

PEMBANGUNAN ONTOLOGI KANSER PAYUDARA BAGI PEMILIHAN
DATA DALAM MERAMAL RISIKO

FATIMATUFARIDAH BINTI JUSOH

UNIVERSITI TEKNOLOGI MALAYSIA

PEMBANGUNAN ONTOLOGI KANSER PAYUDARA BAGI PEMILIHAN
DATA DALAM MERAMAL RISIKO

FATIMATUFARIDAH BINTI JUSOH

Laporan projek ini dikemukakan sebagai memenuhi
syarat bagi penganugerahan ijazah
Sarjana Sains (Sains Komputer)

Fakulti Komputeran
Universiti Teknologi Malaysia

MAC 2014

*Buat ibubapa yang tersayang, Jusoh bin Mat Zain dan Latifah binti Razali,
Tiada kata-kata yang dapat ku ukirkan sebagai gambaran akan segala
pengorbanan yang telah dicurahkan. Yang sentiasa mempercayai tanpa ragu,
meminta tanpa memaksa, menerima tanpa rasa kekurangan dan memahami
tanpa alasan.*

*Tidak dilupakan kepada semua adik-beradikku dan rakan-rakan.
Terima kasih atas doa yang diberikan serta sokongan kalian semua.*

PENGHARGAAN

Syukur ke hadrat Ilahi kerana dengan izinNya, saya akhirnya dapat menyiapkan laporan projek ini dalam tempoh yang ditetapkan.

Sekalung penghargaan dan terima kasih buat kedua-dua penyelia projek saya, Dr Mohd Shahizan bin Othman dan Dr Roliana binti Ibrahim atas bimbingan, dorongan dan segala tunjuk ajar yang diberikan oleh beliau sepanjang tempoh laporan projek ini disiapkan.

Terima kasih kepada Kementerian Pelajaran Tinggi kerana membiayai pengajian saya di sini dan juga kepada pihak Universiti Teknologi Malaysia (UTM) kerana menyediakan persekitaran pembelajaran yang kondusif dan kemudahan seperti makmal yang lengkap untuk memudahkan proses menyiapkan laporan ini.

ABSTRACT

Breast cancer is a deadly disease caused by the uncontrolled growth of cells that starts in the breast. Therefore, the accurate risk prediction is crucial in assisting the selection for the suitable prevention treatment, depending on the level of the risk. However, the abundance of biomedical data from various sources creates difficulty in data organizing. In addition, the big challenge in predicting the risk of breast cancer is the different attributes of the datasets which make it inscrutable for someone who are not from the domain background. Ontology is a new method introduced to improve the knowledge discovery in complex database. Ontology approach was applied in this study to resolve this problem by providing clearer understanding of the data. In this study, ontology was also used to select important features for data analysis. Classification technique of Sequential Minimal Optimization (SMO) was also applied in this study. SMO is a fast learning algorithm of Support Vector Machine (SVM) and able to provide high accuracy results. However, the analysis of breast cancer risk shows that data analysis without ontology has slightly higher accuracy compared to data analysis with ontology, where, the first dataset is 94.7% compared to 92.1% and the accuracy for the second dataset is 96.7% compared to 96.6%. These results were different from expectation, which the application of ontology was supposed to be able to provide higher accuracy results. This is caused by the limitation of data available for this study. Therefore, the study on breast cancer risk prediction by using ontology can be improved in the future by using broader cancer data and consistent cancer data type.

ABSTRAK

Kanser payudara merupakan penyakit pembunuh yang berpunca daripada pertumbuhan sel yang tidak terkawal dibahagian dada. Oleh itu, ketepatan dalam meramal risiko adalah penting bagi memilih rawatan yang sesuai pada setiap tahap risiko. Namun begitu, lambakan data bioperubatan yang datang dari pelbagai sumber menyukarkan pengurusan data. Selain itu, cabaran besar dalam meramal risiko adalah data yang digunakan mempunyai pelbagai atribut yang sukar difahami bagi penganalisa data yang bukan dari domain tersebut. Ontologi merupakan satu kaedah baru bagi menambah baik penemuan pengetahuan dalam pangkalan data yang kompleks. Kajian ini menggunakan pendekatan ontologi bagi menyelesaikan masalah ini dengan memberikan pemahaman data yang lebih jelas. Selain itu, ontologi turut digunakan bagi memilih ciri yang penting bagi melakukan analisis data. Kaedah klasifikasi Pengoptimuman Jujukan Minimum (SMO) telah digunakan dalam kajian ini. Teknik SMO merupakan algoritma pembelajaran bagi Mesin Sokongan Vektor (SVM) yang pantas dan mampu memberikan hasil dengan ketepatan yang tinggi. Walau bagaimanapun, hasil analisis risiko kanser payudara mendapati bahawa ketepatan analisis data tanpa menggunakan ontologi adalah lebih tinggi sedikit berbanding analisis data dengan menggunakan ontologi, di mana, bagi set data pertama adalah 94.7% berbanding 92.1%. dan ketepatan bagi set data kedua adalah 96.7% berbanding 96.6%. Hasil yang didapati adalah berbeza daripada jangkaan, di mana ontologi sepatutnya memberikan ketepatan hasil yang lebih tinggi. Hal ini adalah disebabkan oleh kekangan data yang didapati di dalam kajian ini. Oleh itu, kajian ramalan risiko kanser payudara berasaskan ontologi boleh ditambah baik pada masa akan datang dengan menggunakan data yang lebih banyak dan jenis data kanser yang lebih konsisten.

ISI KANDUNGAN

BAB	TAJUK	HALAMAN
	PENGAKUAN	ii
	DEDIKASI	iii
	PENGHARGAAN	iv
	ABSTRACT	v
	ABSTRAK	vi
	ISI KANDUNGAN	vii
	SENARAI JADUAL	xii
	SENARAI RAJAH	xiv
	SENARAI SINGKATAN	xvi
	SENARAI ISTILAH	xviii
	SENARAI LAMPIRAN	xix
1	Pengenalan	
	1.1 Pengenalan	1
	1.2 Latar Belakang Masalah	2
	1.3 Pernyataan Masalah	5
	1.4 Objektif	6
	1.5 Skop	6
	1.6 Kepentingan Kajian	7
	1.7 Kesimpulan	8
2	KAJIAN LITERATUR	
	2.1 Pengenalan	9
	2.2 Kanser	9

2.3	Kanser Payudara	11
2.4	Tahap Kanser Payudara	13
2.5	Meramal Risiko Kanser	15
2.6	Kaedah Klasifikasi	16
2.6.1	Mesin Sokongan Vektor	19
2.6.2	Pengoptimuman Jujukan Minimum	21
2.7	Kumpulan Data Kanser dan Permasalahan Data	22
2.8	Integrasi Data	23
2.9	Ontologi	25
2.10	Pendekatan Ontologi Berasaskan Integrasi Data	30
2.11	Pemilihan Data Berasaskan Ontologi	35
2.12	Ontologi Dalam Bidang Perubatan	35
2.13	Justifikasi bagi Pemodelan Ontologi dan Meramal Risiko Kanser	37
2.14	Kesimpulan	40
3	METODOLOGI KAJIAN	
3.1	Pengenalan	41
3.2	Aliran kerja Penyelidikan	41
3.3	Fasa Penentuan Masalah dan Pengumpulan Data	43
3.4	Pembangunan Model Ontologi Kanser Payudara	45
3.4.1	Prapembangunan Ontologi Hibrid	48
3.4.2	Pembangunan Ontologi Hibrid	49
3.4.3	Selepas Pembangunan Ontologi Hibrid	51
3.4.4	Penilaian Ontologi	51
3.5	Pembangunan Model Ramalan Risiko Kanser Payudara	52
3.6	Penilaian	54
3.7	Kesimpulan	55

4	PEMODELAN ONTOLOGI KANSER PAYUDARA	
4.1	Pengenalan	56
4.2	Pelaksanaan Pemodelan Ontologi	56
4.2.1	Pelaksanaan Pembangunan Ontologi Hibrid Berasaskan Integrasi Data	57
4.2.2	Pemadanan Istilah Melalui Pemetaan Ontologi	86
4.2.3	Implementasi Ontologi Menggunakan Protégé	87
4.3	Algoritma Pembangunan Ontologi Berasaskan Integrasi Data	92
4.4	Penilaian Ontologi	93
4.4.1	Pakar Domain dan <i>Reasoner</i>	93
4.4.2	Metrik Ontologi	94
4.4.3	Aplikasi Ramalan Risiko Kanser Payudara Berasaskan Ontologi	95
4.5	Kesimpulan	99
5	PEMBANGUNAN MODEL RAMALAN RISIKO KANSER PAYUDARA	
5.1	Pengenalan	101
5.2	Pemilihan Atribut Bagi Meramal Risiko Kanser Payudara	102
5.3	Pelaksanaan Analisis Risiko Kanser Payudara	104
5.3.1	Analisis Set Data 1 Tanpa Ontologi	109
5.3.2	Analisis Set Data 2 Tanpa Ontologi	110
5.3.3	Analisis Set Data 1 Dengan Ontologi	112
5.3.4	Analisis Set Data 2 Dengan Ontologi	113
5.4	Analisis Prestasi	115
5.5	Perbincangan Analisis Risiko Kanser Payudara	117
5.6	Kesimpulan	119

6	KESIMPULAN KAJIAN	
6.1	Pengenalan	120
6.2	Rumusan Perbincangan	121
6.3	Kekangan Kajian	122
6.4	Sumbangan Kajian	122
6.5	Kesimpulan	123
	RUJUKAN	124
	LAMPIRAN	134

SENARAI JADUAL

BIL	TAJUK	HALAMAN
1	Kajian lepas dalam ramalan risiko kanser	17
2	Kajian lepas dalam integrasi data	24
3	Kajian lepas tentang ontologi berasaskan integrasi	28
4	Langkah pembangunan ontologi	32
5	Senarai istilah dalam set data <i>Wisconsin Breast Cancer</i>	44
6	Senarai istilah dalam laman sesawang	44
7	Matriks ramalan	54
8	Pengiraan bagi analisis prestasi	55
9	Data dari <i>PubMed Health</i>	60
10	Pemetaan ontologi global kepada ontologi lokal bagi data <i>Wikipedia Breast Cancer</i>	70
11	Pemetaan ontologi global kepada ontologi lokal bagi data <i>PubMed Health</i>	72
12	Pemetaan ontologi global kepada ontologi lokal bagi data <i>MedicineNet</i>	73
13	Pemetaan ontologi global kepada ontologi lokal bagi data <i>Breastcancer.org</i>	74
14	Pemetaan ontologi global kepada ontologi lokal bagi data <i>Resource for Health Professionals, California</i>	75
15	Pemetaan ontologi global kepada ontologi lokal bagi set data <i>Wisconsin Breast Cancer</i>	77
16	Pemetaan ontologi global kepada ontologi lokal bagi set data <i>Wisconsin Breast Cancer</i>	78
17	Pemetaan ontologi lokal kepada sumber data bagi data <i>Wikipedia</i>	79
18	Pemetaan ontologi lokal kepada sumber data bagi data <i>PubMed Health</i>	80

19	Pemetaan ontologi lokal kepada sumber data bagi data <i>MedicineNet</i>	81
20	Pemetaan ontologi lokal kepada sumber data bagi data <i>Breastcancer.org</i>	82
21	Pemetaan ontologi lokal kepada sumber data bagi data <i>Resource for Health Professionals, California</i>	84
22	Pemetaan ontologi lokal kepada sumber data bagi set data <i>Wisconsin Diagnostic Breast Cancer</i>	85
23	Pemetaan ontologi lokal kepada sumber data bagi set data <i>Wisconsin Breast Cancer</i>	86
24	Senarai atribut sel kanser bagi analisis data	103
25	Jumlah data yang digunakan dalam setiap larian	106
26	Ketepatan bagi data pengujian set data 1 tanpa ontologi	109
	Ringkasan SMO set data 1 tanpa ontologi	1010
	Ketepatan berdasarkan kelas set data 1 tanpa ontologi	1010
27	Matrik set data 1 tanpa ontologi	1010
28	Ketepatan bagi data pengujian set data 2 tanpa ontologi	111
29	Ringkasan SMO set data 2 tanpa ontologi	111
30	Ketepatan berdasarkan kelas set data 2 tanpa ontologi	111
31	Matrik set data 2 tanpa ontologi	112
32	Ketepatan bagi data pengujian set data 1 dengan ontologi	112
33	Ringkasan SMO set data 1 dengan ontologi	113
34	Ketepatan berdasarkan kelas set data 1 dengan ontologi	113
35	Matrik set data 1 dengan ontologi	113
36	Ketepatan bagi data pengujian set data 2 dengan ontologi	114
37	Ringkasan SMO set data 2 dengan ontologi	114
38	Ketepatan berdasarkan kelas set data 2 dengan ontologi	115
39	Matrik set data 2 dengan ontologi	115
40	Perbandingan hasil analisis risiko kanser	117

SENARAI RAJAH

BIL	TAJUK	HALAMAN
1	Kadar kematian pada tahun 2008	10
2	Peratusan kanser di Semenanjung Malaysia pada tahun 2006	11
3	Tahap serangan kanser payudara	13
4	Mesin Sokongan Vektor Linear	20
5	Fungsi objektif yang optimum akan dicari menggunakan SMO	21
6	Struktur Ontologi	26
7	Penggunaan ontologi dalam PPPD	27
8	Pendekatan ontologi sedia ada	30
9	Aliran kerja penyelidikan	42
10	Senibina model ramalan risiko kanser payudara berasaskan ontologi	47
11	Langkah bagi meramal risiko dengan menggunakan SMO	53
12	Langkah pembangunan ontologi hibrid	57
13	Perwakilan konsep, atribut dan hubungan	58
14	Contoh konsep dan atribut ontologi global	59
15	Set data <i>Wisconsin Breast Cancer</i>	59
16	Ontologi lokal set data <i>Wisconsin Breast Cancer</i>	60
17	Ontologi lokal bagi <i>PubMed Health</i>	61
18	Ontologi global bagi kanser payudara	62
19	Ontologi lokal bagi data <i>Wikipedia Breast Cancer</i>	63
20	Ontologi lokal bagi data <i>pubmed health</i>	64
21	Ontologi lokal bagi data <i>medicineNet</i>	65

22	Ontologi lokal bagi data <i>BreastCancer.org</i>	66
23	Ontologi lokal bagi data <i>Resource for Health Professionals, California</i>	67
24	Ontologi lokal bagi set data <i>Wisconsin Diagnostic Breast Cancer</i>	68
25	Ontologi lokal bagi set data <i>Wisconsin Breast Cancer</i>	69
26	Pemadanan istilah dalam ontologi	87
27	Sebahagian ontologi kanser payudara yang dibangunkan dengan format RDFS	88
28	Panel navigasi	89
29	<i>Object Property Hierarchy</i>	90
30	Visual bagi istilah kanser payudara menggunakan ontoGraf	91
31	Fungsi carian bagi ontoGraf	91
32	<i>Inferred hierarchy</i>	94
33	Metrik Ontologi	95
34	<i>Wisconsin Diagnostic Breast Cancer</i>	96
35	Sifat fizikal saiz <i>Wisconsin Diagnostic Breast Cancer</i>	97
36	Sifat fizikal bentuk <i>Wisconsin Diagnostic Breast Cancer</i>	97
37	<i>Wisconsin Breast Cancer</i>	98
38	Sifat fizikal <i>Wisconsin Diagnostic Breast Cancer</i>	98
39	Ciri sel <i>Wisconsin Diagnostic Breast Cancer</i>	99
40	Sifat fizikal data kanser	102
41	Ciri-ciri sel kanser	103
42	Data sebelum penormalan	105
43	Data selepas penormalan	105

SENARAI SINGKATAN

SVM	-	<i>Support Vector Machine</i>
SMO	-	<i>Sequential Minimal Optimization</i>
WHO	-	<i>World Health Organization</i>
NCR	-	<i>National Cancer Registry</i>
WBC	-	<i>Wisconsin Breast Cancer</i>
WDBC	-	<i>Wisconsin Diagnostic Breast Cancer</i>
Medht	-	<i>Mediated Integration architecture</i>
INDUS	-	<i>Intelligent Data Understanding System</i>
BOVIS	-	<i>BOVine Information System</i>
PPPD	-	Penemuan Pengetahuan Dalam Pangkalan Data
MIOD	-	<i>Methodology of Integration-oriented Ontology Development</i>
PCO	-	<i>Prostate Cancer Ontology</i>
PCIS	-	<i>Prostate Cancer Information System</i>
QP	-	<i>Quadratic Programming</i>
DL	-	<i>Description Logic</i>
W3C	-	<i>World Wide Web Consortium</i>
OWL	-	<i>Web Ontology Language</i>
NCBO	-	<i>National Center for Biomedical Ontology,</i>
FNA	-	<i>Fine-Needle Aspiration</i>
DCIS	-	<i>Ductal Carcinoma In Situ</i>
LCIS	-	<i>Lobular Carcinoma In Situ</i>
IDC	-	<i>Invasive Ductal Carcinoma</i>
ILC	-	<i>Invasive Lobular Carcinoma</i>
SNB	-	<i>Sentinel Node Biopsy</i>

ALND	- <i>Axillary Clearance / Axillary Lymph Node Dissection</i>
IHC	- <i>Immunohistochemistry</i>
HRT	- <i>Hormone Replacement Therapy</i>
DES	- <i>Diethylstilbestrol Exposure</i>
MRI	- <i>Magnetic Resonance Imaging</i>
CBE	- <i>Clinical Breast Exam</i>
BSE	- <i>Breast Self-Examination</i>
TP	- <i>True Positive</i>
FP	- <i>False Positive</i>
FN	- <i>False Negative</i>
TN	- <i>True Negative</i>
RDFS	- <i>Resource Description Framework Scheme</i>
HRT	- <i>Hormone Replacement Therapy</i>

SENARAI ISTILAH

Ketepatan	- <i>Accuracy</i>
Kepekaan	- <i>Sensitivity / Recall</i>
Kespesifikan	- <i>Specificity</i>
Kepersisan	- <i>Precision</i>
Hibrid	- <i>Hybrid</i>
Tunggal	- <i>Single</i>
Berbilang	- <i>Multiple</i>
Positif betul	- <i>True positive</i>
Negatif salah	- <i>False negative</i>
Positif salah	- <i>False positive</i>
Negatif betul	- <i>True negative</i>
Persekutuan data	- <i>Data Federation</i>
Malignan	- <i>Malignant</i>
Benigna	- <i>Benign</i>
Pelanggan	- <i>Client</i>
Pertanyaan	- <i>Query</i>
Pengantaraan	- <i>Mediation</i>
Pembalutan	- <i>Wrapping</i>
Kejadian	- <i>Instance</i>
Rangkaian neural	- <i>Neural network</i>
Regresi linear	- <i>Linear regression</i>
Pengaturcaraan kuadratik	- <i>Quadratic programming</i>
Mesin Sokongan vektor	- <i>Support Vector Machine</i>
Pengoptimuman Jujukan Minimum	- <i>Sequential Minimal Optimization</i>

SENARAI LAMPIRAN

LAMPIRAN	TAJUK
A	Carta Gantt
B	Dataset 1 (WBC)
C	Dataset 2 (WDBC)
D	Format RDFS
E	Pembangunan Menggunakan Perisian Protégé 4.2
F	Kaji Selidik Pakar Domain

BAB 1

PENGENALAN

1.1 Pengenalan

Kanser merupakan penyakit yang berpunca dari pertumbuhan sel yang tidak terkawal di dalam badan manusia. Umumnya, punca utama kematian di dunia adalah disebabkan oleh kanser. Malah, kanser payudara merupakan penyebab utama kematian dikalangan wanita (*Cancer Research UK*, 2013). Kebiasaannya, kanser payudara terdapat dibahagian saluran susu dan kelenjar yang menghasilkan susu. Peningkatan jumlah kes kanser payudara adalah sangat membimbangkan. Justeru itu, analisis risiko kanser payudara perlu dilakukan. Meramal risiko kanser dapat membantu pakar perubatan membuat keputusan tentang rawatan bagi setiap pesakit dan meningkatkan keberkesanan perubatan dalam bidang kanser (Ali *et al.*, 2010). Namun begitu, lambakan data kanser yang datang dari pelbagai sumber menyebabkan kesukaran dalam mendapatkan data kanser yang bermakna. Selain itu, perbezaan istilah perubatan yang digunakan menyebabkan masalah ini semakin meruncing (Michael dan Yves, 2003). Menganalisis data yang berkaitan dan bersesuaian adalah perlu bagi mendapatkan hasil analisis yang tepat. Oleh itu, penyelidik perlu mengetahui dan memahami domain kajian yang dilakukan secara jelas dan terperinci.

Ontologi merupakan satu pendekatan yang mampu memberikan pemahaman bagi sesuatu domain dengan memberikan komunikasi yang baik antara manusia dan mesin dengan memformalkan perkataan berdasarkan konsep yang berkaitan (Brisson, 2007; Using *et al.*, 2010). Oleh itu, ontologi digunakan dalam kajian ini bagi mendapatkan pemahaman yang jelas dan membantu dalam pemilihan data yang akan digunakan dalam analisis ramalan risiko kanser payudara.

1.2 Latar Belakang Masalah

Kanser payudara terdapat dibahagian dada, dimana, kebiasaannya terbentuk dibahagian saluran susu dan kelenjar penghasilan susu. Kanser ini tidak hanya menyerang wanita, malah lelaki turut berpotensi mendapat kanser payudara walaupun kes kanser payudara dikalangan lelaki adalah jarang berlaku (*MedicineNet*, 2011). *National Cancer Institute* (2011), menyatakan, kes serangan kanser payudara yang dianggarkan pada tahun 2011 di Amerika Syarikat bagi kaum wanita adalah 230,480 kes dan 2,140 kes bagi lelaki. Manakala anggaran jumlah kematian adalah sebanyak 39,520 bagi wanita dan 450 bagi lelaki.

Berdasarkan *PubMed Health* (2010), purata bagi wanita yang didiagnosis dengan kanser payudara adalah sebanyak seorang daripada lapan wanita. Risiko berlakunya kanser payudara adalah lebih tinggi dengan peningkatan usia, malah, kebarangkalian bagi wanita untuk mendapat kanser payudara adalah seratus kali berbanding lelaki. Sejarah kanser payudara dalam keluarga turut meningkatkan risiko serangan kanser payudara, dimana, jumlah pesakit wanita yang mempunyai sejarah kanser payudara dikalangan ahli keluarga adalah sebanyak 20% hingga 30%.

Secara umumnya, kadar terjadinya kanser payudara di Asia adalah lebih rendah berbanding di negara barat. Namun begitu, kejadian kanser payudara di Asia meningkat dengan tinggi pada kadar masa yang cepat berbanding di negara barat (Yip, 2010). Di Semenanjung Malaysia, sebanyak 21,773 kes kanser telah

didiagnosis dan didaftarkan di *National Cancer Registry* (NCR) pada tahun 2006. Kes yang melibatkan kaum lelaki adalah sebanyak 9,974 kes dan 11,799 kes pula adalah dari golongan wanita. Kanser payudara merupakan kanser tertinggi yang direkodkan di Semenanjung Malaysia pada tahun 2006, iaitu sebanyak 3,525 kes (16.5%) daripada keseluruhan kes kanser yang didaftarkan (Kementerian Kesihatan Malaysia, 2006).

Menurut *National Cancer Institute* (2011), penggunaan teknologi maju adalah perlu bagi mengesan kanser dip peringkat awal dan mendapatkan maklumat risiko kanser bagi setiap pesakit. Justeru, memahami secara mendalam mengenai kanser payudara dan mengurangkan serta mengelakkan risiko yang boleh membawa kepada kematian adalah sangat diperlukan (*National Cancer Institute*, 2011). Kesedaran orang ramai dalam mencegah dan membendung masalah serangan kanser payudara masih membimbangkan. Dengan memberikan kefahaman yang mendalam tentang kanser payudara, kesedaran orang ramai akan meningkat dan seterusnya membantu mencegah dan membendung serangan kanser payudara pada peringkat awal. Mengesan serangan kanser payudara pada peringkat awal dapat mengurangkan risiko dan kos rawatan kanser. Selain itu, ia turut mampu mengurangkan jumlah pesakit kanser payudara serta mengurangkan kadar kematian yang diakibatkan oleh kanser tersebut (Shweta, 2012).

Pelbagai kaedah yang telah dikenalpasti dan boleh digunakan bagi analisis data seperti rangkaian neural, keputusan pepohon, *naïve bayes fusion* dan mesin sokongan vektor (SVM). Klasifikasi merupakan kaedah yang terkenal dan sering digunakan (Ayer *et al.*, 2010; Dancea *et al.*, 2008; Petalas *et al.*, 2003; Bellaachia dan Guven 2006) untuk membangunkan model ramalan yang tepat. Teknik SVM merupakan teknik yang baik dan mampu memberikan hasil ketepatan yang tinggi (Mert *et al.*, 2011; Ali *et al.*, 2010; Mallika dan Saravanan, 2010; Liu *et al.*, 2003).

Namun demikian, jumlah data yang banyak dan datang dari pelbagai sumber yang berlainan telah menyebabkan data menjadi tidak tersusun dan sukar untuk mendapatkan maklumat serta berkongsi pengetahuan dalam pangkalan data (Kusiak

et al., 2000; Sonia dan Khan 2007). Hal ini memberi cabaran baru bagi penganalisa untuk memahami istilah dan data dalam domain yang dianalisa bagi memilih data yang berkaitan (Coulet *et al.*, 2008; Gottgroy *et al.*, 2004). Rekod pesakit yang banyak dan mempunyai atribut yang pelbagai akan dianalisis bagi mengelaskan pesakit di dalam kelas risiko yang tertentu. Kajian ini membantu pakar perubatan dan penyelidik bagi membuat keputusan dalam rawatan dan meramal risiko kanser payudara tepat dan pantas.

Terdapat pelbagai kaedah penyelesaian bagi masalah jumlah data yang banyak, tidak tersusun dan datang dari sumber yang berbeza, seperti persekutuan data, pendekatan berorientasikan objek, capaian maklumat dan pengekstrakan. Namun begitu, pendekatan tersebut tidak memberikan hasil yang memberangsangkan selagi mempunyai rekabentuk pangkalan data yang terhad (Hua Min *et al.*, 2009). Tambahan lagi, pendekatan tersebut hanya menumpukan kepada masalah perbezaan struktur, tetapi tidak menyelesaikan masalah perbezaan semantik (Xiufen dan Yabin 2011). Menyatukan data dari sumber yang berlainan merupakan langkah yang penting bagi melancarkan proses analisis data dan menentukan hasil yang akan didapati. Pendekatan ontologi dalam bidang integrasi data mampu menyelesaikan masalah kepelbagaian semantik dan telah menarik perhatian ramai penyelidik. Umumnya, ontologi menggunakan skema istilah yang jelas bagi menyelesaikan masalah semantik (Buccella *et al.*, 2003; Wache *et al.*, 2001).

Ontologi mampu membantu penganalisa dalam memahami domain yang dilakukan, lantas dapat membantu memberikan hasil analisa yang memberangsangkan (Brisson, 2007). Kajian yang dilakukan oleh (Coulet *et al.*, 2008; Gottgroy *et al.*, 2004; Qi dan Tang, 2006; Brisson, 2007) menyatakan, pemilihan data yang dilakukan menggunakan ontologi dan pakar domain dapat memberikan data yang relevan dan hasil yang baik dalam melakukan analisis data. Kajian ini membangunkan ontologi kanser payudara bagi mendapatkan data yang berkualiti. Data yang berkualiti dalam kajian ini merupakan data yang berkaitan dan penting bagi melakukan ramalan risiko kanser. Data ini dipilih melalui hubungan antara atribut yang dilakukan dalam ontologi kanser payudara.

1.3 Penyataan Masalah

Membangunkan peramalan kanser payudara yang tepat amatlah diperlukan kerana kanser payudara merupakan penyebab utama kematian terutamanya dikalangan wanita. Peramalan kanser payudara yang tepat merupakan satu cabaran kerana permasalahan data kanser yang banyak dan diperolehi dari pelbagai sumber (Coulet *et al.*, 2008; Gottgroy *et al.*, 2004). Selain itu, kanser payudara merupakan satu domain yang kompleks dimana terdapat pelbagai istilah yang sukar difahami dan pelbagai ukuran yang digunakan bagi mengelas risiko kanser payudara.

Kavitha *et al.*, (2011) menyatakan, ontologi yang dibangunkan dapat menyelesaikan masalah semantik dan menyusun data dengan lebih konsisten. Kajian yang dilakukan oleh Brisson (2007), Coutlet *et al.* (2008), Qi dan Tang (2006) menyatakan bahawa ontologi dapat membantu dalam pemilihan data bagi mendapatkan analisis yang baik. Terdapat beberapa domain ontologi yang telah dibangunkan dalam bidang perubatan seperti *National Center for Biomedical Ontology* (NCBO) dan *Disease Ontology* (DO). Oleh itu, pembangunan ontologi kanser payudara adalah perlu bagi membantu meramal risiko kanser payudara dengan tepat serta membantu dalam pembuat keputusan terutamanya kepada doktor.

"Adakah ontologi yang dibangunkan mampu membantu menyusun data dengan lebih sistematik bagi membantu pemilihan data dengan lebih baik dalam meramal risiko kanser payudara?"

- i. Apakah sumber pengetahuan yang boleh digunakan bagi membangunkan ontologi kanser payudara?
- ii. Apakah kumpulan data yang sering digunakan bagi melakukan analisis risiko kanser payudara?
- iii. Apakah pendekatan pembangunan ontologi yang sesuai digunakan untuk membangunkan ontologi kanser payudara?
- iv. Bagaimanakah pendekatan ontologi kanser payudara dilakukan untuk mendapatkan data yang relevan bagi analisis data?

- v. Apakah teknik yang baik dan sering digunakan dalam meramal risiko kanser?
- vi. Bolehkah model klasifikasi SMO digunakan bagi mengelas risiko kanser dan memberikan hasil yang baik?
- vii. Sejauh manakah ontologi kanser payudara yang dibangunkan dapat memberikan data yang relevan?

1.4 Objektif

Objektif utama yang perlu dicapai dalam kajian ini adalah untuk mendapatkan data yang berkualiti bagi meramal risiko kanser payudara dan menyatukan data dari sumber yang berlainan. Selain itu, kajian ini juga bertujuan untuk mencapai objektif berikut:

- i. Membangunkan ontologi kanser payudara dari sumber yang berlainan bagi pemilihan data.
- ii. Membangunkan model ramalan risiko kanser payudara menggunakan kaedah klasifikasi berdasarkan ontologi kanser payudara.

1.5 Skop

Skop telah ditentukan dalam kajian ini bagi memastikan kajian ini mengikut spesifikasi tertentu dan mencapai kehendak kajian. Berikut merupakan skop bagi kajian ini yang mana memfokuskan kepada:

- i. Kajian ini menggunakan dua set data kanser payudara yang diambil dari *Wisconsin Breast Cancer* dan *Wisconsin Diagnostic Breast Cancer*.
- ii. Pengumpulan istilah kanser payudara turut diambil daripada *Wikipedia Breast Cancer*, *PubMed Health*, *Breastcancer.org*, *MedicineNet* dan *Resource for Health Professionals, California*.

- iii. Kajian ini menumpukan kepada pembangunan ontologi kanser payudara dalam fasa pra-pemprosesan data bagi melakukan integrasi data.
- iv. Kajian ini menggunakan pendekatan klasifikasi SMO bagi meramal risiko kanser payudara.

1.6 Kepentingan Kajian

Peramalan data telah digunakan secara meluas dalam pelbagai bidang dan telah berkembang dalam bidang perubatan. Bagi menguruskan data kanser payudara, kajian ini menggunakan pendekatan ontologi dalam melakukan integrasi data. Lambakan data kanser menyebabkan data tersebut sukar difahami selain pengetahuan berguna mungkin tersembunyi. Analisis ini dilakukan bagi mengelaskan pesakit kanser payudara ke dalam kelas risiko kanser yang tertentu. Kajian ini menggunakan data kanser payudara yang diambil dari sumber yang berbeza. Menyatukan data dari sumber yang berbeza adalah penting untuk mendapatkan pemahaman data yang lebih baik bagi melancarkan proses analisis data. Ontologi menyatukan data ke dalam satu hubungan dan mampu menyelesaikan masalah semantik selain mampu memberikan hasil yang lebih memberangsangkan. Dengan pendekatan pembangunan ontologi kanser payudara ini, ia dapat menyatukan data dari sumber berbeza dan membantu penganalisa memahami data yang akan digunakan.

Mengesan kanser pada peringkat awal adalah penting bagi meningkatkan kelangsungan hidup pesakit. Berdasarkan kelas risiko yang dilakukan dalam kajian ini, ia mampu membantu bagi mempercepatkan penentuan kaedah rawatan yang akan dilakukan. Dengan pengesanan awal yang dilakukan, pesakit mampu mendapatkan rawatan segera sebelum kanser payudara tersebut menjadi lebih parah dan membantu mengurangkan kos perubatan yang lebih tinggi selain dapat mengurangkan jumlah pesakit kanser payudara di Malaysia. Pembangunan ontologi mampu menyusun data perubatan dengan lebih baik dan ini memudahkan proses capaian maklumat dilakukan. Dalam kajian ini, ontologi dibangunkan bagi membantu pemilihan data dalam melakukan analisis risiko kanser payudara.

1.7 Kesimpulan

Punca utama kematian dikalangan wanita diseluruh dunia adalah berpunca dari serangan kanser payudara. Ramalan risiko kanser payudara merupakan satu langkah yang penting bagi mencegah risiko kanser pada tahap risiko yang berbeza serta meningkatkan tahap kesihatan di Malaysia. Ontologi merupakan satu metodologi baru yang mampu membantu analisis data dalam pangkalan data yang kompleks. Ontologi menyatukan data daripada sumber yang berlainan bagi melicinkan proses analisis data. Salah satu masalah yang mencabar bagi meramal risiko kanser payudara adalah kesukaran memahami data dari domain yang bukan dalam kepakaran penganalisa. Oleh itu, adalah penting untuk melakukan pemodelan ontologi bagi mendapatkan data yang relevan dan hasil ramalan risiko yang tepat dengan menggunakan kaedah klasifikasi SMO.

RUJUKAN

- Ali, A., Khan, U., Tufail A. dan Kim, M. 2010. Analyzing Potential of SVM based Classifiers for Intelligent and Less Invasive Breast Cancer Prognosis. *Second International Conference on Computer Engineering and Applications*. vol.2, no., pp.313-319, 19-21 March 2010.
- American Cancer Society. 2012a. Breast Cancer. Tidak diterbitkan.
<http://www.cancer.org/cancer/cancerbasics/what-is-cancer>
- American Cancer Society. 2012b. Breast Cancer: Early Detection. Tidak diterbitkan.
<http://www.cancer.org/acs/groups/cid/documents/webcontent/003165-pdf.pdf>
- Ayer, T., Chhatwal, J., Alagoz, O., Khan Jr, C. E., Woods, R. W. dan Burnside, E. S. 2010. Comparison of Logistic Regression and Artificial Neural Network Models in Breast Cancer Risk Estimation, *Journal of Radiographics*, 30(1);13-22.
- Aykut Firat, Stuart E. Madnick, dan Benjamin N. Grosz. 2002. Knowledge Integration to Overcome Ontological Heterogeneity: Challenges from Financial Information Systems. *ICIS, page 17. Association for Information Systems*
- Bellaachia, A. dan Guven, E. 2006. Predicting Breast Cancer Survivability using Data Mining Techniques. *Ninth Workshop on Mining Scientific and Engineering Datasets in conjunction with the Sixth SIAM International Conference on Data Mining (SDM 2006)*, 1-4.
- Bellcross, C. 2009. Approaches to applying breast cancer risk prediction models in clinical practice, *Community Oncology*, 6;373-379. 2009.
- Bermejo J., 2007, A Simplified Guide to Create an Ontology. *Madrid University*
- Bianchi, S.; Burla, A.; Conti, C.; Farkash, A.; Kent, C.; Maman, Y.; Shabo, A.; 2009 "Biomedical data integration – capturing similarities while preserving disparities," *Engineering in Medicine and Biology Society, 2009. EMBC*

2009. *Annual International Conference of the IEEE*, vol., no., pp.4654-4657, 3-6 Sept. 2009

Breast Cancer Wisconsin (Diagnostic) Data Set. Didapati:

<http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29> [Di akses September 2011]

Breast Cancer Wisconsin Data Set. Didapati:

<http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29> [Di akses September 2011]

Breastcancer.org. 2011 Didapati:

<http://www.breastcancer.org/symptoms/diagnosis/staging>. [Di akses September 2011]

Brisson, L. 2007. Knowledge extraction using a conceptual information system (ExCIS). *Proceedings of the First and Second VLDB conference on Ontologies-based databases and information systems*. Trondheim, Norway. P 119-134. 2007. Springer-Verlag.

Buccella, A., Cechich, A., Cechich, R. dan Brisaboa, N. R. 2003. An Ontology Approach to Data Integration. *Journal of Computer Science and Technology*. Vol (3). Pp. 62—68.

Cancer Research UK. 2013 <http://www.cancerresearchuk.org/cancer-info/cancerstats/world/>

Cancer.net. 2011 Didapati: <http://www.cancer.net/cancer-types/breast-cancer/staging-illustrations> [Di akses September 2011]

Castillo, J. A. R., Silvescu, A., Caragea, D., Pathak, J., dan Honavar, V. G. 2003. Information Extraction and Integration from Heterogeneous, Distributed, Autonomous Information Sources- A Federated -Ontology-Driven QueryCentric Approach. *IEEE International Conference on Information Reuse and Integration*. 183-191.

Chen, M. C., Liao, H. C. dan Huang, C. L. 2006. Predicting Breast Tumor via Mining DNA Viruses with Decision Tree, *IEEE International Conference on Systems, Man and Cybernetics*, 2006. SMC '06. pp.3585-3589, 8-11 Oct. 2006.

Chen, M. C., Liao, H. C. dan Huang, C. L. 2006. Predicting Breast Tumor via Mining DNA Viruses with Decision Tree, *IEEE International Conference on*

- Systems, Man and Cybernetics*, 2006. SMC '06. pp.3585-3589, 8-11 Oct. 2006.
- Ching-Chieh, K. dan Chien-Sing, L. 2006. A Data Mining Approach for Managing Shared Ontological Knowledge, *Sixth International Conference on in Advanced Learning Technologies*, 2006. pp. 16-18.
- Chirathamjaree, C. dan Mukviboonchai, S. 2002, "The mediated integration architecture for heterogeneous data integration," *TENCON '02. Proceedings. 2002 IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering* , vol.1, no., pp. 77- 80 vol.1, 28-31 Oct. 2002
- Compumine. Didapati:
<http://www.compumine.com/web/public/newsletter/20071/precision-recall>.
 [Di akses September 2011]
- Coulet, A., Tabbone, M. S., Benlian, P., Napoli, A. dan Devignes, M. D. 2008. Ontology-guided data preparation for discovering genotype-phenotype relationships. *BMC Bioinformatics*, Vol. 9, No. Suppl 4.
- Cruz, I. F. dan Xiao, H. 2005. The Role of Ontologies in Data Integration. *Journal Of Engineering Intelligent Systems*. V(13). 245-252.
- Dai, H. 1997 "An object-oriented approach to schema integration and data mining in multiple databases," *Technology of Object-Oriented Languages, 1997. TOOLS 24. Proceedings* , vol., no., pp.294-303, Sep 1997
- Dancea, O., Gordan, M., Dragan, M., Stoian, I. dan Nedevschi, S. 2008. Postoperative risk classification of prostate cancer patients using support vector machines, *IEEE International Conference on Automation, Quality and Testing, Robotics, 2008. AQTR 2008.*, pp.53-56, 22-25 May 2008
- Dean Jones, Trevor Bench-capon dan Pepijn Visser. 1998. Methodologies For Ontology Development. P. 62-75. *In Proc. IT&KNOWS Conference of the 15th IFIP World Computer Congress, Budapest, 1998*. Chapman-Hall.
- Disease Ontology. Didapati: <http://www.disease-ontology.org/>. [Di akses Julai 2012].
- Djedidi, R. dan Aaufaure, M., 2007 Medical Domain Ontology Construction Approach: A Basis for Medical Decision Support, *Computer-Based Medical Systems, 2007. CBMS '07. Twentieth IEEE International Symposium on*, vol., no., pp.509,511, 20-22 June 2007

- Elsayed, A. E., El-Beltagy, S. R., Rafea, M. dan Hegazy, O. 2007. Applying Data Mining for Ontology Building. *The 42nd Annual Conference On Statistics, Computer Science, and Operations Research*.
- Erin Davis, B.A., Martha Schechtel, R.N., Bruin Ruge, M.D., and David Hickam, M.D., 2010. Reducing the Risk of Breast Cancer With Medicine: A Guide for Women. *Eisenberg Center*.
- Eshlaghy AT, Poorebrahimi A, Ebrahimi M, Razavi AR, Ahmad LG 2013 Using Three Machine Learning Techniques for Predicting Breast Cancer Recurrence. *J Health Med Inform* 4: 124.
- Fonseca, F., Egenhofer, M. J., Agouris, P., Cmara, C., Fonseca, F., Egenhofer, M. J. Agouris, P. dan Câmara, G. 2002. Using Ontologies for Integrated Geographic Information Systems. *Transactions in GIS* 6: 231–57
- Furey, T. S., Cristianini, N., Duffy, N., Bednarski, D. W., Schummer, M. dan Haussler, D. 2000. Support Vector Machine Classification And Validation Of Cancer Tissue Samples Using Microarray Expression Data. *Bioinformatics*. 2000 Oct;16(10):906-14.
- García, J., Jose, F., Peñalvo, G. and Therón, R. 2010. A Survey on Ontology Metrics. *Proceedings Third World Summit on the Knowledge Society, WSKS 2010, Corfu, Greece. Springer Berlin Heidelberg*. pp22-27, 22-24September 2010
- Gong, P., Qu, W. and Feng, D.D.; 2006 , "An Ontology for the Integration of Multiple Genetic Disorder Data Sources," *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the* , vol., no., pp.2824-2827, 17-18 Jan. 2006
- Gottgroy, P., Kasabov, N., and MacDonell, S. G. 2004. An Ontology Driven Approach for Knowledge Discovery in Biomedicine. *in Proceeding of the 8th Pacific Rim International Conference on Artificial Intelligence (PRICAI)*. Auckland, New Zealand, Springer-Verlag, Vol, 3157, pp.53-67
- Goyal, S., Kumari, L. dan Khandelwal, P. 2005. Data Integration through Data Federation. Tidak diterbitkan
- Guangfei, Y., Shimada, K., Mabu, S., Hirasawa, K. dan Jinglu, H. 2007. A genetic network programming based method to mine generalized association rules with ontology. *Annual Conference in SICE, 2007*. pp. 2715-2722.
- Haas, L. M.; Lin, E. T.; Roth, M. A.; 2002. "Data integration through database federation," *IBM Systems Journal* , vol.41, no.4, pp.578-596, 2002

- Hadzic, M.; Dillon, T.; Chang, E., 2007. Use of Digital Ecosystem and Ontology Technology for Standardization of Medical Records, *Digital EcoSystems and Technologies Conference, 2007. DEST '07. Inaugural IEEE-IES*, vol., no., pp.595,601, 21-23 Feb. 2007
- Han, J. dan Kamber, M. 2006. Data Mining: Concepts and Techniques. 2nd ed. *United States of America. Diane Cerra. 743.*
- Horridge, M., Knublauch, H., Rector, A., Stevens, R. dan Wroe, C. 2011. A Practical Guide To Building OWL Ontologies Using Protégé 4 and CO-ODE Tools, Edition 1.3, *published by the University of Manchester*, 24 Mar 2011, 108 pp.
- Hua Min, Frank J. Manion, Elizabeth Goralczyk, Yu-Ning Wong, Eric Ross, and J. Robert Beck. 2009. Integration of Prostate Cancer Clinical Data Using an Ontology. *J Biomed Inform.* 2009 December ; 42(6): 1035–1045.
- Jonsdottir, T., Hvannberg, E. T., Sigurdsson, H. dan Sigurdsson, S. 2008. The feasibility of constructing a Predictive Outcome Model for breast cancer using the tools of data mining. 34(1). 2008. 108-118.
- Kates, R., Harbeck, v. dan Schmitt, M. 2000. Prospects for clinical decision support in breast cancer based on neural network analysis of clinical survival data, *Fourth International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies*, 2000., pp.764-767.
- Kavitha, C., Sadasivam, G. S., dan Shenoy, S. N. 2011. *Ontology Based Semantic Integration of Heterogeneous Databases*. European Journal of Scientific Research. Vol.64 No.1, pp. 115-122.
- Kementerian Kesehatan Malaysia. 2006. Statistics-Data And Figure Peninsular Malaysia 2006. *National Cancer Registry*, Malaysia. 1-112.
- Khan, F.M.; Arnold, M.G.; Pottenger, W.M.; 2004. "Finite precision analysis of support vector machine classification in logarithmic number systems," *Symposium on Digital System Design, 2004. DSD 2004. Euromicro*, vol., no., pp. 254- 261, 31 Aug.-3 Sept. 2004
- Kim, S dan Song, J. 2007. A Study on Ontology Based Medical Information System for Home Healthcare, *Intelligent Pervasive Computing, 2007. IPC. The 2007 International Conference on* , vol., no., pp.364,367, 11-13 Oct. 2007
- Kusiak, A., Kern, J.A., Kernstine, K.H. and Tseng, B.T.L. 2000. Autonomous decision-making: a data mining approach, *IEEE Transactions on Information Technology in Biomedicine*. Dec. 2000. pp.274-284

- Lee J. R. dan Payam, B. M. 2007 *Semantic Media in Medical Information Systems, Semantic Media Adaptation and Personalization, Second International Workshop on* , vol., no., pp.27,31, 17-18 Dec. 2007
- Leung, N. K. Y., Lau, S. K., Fan, J. dan Tsang, N, 2011. An integration-oriented ontology development methodology to reuse existing ontologies in an ontology development process. *Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services. Ho Chi Minh City, Vietnam.* 174-181
- Liu Y. Q., Wang C. dan Zhang L., 2009. Decision Tree Based Predictive Models for Breast Cancer Survivability on Imbalanced Data, *3rd International Conference on Bioinformatics and Biomedical Engineering* , 2009. ICBBE 2009., vol., no., pp.1-4, 11-13 June 2009
- Liu, H. X., Zhang, R. S., Luan, F., Yao, X. J., Liu, M. C., Hu, Z. D., dan Fan, B. T. 2003. Diagnosing Breast Cancer Based on Support Vector Machines. *Journal of Chemical Information and Modeling*, Vol. 43, No. 3., pp. 900-907.
- Mallika, R. dan Saravanan, V. 2010. An SVM based Classification Method for Cancer Data using Minimum Microarray Gene Expressions. *World Academy of Science, Engineering and Technology* 62 2010.
- Mark A. Musen. 2013. National Center for Biomedical Ontology (NCBO) Receives Funding through 2015. Professor of Medicine (Biomedical Informatics Research) and of Computer Science.
http://medicine.stanford.edu/news_events/ncbopress.html
- Medic on Web 2011 Didapati: <http://mediconweb.com/cancer/breast-cancer-stages/2/>
[Di akses September 2011]
- MedicineNet.com 2011 Didapati <http://www.medicinenet.com/script/main/hp.asp>
[Di akses Mac 2011]
- Mert, A., Kilic, N., dan Akan, A. 2011. Breast Cancer Classification by Using Support Vector Machines with Reduced Dimension. *53rd International Symposium ELMAR-2011*, 14-16 September 2011, Zadar, Croatia.
- Michael N. Cantor, dan Yves A. Lussier, 2003 Putting Data Integration into Practice: Using Biomedical Terminologies to Add Structure to Existing Data Sources. *AMIA Annual Symposium Proceedings.* 2003; 125–129.

- Michel Gagnon. 2007. Ontology-Based Integration of Data Sources. *Conference on Information Fusion, 2007 10th International*, vol., no., pp.1-8, 9-12 July 2007.
- Natalya, F. N. dan Deborah, L. M. 2011. Ontology Development 101: A Guide to Creating Your First Ontology
http://protege.stanford.edu/publications/ontology_development/ontology_101-noy-mcguinness.html. [3 Mac 2011].
- National Cancer Institute. 2011. Didapati:
<http://www.cancer.gov/cancertopics/types/breast>. [Di akses September 2011]
- National Center for Biomedical Ontology. Didapati:
<http://bioportal.bioontology.org/>. [Di akses February 2012].
- Nigro, H.O., Císaro, S.E.G. dan Xodo, D.H.. 2008 Data Mining with Ontologies: Implementations, Findings, and Frameworks. *Information Science Reference. the University of Michigan*. 289 pages
- Nimmagadda, S. L., Dreher, H., dan Rudra, A. 2005. Ontology of Western Australian petroleum data for effective data warehouse design and data mining, *3rd IEEE International Conference on in Industrial Informatics, 2005. INDIN '05*. 2005, pp. 584-592.
- Patrick, F. dan Florian, H. 2010. Ontology Evaluation. *Seminar in Applied Ontology Engineering WS 10/11*.
- Perez-Rey D, Maojo V, García-Remesal M, Alonso-Calvo R, Billhardt H, Martín-Sánchez F, dan Sousa A. 2006a. ONTOFUSION: ontology-based integration of genomic and clinical databases. *Comput Biol Med.* 2006 Jul-Aug;36(7-8):712-30.
- Perez-Rey,D., Anguita, A. dan Crespo, J. 2006b. OntoDataClean: Ontology-Based Integration and Preprocessing of Distributed Data. *Biological And Medical Data Analysis Lecture Notes in Computer Science, 2006*, Volume 4345/2006, 262-272
- Petalas, P., Spyridonos, P., Glotsos, D., Cavouras, D., Ravazoula, P. dan Nikiforidis, G. 2003. Probabilistic neural network analysis of quantitative nuclear features in predicting the risk of cancer recurrence at different follow-up times, *Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis, 2003. ISPA 2003.*, pp. 1024- 1027, 18-20 Sept. 2003.

- Pinto, F. M. dan Santos, M. F. 2009. Considering Application Domain Ontologies for Data Mining. *World Scientific and Engineering Academy and Society, (WSEAS)WSEAS Trans. Info. Sci. and App.* 478-1492.
- Pinto, F., Santos, M. F. dan Marques, A. 2009. Ontology based data mining: a contribution to business intelligence, *Proceedings of the 10th WSEAS international conference on Mathematics and computers in business and economics*, Prague, Czech Republic.
- Platt, J. C. 1999. Fast training of support vector machines using sequential minimal optimization. *Advances in kernel methods*. 185--208. MIT Press. Cambridge, MA, USA.
- PubMed Health. 2010. Didapati:
<http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0001911/>. [Di akses September 2011]
- Qi, J. dan Tang, J. 2006. Gene Ontology Driven Feature Selection from Microarray Gene Expression Data, *Computational Intelligence and Bioinformatics and Computational Biology, 2006. CIBCB '06. 2006 IEEE Symposium on* , vol., no., pp.1,7, 28-29 Sept. 2006
- Rajpathak, D dan Chougule, R. 2011: A generic ontology development framework for data integration and decision support in a distributed environment, *International Journal of Computer Integrated Manufacturing*, 24:2, 154-170.
- Resource for Health Professionals, California.. Didapati:
<http://qap.sdsu.edu/resources/websites/> [Di akses September 2011]
- Santoso, H. A., Haw, S. C. dan Abdul-Mehdi, Z. T. 2010. Ontology extraction from relational database: Concept hierarchy as background knowledge, *Knowledge-Based Systems, Volume 24, Issue 3*, Pages 457-464
- Shweta Kharya. 2012. USING Data Mining Techniques For Diagnosis And Prognosis Of Cancer Disease. *International Journal of Computer Science, Engineering and Information Technology (IJCSEIT)*, Vol.2, No.2, April 2012
- Simon Twigger. 2013. Data Management and Integration in the Biological Sciences. Medical College of Wisconsin.
<http://www.bioontology.org/NCBO%20user%20Simon%20Twigger>
- Sonia, K.; dan Khan, S. 2007 "Building Local Ontology from Database Relations in Data Integration," *Emerging Technologies, 2007. ICET 2007. International Conference on* , vol., no., pp.108-113, 12-13 Nov. 2007

- Stephen P. Gardner, Ontologies and semantic data integration, Drug Discovery Today, Volume 10, Issue 14, 15 July 2005, Pages 1001-1007, ISSN 1359-6446.
- Sufyian, M. A. M. dan Zaihisma, C. C. 2010. Breast Cancer prediction based on Backpropagation Algorithm. *Conference on Research and Development (SCORED), 2010 IEEE Student*, vol., no., pp.164-168, 13-14 Dec. 2010
- Thu, T.N.T. dan Davis, D. N. 2006. Predicting Cardiovascular Risks using POSSUM, PPOSSUM and Neural Net Techniques. *Proceedings of the Eighth International Conference on Enterprise Information Systems: Databases and Information Systems Integration*, Paphos, Cyprus, May 23-27. 230-234.
- Ushold, M. dan King, M. 1995. Towards a Methodology for Building Ontologies. *Workshop on Basic Ontological Issue in Knowledge Sharing*. Conjunction with IJCAI-95
- Using, S. N. M., Ahmad, R. dan Taib, S. M. 2010. Ontology of programming resources for semantic searching of programming related materials on the Web. *International Symposium in Information Technology (ITSim)*, 2010 pp. 698-703.
- Wache, H., Vögele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H. dan Hübner, S. 2001. Ontology-Based Integration of Information - A Survey of Existing Approaches. Pp 108—117.
- Wafa Mokharrak, Nedhal Al Khalaf, Tom Altman. 2012. Application of Bioinformatics and Data Mining in Cancer Prediction. *WORLDCOMP'12 - The 2012 World Congress in Computer Science, Computer Engineering, and Applied Computing* July 18, 2012.
- Wang, H. dan Ye, Z. 2009 Building Multi-Level Data Warehouse Based on Hybrid-Ontology, *International Symposium on Computer Network and Multimedia Technology, 2009. CNMT 2009.*, vol., no., pp.1,4, 18-20 Jan. 2009
- Wikipedia Breast Cancer. Didapati: http://en.wikipedia.org/wiki/Breast_cancer [Di akses September 2011]
- Wisconsin University, Science Computer. Didapati: <http://pages.cs.wisc.edu/~bolo/shipyard/neural/local.html> [Di akses September 2011]

- Wolberg, W.H., Street, W.N., Heisey, D.M. dan Mangasarian. O.L. 1995
Computerized breast cancer diagnosis and prognosis from fine needle aspirates. *Archives of Surgery* 1995;130:511-516.
- World Health Organization. 2011. Didapati
<http://www.who.int/mediacentre/factsheets/fs310/en> [Di akses September 2011]
- Wu, K. P. dan Wang, S. D. 2009.Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space, *Pattern Recognition*, Volume 42, Issue 5, May 2009, Pages 710-717
- Xiufen; C. dan Yabin; X. 2011 , "Computer-based patient record data integration method based on ontology," *International Symposium on IT in Medicine and Education (ITME)*, 2011, vol.2, no., pp.551-554, 9-11 Dec. 2011
- Yip, C. H. 2010. Breast Cancer in Asia. *Department of Surgery, University Malaya Medical Centre*. Current issues in Breast Cancer, 3 July 2010.