 9 citra normal dan 9 citra PCOS.

Jumlah Data Normal	Jumlah Data PCOS	Iterasi	W Terbaik	Akurasi
3	3	2	2	55,56%
5	5	2	5	72,22%
10	20	2	2	94,44%

Tabel 1 merupakan hasil pengujian dengan nilai akurasi terbesar untuk tiap-tiap parameter yang di ujikan, yaitu pembobot (w), nilai iterasi, dan jumlah data latih.

Dari hasil pengujian terlihat bahwa perbedaan jumlah data dapat mempengaruhi tingkat akurasi dan performansi dari sistem. Ketika jumlah data latih normal berjumlah 3 dan data latih PCOS berjumlah 3, akurasi yang dihasilkan hanya mencapai 55,56%.

Sedangkan ketika jumlah data latih normal berjumlah 10 dan data latih PCOS berjumlah 20, akurasi yang dihasilkan bisa mencapai 94,44%. Hal ini menunjukkan bahwa jumlah data latih dapat mempengaruhi tingkat akurasi dan performansi dari sistem.

Perubahan akurasi sistem berdasarkan jumlah data tidak terlepas dari pengaruh w yang digunakan.

5. Kesimpulan

Metode *Linear Discriminant Analysis* (LDA) sebagai ekstraksi ciri dan *Fuzzy C-Mean Clustering* sebagai klasifikasi data dapat diimplementasikan dalam sistem deteksi gangguan *Polycystic Ovary Syndrom* (PCOS). Pemilihan nilai w dan jumlah data pelatihan sangat berpengaruh terhadap akurasi. Sehingga perlu dilakukan pemilihan nilai w yang tepat untuk mendapatkan performansi yang optimal.

Referensi

- Laboratorium Klinik Prodia. USG. Diperoleh dari <http://prodia.co.id/pemeriksaan-penunjang/usg> (Diakses 3 Oktober 2013)
- Santosa Budi. *DATA MINING TERAPAN DENGAN MATLAB*. Yogyakarta: Graha Ilmu, 2007.
- Santosa, Budi, "Data mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis", Graha Ilmu : Bandung, 2007.
- Lim, Resmana, "Face Recognition Menggunakan Metode LINEAR DISCRIMINANT ANALYSIS (LDA)". 2002, Universitas Gunadarma. Jakarta
- Wijaya, Marvin.Ch. & Agus Prijono. *Pengolahan Citra Digital Menggunakan Matlab*, Bandung: Informatika, 2007
- Raymond, *Face Recognition Menggunakan Metode Linear Discriminant Analysis (LDA)* [online]. Tersedia : <http://dewey.petra.ac.id>. [8 November 2013]
- Yang, Yong. 2007. *Image Segmentation by Fuzzy C-Means Clustering Algorithm with a Novel Penalty Term*. Tersedia di <http://www.cai.sk/ojs/index.php/cai/article/download/296/228> diunduh pada tanggal 7 November 2013