

Sistem Dapatan Semula Imej untuk Aplikasi Perubatan

(Content-based Image Retrieval for Medical Application)

Wan Siti Halimatul Munirah Wan Ahmad^a, Wan Mimi Diyana Wan Zaki^{a,*}, Aini Hussain^a, Ling Chei Siong^a

^aPusat Kejuruteraan Sistem Bersepadu dan Teknologi Termaju (INTEGRA),
Fakulti Kejuruteraan dan Alam Bina, Universiti Kebangsaan Malaysia, Bangi, Malaysia
Wong Erica Yee Hing^b

^bJabatan Radiologi, Pusat Perubatan Universiti Kebangsaan Malaysia, Cheras, Malaysia

ABSTRAK

Dapatan semula imej (DSI) adalah sistem pencarian imej yang menggunakan ciri-ciri tertentu atau konteks khusus dalam sesuatu imej. Dalam bidang perubatan, sistem DSI digunakan untuk menyediakan imej yang diperlukan secara tepat dan pantas kepada pakar perubatan. Proses itu biasanya berlaku pada dan ketika diagnosis dan rawatan penyakit dilakukan. Sistem dapatan semula yang awal dan masih digunakan dengan meluas dalam bidang perubatan adalah sistem DSI berdasarkan teks (TBIRS). TBIRS menggunakan kata kunci dalam konteks sesuatu imej dan ia memerlukan anotasi teks secara manual. Proses anotasi teks adalah tugas yang memerihkan lebih-lebih lagi jika melibatkan pangkalan data yang besar. Ini memungkinkan kebarangkalian berlakunya kesilapan manusia adalah tinggi. Untuk mengatasi masalah yang dinyatakan, sistem DSI berdasarkan kandungan (CBIRS) dengan pengindeksan automatik adalah dicadangkan. Kaedah ini melibatkan pemprosesan imej perubatan berdasarkan komputer yang menggunakan fitur visual imej seperti warna, bentuk dan tekstur. Namun begitu, umum mengetahui bahawa suatu algoritma tertentu dalam CBIRS adalah khusus untuk satu modaliti sahaja dan melibatkan bahagian yang tertentu. Ini ditambahkan pula bahawa CBIRS telah mengabaikan persepsi manusia dalam tugas menakrif sesuatu imej dan akibatnya, menyebabkan wujudnya masalah jurang semantik. Oleh itu, sistem DSI hibrid (HBIRS) yang menggabungkan kekuatan kedua-dua TBIRS dan CBIRS telah diperkenalkan bagi menangani masalah jurang semantik khususnya dan sekaligus memantapkan sistem DSI amnya. Satu kerangka sistem DSI yang cekap iaitu HBIRS juga telah dicadangkan. Walau bagaimanapun, kajian ini hanya melibatkan TBIRS dan CBIRS bagi aplikasi perubatan, dan prototaip TBIRS yang dikaji menggunakan imej X-Ray turut dicadangkan.

Kata kunci: Imej perubatan; sistem dapatan semula imej; aplikasi perubatan; modaliti x-ray; TBIRS

ABSTRACT

Image retrieval is an image searching system that uses certain characteristics or context in an image. In a medical field, image retrieval system has been used to provide the needed correct images to the physicians while the diagnosis and treatment process is being conducted. The earlier image retrieval system was a text-based image retrieval system (TBIRS) which using keyword for the image context and it requires human's help to manually make text annotation on the images. This system is still being widely used in the hospitals nowadays. The text annotation process is a laborious task especially when dealing with a huge database and is prone to human errors. To overcome the aforementioned issues, the approach of a content-based image retrieval system (CBIRS) with automatic indexing using visual features such as color, shape and texture is introduced. However, it is well known fact that a specific algorithm in CBIRS is only applicable to one specific modality and body part. In addition, CBIRS ignores a human perception of an image semantic that leads to a semantic gap problem. Hence, hybrid based image retrieval system (HBIRS) which combines both TBIRS and CBIRS has been proposed to reduce the semantic gap. A framework of efficient retrieval system which is HBIRS are presented, however the development of HBIRS could not be implemented due to the weaknesses in TBIRS. Hence, this study only considers both TBIRS and CBIRS for medical applications, and a TBIRS prototype tested using various X-Ray images has been proposed.

Keywords: Medical images; content-based image retrieval; medical applications; x-ray modality; TBIRS

PENGENALAN

Imej perubatan memainkan peranan yang penting dalam diagnosis, kajian perubatan dan pendidikan perubatan (Bhadoria and Dethé 2010). Kajian imej dalam bidang perubatan merupakan bidang kajian yang sangat aktif dan

pelbagai kaedah untuk analisis imej-imej perubatan telah diperkenalkan (Pilevar 2011). Contohnya, imej perubatan amat penting kerana ia boleh membekalkan maklumat yang berguna untuk diagnosis, rawatan dan pengurusan penyakit dengan berkesan dalam masa yang singkat. Perkembangan teknologi yang semakin pesat pada masa

kini telah menyebabkan peningkatan penggunaan imej-imej perubatan digital dalam salinan lembut dengan mendadak. Menurut (Müller et al. 2010), Jabatan Radiologi dari hospital Universiti Geneva, sebanyak 114,000 imej perubatan telah dihasilkan dalam masa sehari pada tahun 2009 dan telah mencecah 200,000 imej sehari pada tahun 2011 (Herrera De et al. 2013; MedGift 2017). Penghasilan imej perubatan meningkat dengan mendadak secara berterusan dalam masa sepuluh tahun yang lepas. Satu cara yang cepat dan efektif untuk mendapatkan semula imej dalam pangkalan data yang besar amat diperlukan (Khan et al. 2012). Oleh itu, sistem dapatan semula imej (DSI) memainkan peranan yang penting dalam mengatasi masalah tersebut.

DSI ialah proses melayari, mencari, dan mendapatkan semula imej dari pangkalan data imej yang besar (Murthy et al. 2010; Ho et al. 2012) dengan menggunakan kata-kata kunci atau fitur-fitur imej tersebut (Suganya and Rajaram 2012). Teknologi DSI yang terawal adalah berdasarkan teks (Goodrum 2000; Ahmad 2008). Manakala artikel pertama sistem dapatan semula berdasarkan kandungan iaitu pertanyaan melalui kandungan imej (QBIC) telah pun diterbitkan pada tahun 80-an. (Zhituo et al. 2012; Chuctaya et al. 2012).

Sistem DSI merupakan sistem yang biasa digunakan sekarang tetapi penggunaannya masih tidak meluas dalam bidang perubatan. Dalam bidang perubatan, sistem ini bertujuan untuk menyediakan imej yang diperlukan oleh pakar perubatan pada dan ketika proses diagnosis dan rawatan penyakit dilakukan. Walaupun sistem dapatan semula imej dalam perubatan (DSIP) masih mempunyai banyak kelemahan, banyak kajian-kajian penambahbaikan telah dijalankan secara aktif bagi meningkatkan keupayaan sistem ini untuk menguruskan pangkalan data imej perubatan dengan lebih cekap (Pourghassem and Ghassemian 2008; Amaral et al. 2010; Grace et al. 2014; Arthi and Vijayaraghavan 2013; Jaganathan and Vennila 2013; Syam et al. 2013; Chuctaya et al. 2012; Nakaram and Leauhatong 2012; Ramamurthy and Chandran 2011). Namun, realitinya masih terdapat banyak cabaran yang sedang dihadapi oleh para penyelidik (Müller and Greenspan 2013).

Berbanding dengan sistem bantuan pengkomputeran diagnosis (CADs), sistem DSI amat berbeza dari segi prinsip pembangunannya yang mana ia bertindak menjadi pelengkap kepada CADs dengan menyediakan imej-imej visual yang serupa seperti imej yang ditanya atau diminta oleh pengguna. Kajian oleh Kalpathy-Cramer et al. menyatakan ketepatan keputusan dapatan semula imej boleh dicapai dengan bantuan teknik-teknik yang digunakan oleh CADs (Kalpathy-Cramer et al. 2015). Selain membantu dalam proses diagnosis, sistem DSI juga bertujuan untuk membantu pakar radiologi dengan memberikan maklumat yang sesuai (Ahad et al. 2011) dan berkaitan kes yang sedang dirawat. Sekiranya pemerhati merupakan pengelas, sistem DSI bertindak dengan memberikan contoh-contoh kes yang hampir serupa dengan kes yang sedang dirawat oleh pemerhati. Selain itu, DSI juga amat berguna dalam bidang-bidang yang berkaitan, contohnya seperti berikut:

1. Pengajaran dan penyelidikan – Pencarian sesuatu imej boleh dilakukan dari pangkalan data yang besar untuk digunakan sebagai perbandingan bagi imej yang hampir sama untuk membantu aktiviti pengajaran dan penyelidikan.
2. Rujukan berdasarkan kes – Kes perubatan yang serupa sebagai rujukan diperlukan semasa menjalankan rawatan sesuatu penyakit. Sekiranya beberapa atau lebih banyak imej kes yang serupa boleh digunakan semasa diagnosis, cadangan rawatan atau tindakan susulan yang lebih tepat dapat dilakukan.
3. Patologi, dermatologi atau imbas CT paru-paru yang mempunyai peleraian tinggi – Diagnosis untuk kes-kes ini bergantung kepada ciri-ciri tesktur dan warna imej. Oleh itu, kes yang serupa boleh didapati dari sistem DSI.

Oleh yang demikian, kertas kajian ini akan membentangkan kajian-kajian terdahulu yang telah dilakukan untuk membangunkan sistem DSI yang merupakan suatu alat teknologi pembantu perubatan bagi menangani pangkalan data imej yang besar. Selain itu, prototaip sistem dapatan semula imej perubatan (DSIP) yang telah dibangunkan turut dibentangkan. Sistem ini akan diintegrasikan menggunakan platform android untuk menjadikannya lebih mudah alih dan digunakan.

KAJIAN PERPUSTAKAAN

Penampil imej memainkan peranan yang penting dalam rawatan kesihatan moden. Pembangunan pengimejan digital dan teknologi perangkaan yang pantas telah mendorong penggunaan sistem maklumat visual yang pesat (Syam et al. 2013). Pemprosesan imej adalah salah satu bentuk pemprosesan isyarat di mana input boleh merupakan gambar atau video, dan output adalah imej atau parameter yang berkaitan dengan sesuatu imej. Sistem DSI merupakan suatu sistem yang membolehkan kita melayari, mencari dan mendapat semula imej yang sama atau hampir sama ciri visual dengan imej yang diminta (Arthi et al. 2013).

Penggunaan sistem DSI dalam bidang perubatan merupakan salah satu kajian yang paling popular dalam bidang visi komputer pada dekad yang lepas (Müller et al. 2013; Ghosh et al. 2011). Pangkalan data imej di hospital-hospital telah meningkat dengan mendadak seiring dengan penggunaannya semasa latihan harian, kajian perubatan dan pendidikan. Data-data imej ini telah disimpan dan diarkibkan di pangkalan data imej perubatan yang juga dikenali sebagai Sistem Pengarkiban Gambar dan Komunikasi (PACS) (Arakeri and Reddy 2012). Pakar radiologi sentiasa merujuk dan membandingkan kes-kes yang serupa untuk mendapatkan cara rawatan yang sesuai. Namun, pencarian kes-kes terdahulu yang serupa dalam pangkalan data yang besar adalah sangat leceh dan menyusahkan (Feng et al. 2003). Oleh kerana itu, cara yang efektif diperlukan untuk mengendali dan menguruskan pangkalan data imej yang besar ini.

TBIRS DAN CBIRS

Terdapat dua jenis sistem DSI iaitu DSI berdasarkan teks (TBIRS) dan DSI berdasarkan kandungan (CBIRS) (Wang et al. 2016). *Picture Archiving and Communication System* (PACS) adalah TBIRS yang biasa digunakan dalam hospital dan telah menjadi sistem yang amat penting dalam prosidur klinikal, aktiviti pendidikan dan kajian, penggunaan pangkalan data untuk imej *Digital Imaging and Communication in Medicine* (DICOM) dan DSI (Nakaram et al. 2012). Sistem tersebut melakukan pencarian berdasarkan kata kunci (metadata) yang terkandung dalam imej DICOM seperti nama pesakit, tarikh dan jenis modaliti yang disimpan semasa proses perolehan data dan juga catatan perihai teks oleh pakar-pakar radiologi mengenai kes yang didiagnosis. Perihai teks menghadkan keupayaan sistem pencarian dengan syarat kata kunci mesti diketahui semasa proses pencarian. Sistem ini kurang efektif kerana catatan secara manual boleh menyebabkan kesilapan dan kehilangan data. Catatan oleh setiap pakar radiologi adalah berbeza dan tugas ini amat memerlukan jika melibatkan pangkalan data imej yang besar. Oleh itu, kelemahan-kelemahan dalam TBIRS tersebut telah mendorong ahli-ahli kajian memberikan tumpuan kepada CBIRS (Arakeri et al. 2012; Madugunki et al. 2011).

CBIRS telah dibangunkan untuk mengatasi kelemahan-kelemahan dalam TBIRS. Kini, kajian dalam CBIRS lebih popular berbanding dengan TBIRS (Guo et al. 2016; Arakeri and Ram Mohana Reddy 2013; Madugunki et al. 2011). Menurut (Kumar et al. 2013), staf klinikal memilih kes yang serupa dengan mengutamakan ciri-ciri visual dalam menjalankan diagnosis dan rawatan. CBIRS adalah sistem yang mendapatkan semula imej melalui penyarian fitur-fitur visual imej seperti warna, tekstur dan bentuk. Penyarian fitur boleh dijalankan secara automatik atau separa automatik. Sistem ini mampu mempercepatkan proses pencarian imej di mana ia tidak memerlukan catatan secara manual.

Penyarian fitur merujuk kepada proses mengenal pasti maklumat-maklumat penting dan ciri-ciri dari sesuatu imej melalui pengesanan fitur dan pemerihai fitur. Pengesanan fitur berperanan untuk mengesan fitur-fitur yang hadir pada sesuatu imej dan kesemua fitur yang dikesan akan ditunjukkan dalam nilai angka oleh pemerihai fitur (Yaghoubyan et al. 2016). Fitur boleh dikategorikan kepada dua jenis iaitu fitur global dan fitur tempatan. Fitur global seperti warna, bentuk dan tekstur sesuai digunakan dalam kebanyakan aplikasi, manakala fitur tempatan seperti modaliti imej dan kawasan badan hanya sesuai untuk sebahagian aplikasi sahaja. CBIRS menggunakan fitur global untuk menentukan ciri-ciri subjek yang ingin didapati semula. Fitur global yang paling kerap digunakan dalam CBIRS adalah warna, bentuk dan tekstur (Arthi et al. 2013). Proses pemilihan fitur (Husam et al. 2017) juga boleh digunakan bagi meningkatkan ketepatan ramalan dalam mendapatkan semula imej yang serupa.

Dalam CBIRS, fitur-fitur visual imej pertanyaan akan diekstrak terlebih dahulu dan diikuti oleh pengukuran keserupaan untuk mencari fitur-fitur yang serupa dengan imej

pertanyaan dalam pangkalan data (Syam et al. 2013). Namun begitu, terdapat kekangan yang membataskan keberkesanan CBIRS dalam aplikasi perubatan. Umum mengetahui bahawa suatu algoritma tertentu dalam CBIRS adalah khusus untuk satu modaliti dan bahagian tertentu sahaja. Ini adalah kerana setiap modaliti imej perubatan mempunyai resolusi yang berbeza. Contohnya, antara fitur-fitur yang terdapat dalam imej, fitur warna tidak sesuai digunakan ke atas imej X-Ray disebabkan oleh imej yang berskala kelabu (Silva, Da et al. 2011). Oleh itu, pembangunan CBIRS untuk aplikasi yang spesifik diperlukan (Kaur and Jyoti 2013). Selain itu, penambahbaikan cara penyarian fitur-fitur visual juga diperlukan bagi mengurangkan masa dan ruang untuk penyimpanan fitur-fitur imej (Guo et al. 2016).

Untuk mendapatkan DSI yang lebih efisien, imej perubatan dibahagikan kepada beberapa kategori dengan modaliti imej yang berbeza. Perluasan imej yang baik adalah berdasarkan modaliti, bahagian badan dan orientasi. Proses perluasan adalah penting untuk menapis imej bagi mengurangkan kawasan yang dicari (Pourghassem et al. 2008).

Terdapat beberapa jenis modaliti pengimejan telah dibangunkan untuk sistem DSI, seperti *Computed Tomography (CT)*, *Magnetic Resonance (MR)*, *Ultrasound (US)*, *Mammogram (MG)* and *Digital Radiography (DR)*. Beberapa jenis sistem CBIR yang penting yang telah digunakan dalam bidang perubatan telah dibentangkan (Bhadoria et al. 2010) seperti *Automatic Search and Selection Engine with Retrieval Tools (ASSERT)*, *CasImage*, *medGIFT*, *Image Retrieval in Medical Applications (IRMA)*, *The Second National Health And Nutrition Examination Survey (NHANES II)* dan *Feature Subset Selection using Expectation-maximization Clustering (FSSEM)*. Sesetengah sistem telah berjaya digunakan dalam bidang perubatan tetapi masih lagi mempunyai batasan. Selain itu juga terdapat beberapa sistem dalam talian yang dibincangkan oleh Ghosh et al. dalam kajiannya (Ghosh et al. 2011) iaitu *FigureSearch*, *BioText*, *GoldMiner*, *Yale Image Finder*, *Yottalook*, *Image Retrieval for Medical Applications (IRMA)*, dan *NLM's Medline*. Banyak kajian lain yang turut dilakukan dan kaedah-kaedah yang telah dibuat oleh ahli penyelidik lain dipaparkan di dalam Jadual 1.

Satu cabaran penting yang dihadapi semasa menggunakan sistem CBIR adalah fitur yang digunakan telah mengabaikan persepsi manusia dalam sesuatu imej dan ia dikenali sebagai masalah jurang semantik. Masalah ini telah menyebabkan pembangunan sistem DSI berdasarkan semantik (SBIRS) (Kumar et al. 2013) yang menggabungkan kedua-dua sistem TBIR dan CBIR. Prestasi hibrid adalah lebih baik daripada CBIRS kerana fitur dalam CBIRS hanya memberikan deksripsi imej melalui kandungan visual dan tidak berupaya dalam membezakan imej yang mempunyai fitur yang lebih kurang sama (Hartvedt 2010). Contohnya, dalam membezakan burung dan kapal terbang. Oleh itu, pelbagai sistem DSI hibrid telah diperkenalkan untuk mengatasi masalah jurang semantik dan menapis imej yang tidak diinginkan menggunakan keupayaan daripada TBIRS (Boparai and Chhabra 2015).

JADUAL 1. Kajian yang telah dijalankan oleh ahli penyelidik

Kajian	Kaedah	Modaliti Imej
(Arakeri et al. 2012)	Pengelasan hierarki dan persamaan kandungan	MRI
(Suganya et al. 2012)	<i>SVM</i> digabungkan dengan maklumbalas relevan	Ultrasound
(Nakaram et al. 2012)	<i>Discrete Wavelet Transform (DWT)</i>	X-Ray
(Murala and Jonathan Wu 2013)	<i>Local ternary co-occurrence patterns</i>	MRI dan CT
(Arthi et al. 2013)	<i>CCM (Colour Co-occurrence Matrix)</i> menggunakan peta warna <i>Hue Saturation Value (HSV)</i>	Imej Perubatan
(Grace et al. 2014)	Rangka <i>Apache Hadoop</i>	Imej Perubatan
(Kumar et al. 2014)	Berasaskan kaedah graf menggunakan <i>vertex set</i> dan <i>edge set</i>	PET-CT
(Wan Ahmad et al. 2014)	<i>Gabor transform, Discrete Wavelet Frame, Grey Level Histogram</i> dan kombinasi kaedah ini	X-Ray
(Bergamasco and Nunes 2015)	Fitur global dan tempatan: <i>Distance Histogram Descriptor, Local Distance Histogram Descriptor</i> , dan <i>3D Hough Transform Descriptor</i>	MRI
(Kitanovski et al. 2016)	Berasaskan hasil pengkuantuman dan pengelasan <i>SVM</i>	Imej Perubatan
(Sparks and Madabhushi 2016)	<i>Out-of-Sample Extrapolation</i> menggunakan <i>Semi-Supervised Manifold Learning (OSE-SSL)</i>	Histologi Prostat
(Nowaková et al. 2017)	Pengkuantuman vektor dengan <i>fuzzy S-trees</i>	Mammogram
(Spanier et al. 2017)	Hibrid: gabungan dengan penemuan radiologi dari laporan kes perubatan dan graf <i>Radlex</i>	CT
(Xu et al. 2017)	<i>Bag of visual words (BoVW)</i> dengan maklumat ruang	CT

PROTOTAIP SISTEM DAPATAN SEMULA IMEJ

Bahagian ini memerihalkan kaedah kajian yang digunakan dalam pembangunan sistem DSI untuk aplikasi perubatan. Umum mengetahui bahawa sistem DSI yang efisien adalah HBIRS di mana ia mengintegrasikan dua sistem iaitu CBIRS dan TBIRS. Sebelum membangunkan HBIRS, CBIRS dan TBIRS haruslah dibangunkan terlebih dahulu. Kajian ini hanya memfokuskan pembangunan dan pengujian sistem TBIR sahaja, yang melibatkan tiga modul utama iaitu penyarian metadata, pengurusan pangkalan data dan dapatan semula.

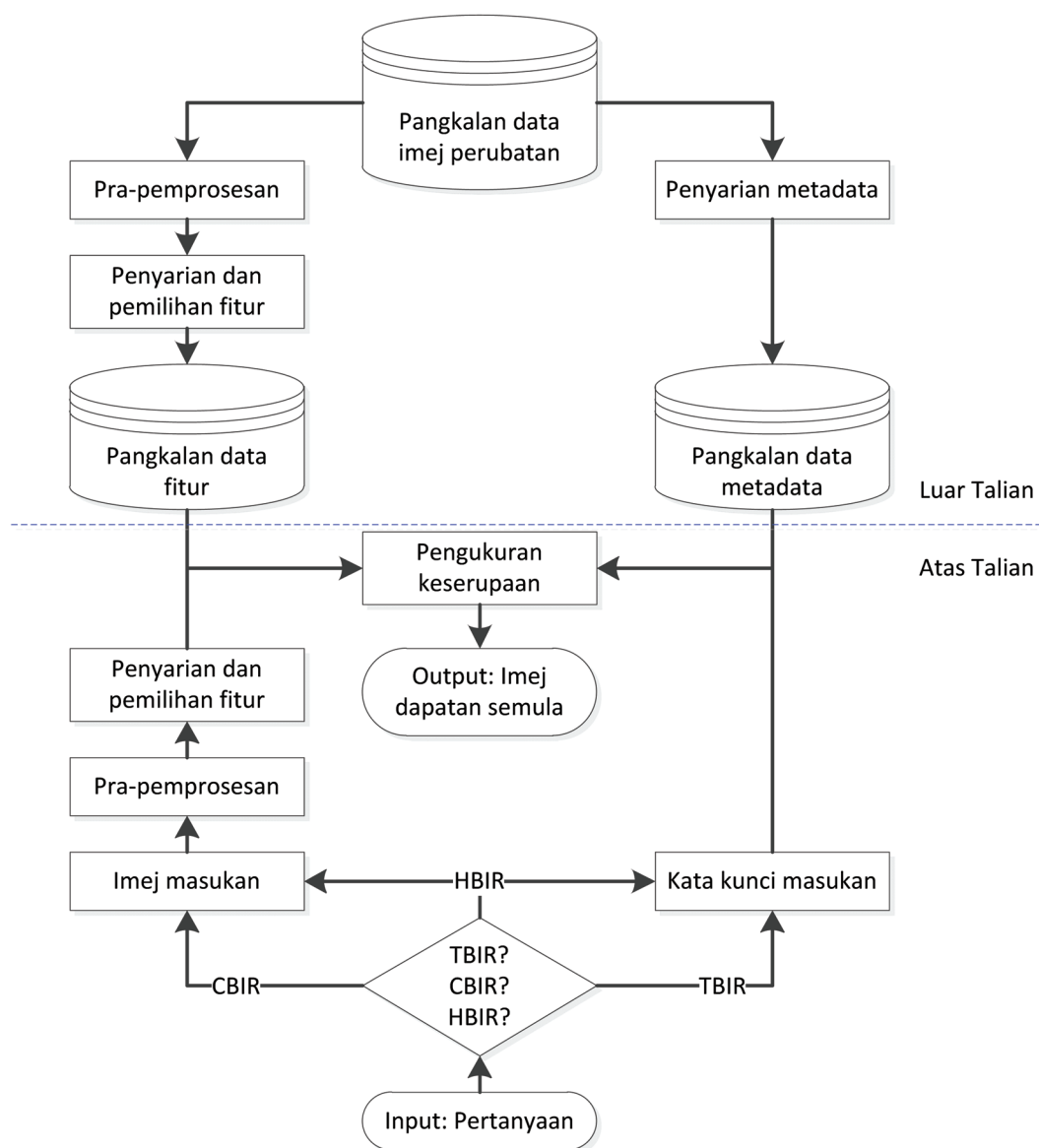
Rajah 1 memaparkan cadangan kerangka kerja sistem DSI perubatan (DSIP) yang terdiri daripada CBIRS dan TBIRS. Pada bahagian CBIRS, fasa pra-pemprosesan dibangunkan bagi memproses imej supaya lebih jelas dan kurang hingar serta meningkatkan kontras imej. Seterusnya, teknik peruasan dilaksanakan untuk meruas ROI sebelum fitur-fitur yang optimum bagi mewakili ROI disari dan dipilih. Fitur-fitur yang telah disari kemudiannya disimpan dalam pangkalan data fitur dan proses ini dilakukan secara luar talian. Sekiranya terdapat pertanyaan dari pengguna untuk mencari imej yang serupa, fasa dapatan semula imej akan menjalankan pengukuran keserupaan ke atas fitur-fitur imej pertanyaan yang dimasukkan dan membandingkannya dengan fitur-fitur dalam pangkalan data yang telah terdahulu disari. Imej-imej serupa dalam pangkalan data sistem ini disusun mengikut tahap keserupaan terdekat dan dipaparkan kepada pengguna. Bagi TBIRS pula, modul penyarian metadata bertujuan untuk mengekstrak maklumat daripada pengepala imej berformat DICOM dalam bentuk deskripsi teks sebelum metadata tersebut dikendali dan disimpan oleh modul pengurusan pangkalan data. Sekiranya terdapat pertanyaan dari pengguna untuk mencari imej yang serupa menggunakan TBIRS, modul dapatan semula imej akan memaparkan imej-imej dalam pangkalan data yang mempunyai persamaan metadata imej yang ditanya oleh pengguna.

Sistem ini juga dapat dibahagikan kepada dua kaedah pengoperasian iaitu secara luar talian dan atas talian. Kaedah secara luar talian merujuk kepada penyarian metadata dan fitur imej-imej yang dilakukan sebelum pencarian oleh pengguna dilakukan dan disimpan dalam pangkalan data, manakala kaedah secara atas talian melibatkan penggunaan aplikasi antara muka pengguna grafik (GUI) sebagai platform oleh pengguna untuk mendapatkan semula imej-imej yang dikehendaki.

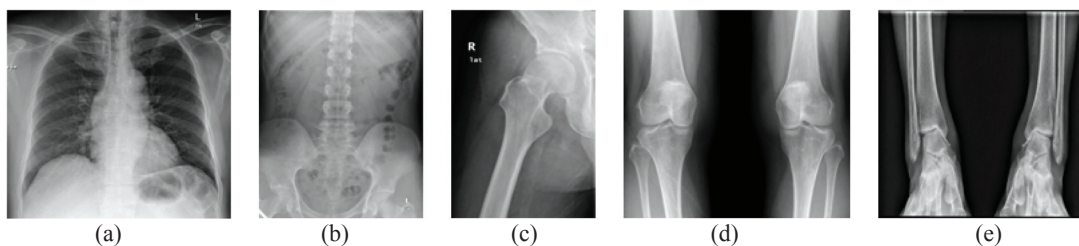
Dalam kajian ini, pembangunan DSI dan semua eksperimen dijalankan menggunakan perisian MATLAB® 2012b dan perkakasan komputer meja berprosesan Intel Xeon CPU E31225 berkelajuan 3.10 GHz dengan kapasiti RAM 8 GB.

PANGKALAN DATA IMEJ

Radiografi x-ray ialah pengimejan diagnostik yang paling sering digunakan untuk pemeriksaan penyakit. Malah ia masih digunakan secara meluas kerana kos yang rendah, penggunaan yang mudah dan radiasi yang rendah, dianggarkan 100 kali lebih rendah daripada CT (Rasheed et al. 2007). Kajian awal telah dibangunkan menggunakan 60 imej x-ray yang terdiri daripada lima bahagian anggota badan yang berlainan iaitu dada, abdomen, pinggul, lutut dan pergelangan kaki, seperti yang ditunjukkan dalam Rajah 2. Imej-imej ini diperolehi dari pangkalan data imej perubatan di Jabatan Radiologi, Pusat Perubatan Universiti Kebangsaan Malaysia (PPUKM) hasil kerjasama penyelidikan yang dijalankan bersama beberapa pakar radiologi di sana. Imej-imej ini adalah dalam format DICOM dan diperolehi daripada 60 pesakit. Kajian yang dijalankan ini menggunakan perisian MATLAB dan Java untuk membangunkan sistem DSI bagi aplikasi imej perubatan.



RAJAH 1. Rajah aliran kerja sistem DSI yang dicadangkan



RAJAH 2. Sampel imej x-ray yang digunakan (a) dada, (b) abdomen, (c) pinggul, (d) lutut dan (e) gelang kaki

CADANGAN TBIRS

Terdapat tiga modul penting dalam TBIRS iaitu penyarian metadata, sistem pengurusan pangkalan data dan proses DSI. Modul penyarian metadata memproses dan mengekstrak maklumat daripada pengepala imej DICOM untuk mendapatkan deskripsi teks yang digunakan dalam proses DSI. Modul pengurusan pangkalan data bertanggungjawab untuk menyimpan dan menyusun deskripsi teks dalam bentuk jadual bagi memudahkan proses DSI. Semasa proses DSI, kata kunci yang digunakan oleh pengguna untuk menanya dan mendapat semula imej serupa dibandingkan dengan metadata imej dalam pangkalan data metadata TBIRS. Selepas itu, hasil DSI menggunakan TBIRS akan dipaparkan kepada pengguna.

Dalam bidang perubatan, imej-imej kebiasaannya disimpan dalam format DICOM. Imej DICOM mengandungi pelbagai maklumat penting mengenai pesakit seperti identiti pesakit, jantina, modaliti imej, bahagian badan dan parameter mesin. Dalam kajian ini, beberapa kategori atau medan penting yang biasanya dirujuk dan digunakan oleh ahli radiologi dan doktor telah diekstrak dan disimpan dalam pangkalan data TBIRS. Kategori-kategori yang dipilih adalah nama fail, modaliti, jantina, bahagian badan dan posisi pandangan seperti yang ditunjukkan dalam Jadual 2.

Microsoft Access (MA) telah digunakan sebagai pangkalan data untuk menyimpan dan menyusun metadata yang diekstrak. MA dipilih kerana ia merupakan sistem pengurusan pangkalan data yang boleh berinteraksi dengan GUI. MA menyimpan metadata dalam bentuk jadual mengikut kategori masing-masing supaya senang dibaca dan diproses

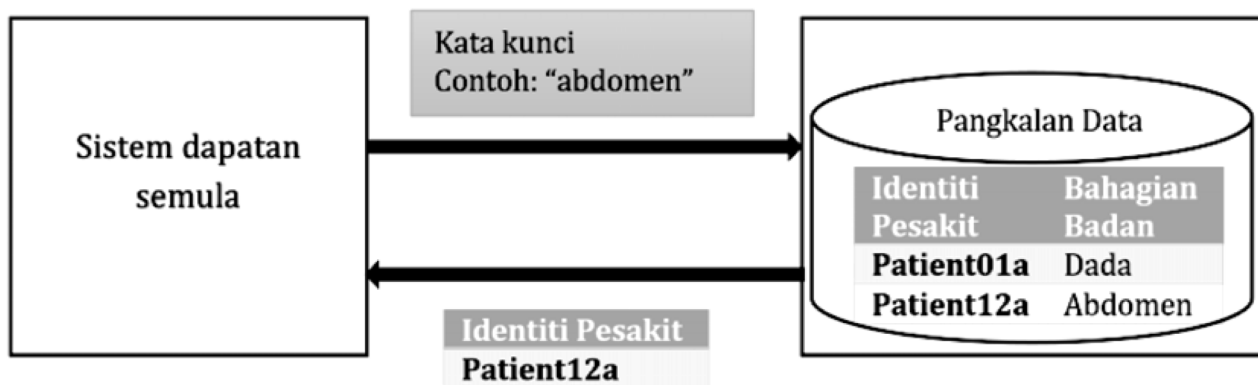
JADUAL 2. Kategori-kategori dalam pangkalan data bagi TBIRS

Kategori	Deskripsi	Contoh Format
Nama fail	Nama fail imej	patient01a.dcm
Bahagian badan	Bahagian anatomi badan	abdomen, dada
Modaliti	Jenis pengimbas yang digunakan	x-ray, MRI
Jantina	Lelaki atau perempuan	M atau F
Posisi pandangan	Posisi imej diambil	AP, PA

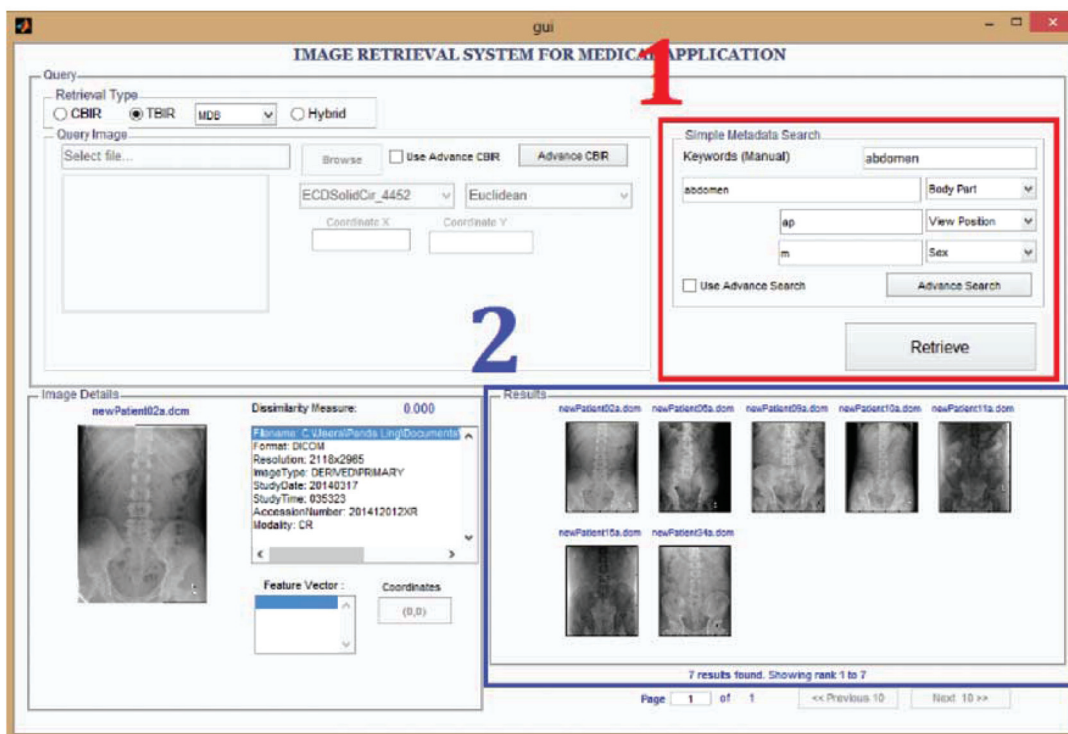
oleh sistem. Rajah 3 menunjukkan interaksi antara pangkalan data dan sistem DSI.

Bagi mendapatkan semula imej dalam TBIRS, kata kunci telah digunakan sebagai data masukan kepada sistem. Sistem ini berinteraksi dengan pangkalan data untuk mencari metadata yang sepadan dengan kata kunci lalu memaparkan imej-imej berkenaan kepada pengguna. Metadata dalam jadual pangkalan data dibaca dari baris pertama hingga ke baris terakhir. Oleh itu, imej-imej yang didapatkan semula kemudiannya disusun dari imej teratas hingga ke bawah. Satu GUI ringkas telah dibangunkan untuk menguji kebolehan dan prestasi TBIRS seperti yang ditunjukkan dalam Rajah 4. Dalam rajah tersebut, medan "1" merupakan ruangan bagi kata kunci dari kategori yang boleh dipilih manakala medan "2" adalah imej-imej dapatan semula berdasarkan kata kunci dan kategori yang dipilih.

Rajah 5 (a) menunjukkan terdapat beberapa bahagian kata kunci yang perlu diisi oleh pengguna. Pengguna boleh mengisi lebih daripada satu kata kunci. Seperti yang ditunjukkan dalam rajah tersebut, pengguna boleh memilih bahagian badan, kedudukan pandangan dan juga jantina untuk imej yang ingin dicari. Imej-imej yang didapat semula akan dipaparkan di bahagian keputusan seperti yang ditunjukkan oleh Rajah 5 (b). Bahagian ini akan memaparkan semua imej yang berkaitan dengan kata kunci yang telah dimasukkan. Imej-imej ini yang terdiri daripada kategori yang sama dengan imej yang ditanya oleh pengguna boleh digunakan sebagai rujukan oleh doktor atau ahli radiologi. Prestasi TBIRS telah diuji dengan menggunakan setiap satu daripada 60 imej-imej x-ray sebagai imej pertanyaan pada sesuatu masa dengan memasukkan kata kunci secara manual seperti yang ditunjukkan dalam Rajah 5. Pemerhatian telah mendapati sistem ini mampu mendapat balik 100% imej-imej dalam kategori yang sama dengan imej yang ditanya oleh pengguna. Walau bagaimanapun, sistem ini memerlukan pengguna memasukkan kata kunci yang betul bagi mendapat balik imej-imej dari kategori yang sama dari pangkalan data. Kekangan ini mendorong pembangunan sistem DSI menggunakan kandungan dan juga mengintegrasikannya dengan penggunaan kata kunci yang dikenali sebagai sistem dapatan balik secara hibrid (HBIRS).



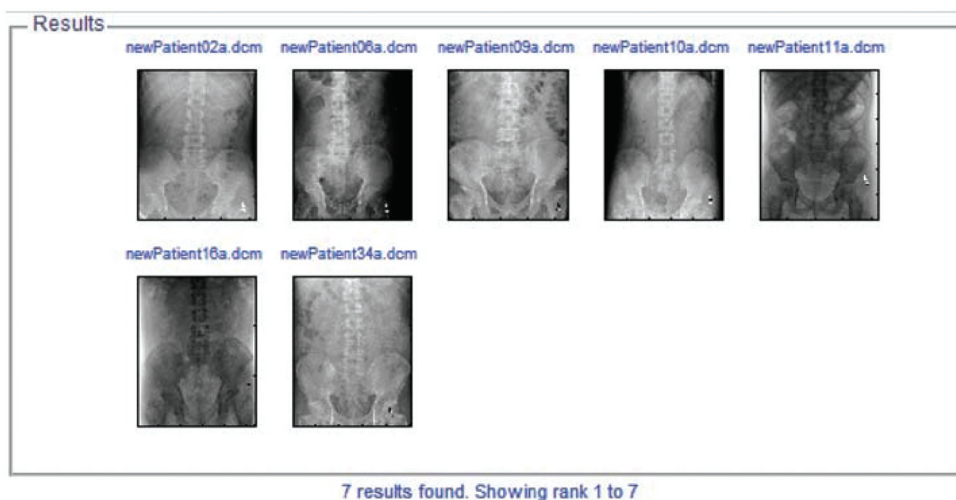
RAJAH 3. Interaksi antara TBIRS dan pangkalan data



RAJAH 4. Prototaip yang menunjukkan imej abdomen yang ditanya pengguna

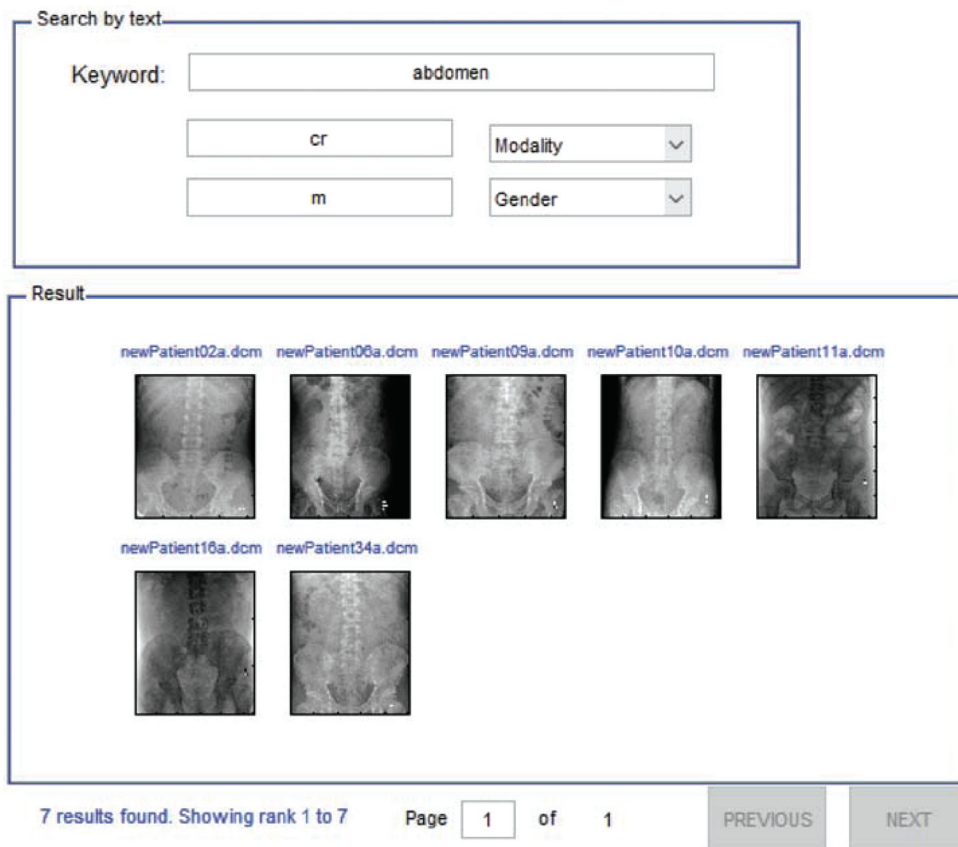


(a)



(b)

RAJAH 5. (a) Bahagian TBIRS untuk memasukkan kata kunci (b) Bahagian sistem DSI yang memaparkan imej yang didapat semula



RAJAH 6. Keputusan TBIRS secara visual dengan dapatan semula pada imej

TBIRS dilaksanakan secara langsung melalui pepadanan kata kunci dengan metadata dalam pangkalan data sahaja dan tidak melibatkan sebarang pemprosesan imej. Oleh itu, penilaian TBIRS dilakukan dengan cara membandingkan jumlah imej daripada dapatan semula dengan jumlah imej dalam pangkalan data. Prestasi sistem terbaik akan mempunyai jumlah imej yang berkaitan daripada dapatan semula sama dengan jumlah imej yang berkaitan dalam pangkalan data seperti pada persamaan (1).

$$\frac{\text{Jumlah imej yang berkaitan daripada dapatan semula}}{\text{Jumlah imej yang berkaitan dalam pangkalan data}} = 1 \quad (1)$$

Jadual 3 menunjukkan keputusan TBIRS bagi kombinasi kata kunci yang berlainan di mana jumlah imej daripada dapatan semula adalah sama dengan jumlah imej dalam pangkalan data. Ia mencatatkan kejituan sempurna sebanyak 100% bagi TBIRS dengan pengujian dapatan semula ini dibuat menggunakan metadata dari medan yang umum iaitu modaliti, jantina, bahagian badan dan posisi pandangan sahaja. Namun begitu, pengujian dapatan semula berdasarkan teks tidak dapat dijalankan menggunakan medan 'study' yang memerlukan ahli radiologi atau doktor memasukkan huraian atau catatan. Ini adalah kerana, tidak semua medan pada imej MRI tersebut mempunyai catatan. Kekangan ini juga menyebabkan DSI tidak berfungsi dengan baik terutamanya apabila berlaku ketidakseragaman catatan pada metadata.

Keputusan TBIRS secara kualitatif juga ditunjukkan pada Rajah 6 dengan kata kunci "abdomen", "cr" dan "m". Keputusan menunjukkan kesemua 7 imej (seperti yang dicatat dalam Jadual 3) telah dipaparkan pada paparan GUI.

JADUAL 3. Keputusan TBIRS bagi kombinasi kata kunci yang berlainan

Kata kunci	Jumlah imej yang berkaitan dalam pangkalan data	Jumlah imej yang berkaitan daripada dapatan semula
Bahagian badan = 'abdomen'	11	11
Bahagian badan = 'abdomen'; Jantina = 'M'	7	7
Bahagian badan = 'abdomen'; Jantina = 'M'; Pandangan = 'AP'	4	4

SISTEM CBIR

Seperti yang ditunjukkan dalam Rajah 1, bahagian CBIRS melibatkan tiga fasa utama iaitu pra-pemprosesan, penyarian dan pemilihan fitur dan dapatan semula. Fasa pra-pemprosesan merangkumi dua modul utama iaitu penambahbaikan imej dan peruasan imej. Fasa ini adalah fasa yang pertama

dalam pembangunan CBIRS bertujuan untuk mengurangkan hingar, menambahbaik pinggir, mengesan tepi, menyelaraskan kontras (Negi and Bhandari, 2014) dan pengekstrakan ROI. Pra-pemprosesan dapat meningkatkan ketepatan dalam diagnosis selain mengurangkan masa dan kos tenaga manusia (Rajamani, S.Jaiganesh, V., 2013). Pangkalan data CBIRS menjalankan proses pemprosesan, peruasan dan penyarian fitur sebelum proses pepadanan. Satu pangkalan data untuk fitur-fitur imej yang telah disarikan akan dibentuk dari proses-proses di atas. Fitur-fitur yang disari adalah seperti warna, bentuk dan tekstur dari imej. Fitur-fitur tersebut akan digunakan untuk mengenal pasti imej yang ingin dicari. Dengan cara yang sama, imej yang dimasukkan oleh pengguna juga akan mengalami proses yang sama sebelum proses pepadanan. Akhirnya, perbandingan antara fitur-fitur imej yang dimasukkan dengan pangkalan data fitur dan imej yang sesuai akan dipaparkan kepada pengguna.

SISTEM HBIR

Dalam sistem ini, kedua-dua fungsi dalam TBIR dan CBIR akan digabungkan, seperti yang digambarkan di dalam Rajah 1. Sistem HBIR membolehkan pengguna memasukkan kata kunci dan imej untuk pemprosesan bagi mendapatkan semula imej yang sama dengan fitur dan kata kunci dari imej pertanyaan. Sistem ini dijangka akan memberikan kejituan yang lebih tinggi. Kejituan dan perolehan imej boleh digunakan untuk mengukur keberkesanan dan ketepatan sistem. Kawasan bawah lengkungan graf kejituan-perolehan juga boleh digunakan untuk mengukur prestasi sistem. Semakin tinggi nilai kawasan bawah lengkungan graf adalah semakin jitu sistem.

KESIMPULAN

Sistem TBIR yang telah dipaparkan didalam kertas kerja ini menggunakan kata kunci yang berkaitan dengan imej yang dicari yang akan dimasukkan oleh pengguna melalui GUI. Sistem akan berinteraksi dengan pangkalan data metadata yang berkenaan dan melakukan proses pepadanan kata kunci dengan metadata dalam pangkalan data tersebut. Akhirnya semua imej yang berkenaan akan dipaparkan kepada pengguna. Adalah diakui terdapat banyak cabaran yang perlu ditangani dalam membangunkan sistem DSI yang lebih efisien yang boleh digunakan oleh doktor dalam bidang perubatan. Oleh itu, sistem DSI berdasarkan kandungan (CBIRS) dalam bidang perubatan akan tetap dijalankan dalam kajian seterusnya. Selain itu, keberkesanan HBIRS juga akan dikaji, serta pengintegrasian sistem tersebut dalam platform android akan dibuat supaya boleh digunakan sebagai alat bantuan mengajar bagi pendidik di dalam bidang perubatan.

PENGHARGAAN

Penyelidikan ini disokong oleh Geran Galakan Penyelidik Muda (GGPM-2013-012) dan Dana Impak Perdana (DIP-2015-012) dari Universiti Kebangsaan Malaysia.

RUJUKAN

- Ahad, N.A., T.S. Yin, A.R. Othman. & C.R. Yaacob. 2011. Sensitivity of normality tests to non-normal data. *Sains Malaysiana* 40(6): 637-41.
- Ahmad, I.S. 2008. Text-based image indexing and retrieval using formal concept analysis. *KSII Transactions on Internet and Information Systems* 2(3):150-70.
- Amaral, I.F., F. Coelho, J.F.P. Da Costa. & J.S. Cardoso. 2010. Hierarchical medical image annotation using SVM-based approaches. *In Proceedings of the IEEE/EMBS Region 8 International Conference on Information Technology Applications in Biomedicine*, ITAB.
- Arakeri, M.P. & G. Ram Mohana Reddy. 2013. An intelligent content-based image retrieval system for clinical decision support in brain tumor diagnosis. *International Journal of Multimedia Information Retrieval* 2 (3):175-88.
- Arakeri, M.P. & G.R.M. Reddy. 2012. Medical image retrieval system for diagnosis of brain tumor based on classification and content similarity. *In 2012 Annual IEEE India Conference (INDICON)* 416-21.
- Arthi, K. & J. Vijayaraghavan. 2013. Content based image retrieval algorithm using colour models. *International Journal of Advanced Research in Computer and Communication Engineering* 2(3): 1343-47.
- Bergamasco, L.C.C. & F.L.S. Nunes. 2015. Three-dimensional content-based cardiac image retrieval using global and local descriptors. *In AMIA Annual Symposium Proceedings American Medical Informatics Association* 1811-1820.
- Bhadoria, S. & C.G. Deth. 2010. Study of medical image retrieval. *In 2010 International Conference on Data Storage and Data Engineering* 192-96.
- Boparai, N.K. & A. Chhabra. 2015. A hybrid approach for improving content based image retrieval systems. *In 1st International Conference on Next Generation Computing Technologies (NGCT-2015)* 944-49.
- Chuctaya, H., C. Portugal, C. Beltran, J. Gutierrez, C. Lopez, & Y. Tupac. 2012. M-CBIR: A medical content-based image retrieval system using metric data-structures. *In Proceedings – International Conference of the Chilean Computer Science Society, SCCC* 135-41.
- Feng, D., W. Siu. & H.J. Zhang. 2003. *Multimedia Information Retrieval and Management: Technological Fundamentals and Applications*. Springer
- Ghosh, P., S. Antani, L.R. Long. & G.R. Springer. 2011. Review of medical image retrieval systems and future directions. *In 2011 24th International Symposium on Computer-Based Medical Systems (CBMS)* 1-6.

- Goodrum, A. 2000. Image information retrieval: An overview of current research. *Informing Science* 3(2): 63-66.
- Grace, R.K., R. Manimegalai, & S.S. Kumar. 2014. Medical image retrieval system in grid using hadoop framework. *In Proceedings – 2014 International Conference on Computational Science and Computational Intelligence, CSCI 2014* 1: 144-148.
- Guo, J.M., H. Prasetyo, H. Lee, & C.C. Yao. 2016. Image retrieval using indexed histogram of void-and-cluster block truncation coding. *Signal Processing* 123: 143-156.
- Hartvedt, C. 2010. Using context to understand user Intentions in image retrieval. *In 2nd International Conference on Advances in Multimedia, MMEDIA 2010* 130-133.
- Herrera, A.G.S. De, D. Markonis, R. Schaer, I. Eggel, & H. Müller. 2013. The medGIFT Group in Image CLEF med 2013. *In CEUR Workshop Proceedings*. Vol. 1179.
- Ho, J.-M., S.-Y. Lin, C.-W. Fann, Y.-C. Wang, & R.-I. Chang. 2012. A novel content based image retrieval system using k-means with feature extraction. 2012 *International Conference on Systems and Informatics (ICSAI2012)*, no. Icsai: 785-790.
- Husam, I.S., A.A.B. Abuhamad, S. Zainudin, M. Sahani, & Z.M. Ali. 2017. Feature selection algorithms for Malaysian dengue outbreak detection model. *Sains Malaysiana* 46(2): 255-265.
- Jaganathan, Y. & I. Vennila. 2013. A hybrid approach based medical image retrieval system using feature optimized classification similarity framework. *American Journal of Applied Sciences* 10(6): 549-562.
- Kalpathy-Cramer, J., A.G.S. de Herrera, D. Demner-Fushman, S. Antani, S. Bedrick, & H. Müller. 2015. Evaluating performance of biomedical image retrieval systems-an overview of the medical image retrieval task at ImageCLEF 2004-2013. *Computerized Medical Imaging and Graphics* 39: 55-61.
- Kaur, H. & K. Jyoti. 2013. Survey of techniques of high level semantic based image retrieval. *International Journal of Research in Computer and Communication Technology* 2(1): 15-19.
- Khan, S.M.H., A. Hussain, & I.F.T. Alshaikhli. 2012. Comparative study on content-based image retrieval (CBIR). *In 2012 International Conference on Advanced Computer Science Applications and Technologies (ACSAT)* 61-66.
- Kitanovski, I., G. Strezoski, I. Dimitrovski, G. Madjarov, & S. Loskovska. 2016. Multimodal medical image retrieval system. *Multimedia Tools and Applications* 1-24.
- Kumar, A., J. Kim, W. Cai, M. Fulham, & D. Feng. 2013. Content-based medical image retrieval: A survey of applications to multidimensional and multimodality data. *Journal of Digital Imaging* 26(6): 1025-1039.
- Kumar, A., J. Kim, J., Wen, L., Fulham, M. & Feng, D. 2014. A graph-based approach for the retrieval of multimodality medical images. *Medical Image Analysis* 18(2): 330-342.
- Madugunki, M., Bormane, D.S., Bhadoria, S. & Dethé, C.G. 2011. Comparison of different CBIR techniques. *In ICECT 2011 - 2011 3rd International Conference on Electronics Computer Technology* 4: 372-375.
- MedGift. 2017. Retrieval of Medical Images by Their Visual Content.
- Müller, H., & H. Greenspan. 2013. Overview of the third workshop on medical content-based retrieval for clinical decision support (MCBR-CDS 2012). *In Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7723 LNCS: 1-9.
- Müller, H., J. Kalpathy-cramer, B. Caputo, T. Syeda-Mahmood, & F. Wang. 2010. Overview of the first workshop on medical content-based retrieval for clinical decision support at MICCAI 2009. *In Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5853 LNCS: 1-17.
- Murala, S. & Q.M. Jonathan Wu. 2013. Local ternary co-occurrence patterns: A new feature descriptor for MRI and CT image retrieval. *Neurocomputing* 119: 399-412.
- Murthy, V.S.V.S., E. Vamsidhar, J.N.V.R.S. Kumar, & P. Sankara Rao. 2010. Content based image retrieval using hierarchical and k-means clustering techniques. *International Journal of Engineering Science and Technology* 2(3): 209-12.
- Nakaram, P. & T. Leauhatong. 2012. A New Content-Based Medical Image Retrieval System Based on Wavelet Transform and Multidimensional Wald-Wolfowitz Runs Test. *In 5th 2012 Biomedical Engineering International Conference, BMEiCON 2012*.
- Negi, S.S. & Y.S. Bhandari. 2014. A hybrid approach to image enhancement using contrast stretching on image sharpening and the analysis of various cases arising using histogram. *In International Conference on Recent Advances and Innovations in Engineering, ICRAIE 2014*.
- Nowaková, J., M. Prilepok, & V. Snášel. 2017. Medical image retrieval using vector quantization and fuzzy s-tree. *Journal of Medical Systems* 41(2):1-16.
- Pilevar, A.H. 2011. CBMIR: Content-based image retrieval algorithm for medical image databases. *Journal of Medical Signals and Sensors* 1(1): 12-18.
- Pourghassem, H. & H. Ghassemian. 2008. Content-based medical image classification using a new hierarchical merging scheme. *Computerized Medical Imaging and Graphics* 32(8): 651-661.
- Rajamani, S.Jaiganesh, P.B. V. 2013. A review of various global contrast enhancement techniques for still images using histogram modification framework. *International Journal of Engineering Trends and Technology* 4(4): 1045-1048.
- Ramamurthy, B. & K.R. Chandran. 2011. Cbmir: Shape-based image retrieval using canny edge detection and k-means clustering algorithms for medical images.

- International Journal of Engineering Science and Technology* 3(3): 1870-1877.
- Rasheed, T., B. Ahmed, M.A.U. Khan, M. Bettayeb, L. Sungyoung. & K. Tae-Seong. 2007. Rib suppression in frontal chest radiographs: a blind source separation approach. In *Signal Processing and Its Applications, 2007. ISSPA 2007. 9th International Symposium on*, 1-4.
- Silva, S.F. Da, M.X. Ribeiro, J.D.E.S. Batista Neto, C. Traina-Jr. & A.J.M. Traina. 2011. Improving the ranking quality of medical image retrieval using a genetic feature selection method. *Decision Support Systems* 51(4): 810-820.
- Spanier, A.B., D. Cohen. & L. Joskowicz. 2017. A new method for the automatic retrieval of medical cases based on the radlex ontology. *International Journal of Computer Assisted Radiology and Surgery* 12(3): 471-484.
- Sparks, R. & A. Madabhushi. 2016. Out-of-Sample extrapolation utilizing semi-supervised manifold learning (OSE-SSL): content based image retrieval for histopathology images. *Scientific Reports* 6 (June). England: 27306.
- Suganya, R. & S. Rajaram. 2012. Content based image retrieval of ultrasound liver diseases based on hybrid approach. *American Journal of Applied Sciences* 9(6): 938-945.
- Syam, B., J.S.R. Victor. & Y.S. Rao. 2013. Efficient similarity measure via genetic algorithm for content based medical image retrieval with extensive features. In *Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), 2013 International Multi-Conference on* 704-11.
- Wan Ahmad, W.S.H.M., Zaki, W.M.D.W., Fauzi, M.F.A. & Haw, T.W. 2014. content-based medical image retrieval system for infections and fluids in chest radiographs. In *Information Retrieval Technology*, edited by Jaafar, A., Mohamad Ali, N., Mohd Noah, S.A., Smeaton, A.F., Bruza, P., Bakar, Z.A., Jamil, N., and Sembok, T.M.T., 14–23. Cham: Springer International Publishing.
- Wang, J., Zhang, D. & Wang, Q. 2016. An image retrieval system with color emotion query. In *Proceedings – International Conference on Natural Computation* 446-550.
- Xu, Y., Lin, L., Hu, H., Wang, D., Zhu, W., Wang, J., Han, X.H. and Chen, Y.W. 2018. Texture-specific bag of visual words model and spatial cone matching-based method for the retrieval of focal liver lesions using multiphase contrast-enhanced CT images. *International Journal of Computer Assisted Radiology and Surgery* 2017.
- Yaghoubyan, S.H., M.A. Maarof, A. Zainal. & M.M. Oghaz. 2016. A survey of feature extraction techniques in content-based illicit image detection. *Journal of Theoretical and Applied Information Technology* 87(1): 110-25.
- Zhituo, X., Hao, R. & Hao, W. 2012. A content-based image retrieval system using multiple hierarchical temporal memory classifiers. In *Proceedings – 2012 5th International Symposium on Computational Intelligence and Design, ISCID 2012 2*: 438-41.

Wan Siti Halimatul Munirah Wan Ahmad, *Wan Mimi Diyana Wan Zaki, Aini Hussain, Ling Chei Siong
 Pusat Kejuruteraan Sistem Bersepadu dan Teknologi Termaju (INTEGRA),
 Fakulti Kejuruteraan dan Alam Bina,
 Universiti Kebangsaan Malaysia,
 43600 UKM Bangi, Selangor, Malaysia

Wong Erica Yee Hing
 Jabatan Radiologi,
 Pusat Perubatan Universiti Kebangsaan Malaysia,
 56000 Cheras, Malaysia.

*Corresponding author; email: wmdiyana@ukm.edu.my

Received date : 14th November 2017
 Accepted date : 1st February 2018
 In Press date : 1st April 2018
 Published date : 30th April 2018