

Representasi Kode IRMA pada Basis Data Mammografi MIAS

Karmilasari¹, Suryarini Widodo², Lussiana ETP³

^{1,2}Jurusan Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Gunadarma
Jl. Margonda Raya No. 100 Pondok Cina Depok, Telp. 021-78881112
{karmila,srini}@staff.gunadarma.ac.id

³Program Studi Sistem Informasi, STMIK Jakarta STI&K
Jl. BRI Radio Dalam No. 17 Kebayoran Baru – Jakarta Selatan 12140, Telp. 021-7210722
lussiana@jak-stik.ac.id

Abstrak

Keterbatasan basis data mammografi yang memiliki pengkodean citra dan adanya karakteristik identifikasi yang beragam, seperti patologi, tipe jaringan payudara dan abnormalitas, merupakan permasalahan dalam pengembangan sistem komputer untuk diagnosis kanker payudara. Sistem pengkodean IRMA dikembangkan untuk memudahkan temu kenali citra berbasis konten (CBIR) sebagai aplikasi prototipe diagnostik medis pada citra radiologi. Kode IRMA dikembangkan mengikuti kode jaringan American College of Radiology (ACR) dan sistem data (BI-RAD). Melalui kode IRMA, diperoleh standarisasi kode untuk tipe jaringan, tingkatan tumor dan deskripsi lesi. Hasil seluruh kode berupa string karakter tidak lebih dari 13 karakter (IRMA: TTTT - DDD - AAA - BBB). Kode dapat diperluas dengan memperkenalkan karakter dalam kode posisi tertentu jika ada modalitas baru yang diperkenalkan. Sistem pengkodean IRMA dapat diterapkan pada Mammogram Digital Mammographic Image Analysis Society (MIAS). Informasi awal yang lengkap dari mammografi menjadi dasar bagi penelitian citra medis kanker payudara, sedangkan informasi akhir yang diperoleh dari sistem pengkodean IRMA dapat menjadi masukan bagi dokter dalam pengambilan keputusan tindakan bagi pasien.

Kata kunci: Mammografi, Sistem Kode IRMA, Basis Data MIAS

Abstract

Limitations of mammography database with image coding and the identification of a variety of characteristics, such as pathology, and abnormal breast tissue types, is an issue in the development of computer systems for the diagnosis of breast cancer. IRMA coding system was developed to facilitate content-based image retrieval identify (CBIR) as a prototype application in medical diagnostic radiology imagery. IRMA Code was developed following the network code American College of Radiology (ACR) and data system (BI-RAD). Through IRMA code, obtained standardized code for the type of tissue, the level of tumor and lesion description. The results of the code in the form of a character string of no more than 13 characters (IRMA: YYYY - DDD - AAA - BBB). The code can be extended by introducing characters in certain positions code if there is a new modality is introduced. IRMA coding system can be applied to mammographic Digital Mammogram Image Analysis Society (MIAS). Complete initial information from mammography is the basis for the study of medical image breast cancer, while the final information obtained from IRMA coding system can be input for clinicians in decision-making for patient action.

Keywords : Mammography, IRMA coding system, MIAS database

1. Pendahuluan

Survey Globocan 2008 yang dilakukan oleh WHO pada tahun 2008, menyebutkan bahwa kanker payudara menempati urutan pertama (13,7%) kasus kanker dan penyebab kematian nomor satu akibat kanker bagi wanita di dunia. Gejalanya adalah timbulnya nodul atau tumor pada payudara. Deteksi dini adalah cara paling efektif untuk menekan jumlah kasus dan angka kematian. Mammografi merupakan metode terbaik dari skrining untuk kanker payudara karena dapat menunjukkan lesi pada tahap awal. Diagnosis lesi dibuat oleh ahli radiologi. Dengan kemajuan teknologi komputer saat ini, mammografi telah dapat dikonversi ke dalam bentuk digital. Hal ini meningkatkan minat akan sistem komputer sebagai alat bantu diagnosis [2].

Suatu sistem diagnosis kanker berbantuan komputer adalah sistem yang dapat mengidentifikasi posisi, ukuran dan tingkatan lesi seperti *microcalcification* (pengapuran) dan massa pada sinar x-mammografi, serta dapat mengevaluasi sejumlah besar citra referensi yang telah disetujui.

Salah satu basis data mammografi yang banyak digunakan oleh peneliti adalah basis data mammogram digital Mammographic Image Analysis Society (MIAS). Beberapa penelitian terdahulu menggunakan sejumlah kecil set data MIAS. Zwiggelaar dkk. menggunakan 15 buah data sintetis

mammografi dari Mammographic Image Analysis Society (MIAS), untuk mendeteksi struktur linier dan pengklasifikasian kanker ke dalam jenis anatomi : pembuluh, saluran dan spikula [13]. Christoyianni dkk menggunakan 22 citra dari basis data MIAS untuk mendeteksi lokasi dari massa yang diindikasikan kanker [3].

Secara umum penanganan dan pengelolaan citra medis dalam jumlah besar cukup sulit. Penelitian menghadapi masalah dalam pemilihan kasus yang tepat dan prosedur penyajian hasil perhitungan visual yang transparan dan *user-friendly*. Beberapa metode manajemen data visual yang saat ini sedang dikembangkan diantaranya adalah temu kenali citra berbasis konten / *content based image retrieval* (CBIR) dan penambangan data citra / *image data mining* [6], tinjauan CBIR untuk citra medis dilakukan oleh Smeulders dkk [10] dan Muller dkk [8].

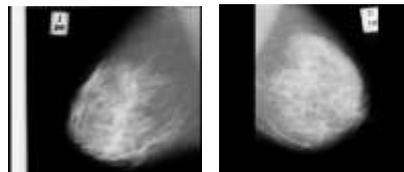
Proyek yang dilakukan oleh *Image Retrieval Medical Application* (IRMA) [7] mendefinisikan struktur basis data terpadu dan skema pengkodean untuk radiografi mammografi terkait dengan diagnostik informasi yang konsisten dengan citra dari berbagai basis data yang tersedia serta mamografi data dari *Picture Archiving and Communication Systems* (PACS) yang biasa digunakan di rumah sakit dan institusi kesehatan. Referensi basis data terpadu digunakan untuk evaluasi sistem diagnostik berbantuan komputer.

Penelitian ini bertujuan merepresentasikan kode IRMA pada basis data mammografi MIAS. Hasil penelitian ini dapat digunakan bagi penelitian yang membutuhkan data mammografi dengan kode yang telah distandarisasi.

2. Metode Penelitian

2.1. Basis Data MIAS

Basis data MIAS [11] tersedia untuk tujuan penelitian dan berisi 322 citra mamografi yang diambil dari sudut pandang mediolateral. Pada awalnya citra dipindai dari film dengan resolusi 50 mikron. Selanjutnya semua citra dipotong dan disimpan dalam matriks 1.024 x 1.024. File citra yang terbentuk berformat *Portable Network Graphics* (PNG) dan dianotasi dengan rincian : referensi basis data dengan nomor yang menunjukkan payudara kiri dan kanan, karakter dari latar belakang jaringan, patologi, kelas lesi dan koordinat serta ukuran lesi. Gambar 1 menunjukkan contoh citra mamografi yang terdapat pada basis data MIAS.



Gambar 1. Citra mamografi MIAS

Tabel 1 menunjukkan informasi mengenai resolusi dan tipe citra basis data, sedangkan tabel 2 menunjukkan anatomi, arah dan informasi biosistem

Tabel 1. Resolusi dan tipe citra basis data MIAS

Resolusi	x-min	334
	x-max	1,000
	y-min	802
	y-max	1,024
	g-min	8 bit
	g-max	8 bit
Tipe File	Format	PNG
	Standar	yes

Tabel 2. Anatomi, arah dan informasi biosistem

Anatomi	Kiri	161
	Kanan	161
Arah	CC	0
	ML	322
Tipe jaringan		Yes
Tingkatan Tumor		Yes
Deskripsi lesi	Posisi	Lingkaran
	Multiple	yes

2.2. Sistem IRMA

Proyek IRMA bertujuan untuk mengembangkan dan menerapkan metode tingkat tinggi untuk temu kenali citra berbasis konten (CBIR) dengan aplikasi prototipe untuk diagnostik medis pada citra radiologi [7]. Di luar citra mammografi, saat ini ada lebih dari 20.000 citra diagnostik dalam basis data IRMA, yang digunakan untuk temu kenali citra [6] [4].

Dalam IRMA, semua citra akan dikodekan dengan mono-hirarkis, skema kode multi-aksial [8]. Empat sumbu, masing-masing memiliki tiga hingga empat posisi hirarkis, yaitu :

- Teknis (T): modalitas citra.
Kode teknis IRMA untuk mamografi adalah TTTT = 11xx, dimana 11 berarti x-ray, radiografi, dan dua posisi kode yang tersisa digunakan untuk menangkap sifat citra (1 = digital secara langsung, 2 = didigitalkan sekunder) dan resolusi (misalnya, 42, 43, 5, 50, atau 200 mikron).
- Arah (*direction/D*): orientasi tubuh.

Skema kode sebagai petunjuk pencitraan payudara, yaitu CC (*craniocaudal*) dan ML (*mediolateral*), dinotasikan dengan DDD = 310 (*aksial - craniocaudal - unspecified*) dan DDD = 410 (*orientasi lain - miring - unspecified*).

- Anatomi (A): wilayah tubuh yang diperiksa.
 Sumbu anatomi kode IRMA digunakan untuk membedakan payudara kanan dengan menggunakan Kode AAA = 610 (payudara – payudara kanan - *unspecified*) dan AAA = 620 (payudara - payudara kiri - *unspecified*).
- Biosistem (B): sistem biologis yang diperiksa.
 Sumbu biosistem diperluas untuk menjelaskan kepadatan jaringan, tingkatan tumor, dan deskripsi lesi (Tabel 3). Posisi pertama menjelaskan jenis jaringan menurut kelas ACR. Misalnya, jika payudara hampir sepenuhnya lemak atau tersebar dengan kepadatan kelenjar fibro, maka ACR-nya adalah ACR-1 atau ACR-2. Posisi kedua kode IRMA menjelaskan tingkatan tumor sesuai dengan laporan ACR citra payudara dan sistem data (BI-RAD) [1], jika ada keperluan untuk penambahan kode evaluasi payudara (BI-RAD 0). Posisi kode ketiga mengacu pada tipe lesi. Sistem BI-RAD, menyediakan delapan deskripsi, seperti tampak pada Tabel 3 yang dapat disesuaikan dengan struktur kode IRMA dan dapat diperluas sesuai kebutuhan.

Hasilnya adalah string unik dengan 13 digit: TTTT-DDD-AAA-BBB

Tabel 3. Kode IRMA untuk biosistem

IRMA code		Tissue density description	IRMA code		Tumor staging	IRMA code BBx	Type of lesion
xBB	ACR		BxB	BI-RADS			
0	<i>unspecified</i>	0	BI-RADS 0	<i>unspecified</i>	1	<i>Calcification, unspecified</i>
1-c	<i>already in use</i>	1	BI-RADS 1	<i>normal</i>	2	<i>Micro-calcification</i>
d	ACR-1	<i>fat transparent system</i>	2	BI-RADS 2	<i>benign</i>	3	<i>Macro-calcification</i>
e	ACR-2	<i>fibroid glands system</i>	3	BI-RADS 3	<i>probably benign</i>	4	<i>Circumscribed mass</i>
f	ACR-3	<i>heterogeneously dense system</i>	4	BI-RADS 4	<i>suspiciously abnormal</i>	5	<i>Spiculated mass</i>
g	ACR-4	<i>extremely dense system</i>	5	BI-RADS 5	<i>malignant</i>	6	<i>Other mass</i>
h	ACR-3/4	<i>dense system</i>				7	<i>Architectural distortion</i>
						8	<i>Aysmmetry</i>

Kerangka IRMA menyediakan basis data citra berikut fitur-fiturnya. Misalnya, tanda tekstur pada citra terkait memungkinkan akses temu kenali citra berbasis konten menjadi cepat. Secara internal, keterkaitan ini dibentuk berdasarkan identifier IRMA, yang unik untuk tiap elemen basis data. Hubungan ini digunakan untuk menyimpan satu atau lebih deskripsi lesi, dalam bentuk :

1. Lingkaran, yang digambarkan oleh koordinat pusat dan jari-jari;
2. Titik kontur, yaitu daftar koordinat (x, y);
3. Rantai kode, koordinat awal (x, y) diikuti dengan urutan angka dalam [1 .. 8] dan menjelaskan arah ke titik terdekat kontur, atau
4. Masking/topeng citra, citra biner dengan dimensi x-, y- yang sama dengan mammografi, di mana 0 dan 1 menunjukkan "latar belakang" dan "lesi".

2.3. Integrasi Basis Data MIAS ke dalam Sistem Pengkodean IRMA

Pengkodean dilakukan secara manual sesuai dengan deskripsi basis data di Internet. Awalnya ACR jenis jaringan diubah sebagai berikut:

- F (MIAS fat/lemak) → d (ACR-1);
- G (MIAS fibroid-glands/lemak-kelenjar) → e (ACR-2);
- D (MIAS dense/padat-kelenjar) → h (ACR-3/4).

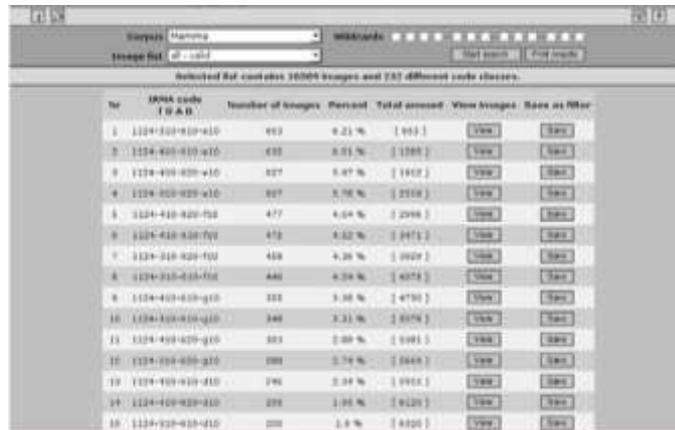
Namun, konversi ini tidak sepenuhnya menghasilkan kode ACR yang benar. Dari 322 citra, 187 harus dikoreksi oleh seorang ahli radiologi yang berpengalaman (Tabel 4). Selanjutnya, tiga citra dikoreksi pada klasifikasi BI-RAD dari normal jinak menjadi mikro-kalsifikasi.

Tabel 4. Klasifikasi jaringan koreksi MIAS

Code alteration	Description	Number
ACR 1 → ACR 2	fatty → fibroid glands	24
ACR 1 → ACR 3	fatty → heterogeneously dense	1
ACR 2 → ACR 3	fatty-grandular → heterogeneously dense	45
ACR 2 → ACR 4	fatty-grandular → extremely dense	4
ACR 3/4 – ACR 2	dense-grandular → fibroid glands	5
ACR 3/4 – ACR 3	dense-grandular → heterogeneously dense	38
ACR 3/4 – ACR 4	dense-grandular → extremely dense	70
Total		187

3. Hasil dan Analisis

Gambar 2,3 dan 4 adalah tampilan sistem IRMA dengan basis data MIAS. Gambar. 2 menunjukkan tampilan sistem IRMA dengan semua kode kelas yang tersedia. Melalui tombol "view", tampilan ini terhubung ke *interface* yang menampilkan citra dan kode IRMA dari citra basis data MIAS (Gambar 3) dan hasil pencarian tampak seperti pada gambar 4.



No	IRMA code (TAB)	Number of Images	Percent	Total amount	View Images	View as Filter
1	1124-110-110-w10	493	0.21 %	(493)	[View]	[View]
2	1124-110-110-w10	433	0.21 %	(1086)	[View]	[View]
3	1124-110-110-w10	327	0.17 %	(1412)	[View]	[View]
4	1124-110-110-w10	327	0.16 %	(2218)	[View]	[View]
5	1124-110-110-w10	477	0.14 %	(2098)	[View]	[View]
6	1124-110-110-w10	473	0.12 %	(2471)	[View]	[View]
7	1124-110-110-w10	459	0.10 %	(2004)	[View]	[View]
8	1124-110-110-w10	440	0.09 %	(1674)	[View]	[View]
9	1124-110-110-w10	351	0.10 %	(4791)	[View]	[View]
10	1124-110-110-w10	349	0.11 %	(2076)	[View]	[View]
11	1124-110-110-w10	333	0.09 %	(1081)	[View]	[View]
12	1124-110-110-w10	309	0.14 %	(2449)	[View]	[View]
13	1124-110-110-w10	296	0.14 %	(1914)	[View]	[View]
14	1124-110-110-w10	279	0.05 %	(6126)	[View]	[View]
15	1124-110-110-w10	279	0.1 %	(8126)	[View]	[View]

Gambar 2. Tampilan sistem IRMA untuk kode statistik.



Gambar 3. Tampilan sistem IRMA untuk mencari citra mammografi



Gambar 4. citra "Mamma" dengan informasi diagnostik lebih rinci.

Skema kode IRMA memungkinkan klasifikasi citra berkaitan dengan empat hal yaitu teknis, arah, anatomi, dan kriteria biologis. Berdasarkan strukturnya, kode IRMA dapat dengan mudah diperluas (menambahkan karakter baru untuk posisi kode tertentu). Sejauh ini, dukungan pengkodean tersebut mudah digunakan pada sekitar 6.000 citra, waktu pemanggilan citra untuk citra berukuran kecil (sekitar 200 piksel) cepat, dukungan dari *javascripts* dalam skema pengkodean sangat membantu. Terdapat kurang dari 0,1% kode regional yang sangat terbatas untuk menggambarkan bagian-bagian tubuh pada radiograf. Kode regional memungkinkan pengkodean bagian tubuh yang sangat umum atau yang sangat spesifik.

Beberapa penelitian mengusulkan untuk membangun database mamografi dalam format XML dengan citra dalam jumlah besar dengan menambahkan informasi tentang pasien, seperti usia, riwayat menstruasi dan kejadian kanker payudara di keluarga. Dengan database IRMA, sampai saat ini semua

informasi yang dibutuhkan oleh ahli radiologi dapat membantu temu kenali dan sistem computer-aided diagnosis (CAD).

4. Simpulan

Meskipun evaluasi analisis citra medis harus mengandalkan set data penelitian yang besar, penelitian di bidang komputer untuk membantu membaca mammografi seringkali didasarkan pada data dengan jumlah kurang dari 30 citra. Berdasarkan kerangka IRMA, dapat didefinisikan sebuah kode skema sesuai dengan standar ACR dan menunjukkan bagaimana sumber daya citra yang ada terintegrasi. Kode IRMA menyediakan skema untuk mengintegrasikan database mamografi yang tersedia menggunakan deskripsi standar dari modalitas citra dan resolusi, orientasi dan pandangan, posisi kiri dan kanan payudara, tipe jaringan, deskripsi tumor dan lesi serta posisi lesi terlepas apakah posisi tersebut dikodekan oleh beberapa titik batas, batas lingkaran, atau kode rantai lengkap. Skema ini didasarkan pada internasional ACR dan kode BI-RADS, dan dapat digunakan untuk evaluasi masa depan sistem CAD. Namun, beberapa masalah muncul dari kenyataan bahwa sumber daya yang ada tidak menyediakan informasi kode yang lengkap, atau menyembunyikan informasi ini di dalam file atau pada tambahan deskripsi file.

skema kode memungkinkan klasifikasi unik untuk citra medis. Hasil seluruh kode dalam string karakter tidak lebih dari 13 karakter. Kode dapat dengan mudah diperluas dengan memperkenalkan karakter dalam kode posisi tertentu, misalnya, jika ada modalitas baru yang diperkenalkan. Berbeda dengan pendekatan lain, perpaduan dari satu dan dua literal posisi kode harus dihindari untuk menyederhanakan pengolahan kode otomatis. Kode IRMA menghindari ambiguitas akibat tumpang tindih elemen kode dalam tingkat yang sama. Dengan mengintegrasikan sumber daya citra berbeda yang tersedia secara bebas di Internet, database IRMA saat ini memiliki 10.509 citra dari 232 kode kelas yang berbeda.

Daftar Pustaka

- [1] American College of Radiology. Breast Imaging Reporting and Data System (BI-RADS®). *Atlas*. 2006.
- [2] Burhenne LJW, Wood SA, D'Orsi CJ, Feig SA, Kopans DB, Castellino RA. "Potential contribution of computer aided detection to the sensitivity of screening mammography". *Radiol*. 2000; 215: 554-62.
- [3] Christoyianni I, Dermantas E, Kokkinakis G. "Automatic detection of abnormal tissue in mammography". *Proceedings ICIP*. 2001; 877-80.
- [4] Deselaers T, Müller H, Clough P, Ney H, Lehmann TM. "The CLEF 2005 automatic medical image annotation task". *Intl J Comp Vis*. 2007; 74(1): 51-8.
- [5] Elter M, Schulz-Wendtlund R, Wittenberg T. "The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process". *Med Phys*. 2007; 34(11): 4164-72.
- [6] Lehmann TM, Güld MO, Deselaers T, Keysers D, Schubert H, Spitzer K, Ney H, Wein BB. "Automatic categorization of medical images for content-based retrieval and data mining". *Comput Med Imaging Graph*. 2005; 29(2): 143-155.
- [7] Lehmann TM, Güld MO, Thies C, Fischer B, Spitzer K, Keysers D, Ney H, Kohnen M, Schubert H, Wein BB. "Content-based image retrieval in medical applications". *Methods Inform Med*. 2004; 43(4): 354-61.
- [8] Lehmann TM, Schubert H, Keysers D, Kohnen M, Wein BB. "The IRMA code for unique classification of medical images". *Proc SPIE*. 2003; 5033:440-51.
- [9] Muller H, Michoux N, Bandon D, Geissbuhler A. "A review of content-based image retrieval systems in medical applications : Clinical benefits and future directions". *Int J Med Inform*. 2004; 73(1): 1-23.
- [10] Smeulders AWM, Worring M, Santini S, Gupta A, Jain R. "Content-based image retrieval at the end of the early years". *IEEE Trans Pattern Anal Mach Intell*. 2000; 22(12): 1349-80.
- [11] Suckling J, et al. "The Mammographic image analysis society digital mammogram database", *Exerpta Medica International Congress*. 1994; 1069: 375-8.
- [12] World Health Organization. *GLOBOCAN 2008*. International Agency for Research on Cancer. 2008.
- [13] Zwiggelaar R, Astley SM, Boggis CRM, Taylor CJ. "Linear structures in mammographic images: detection and classification". *IEEE Trans Med Imaging*. 2004; 23(9): 1077-86.