Effective Clinical Risk Assessment Using Ontology Driven and Machine Learning Approach

UNIVERSITY OF STIRLING

Kamran Farooq (SMIEEE)
IEEE UK&RI Industrial Applications
Doctoral Fellow
COSIPRA Lab
University of Stirling, Scotland

Project Collaborators

EPSRC funded PhD project born from the collaboration between the University of Stirling (Scotland), Sitekit Solutions Ltd. (Scotland), MIT (USA) and the Harvard Medical School (USA).

Project Team:

Kamran Farooq: Doctoral Fellow

Prof. Amir Hussain: Principal Investigator

Chris Eckl : Sitekit Healthcare Labs

Prof. Calum MacRae: Cardiologist, Harvard Medical School

Prof. Warner Slack: Founding Head of Clinical Informatics Division, Harvard Medical School

Prof. Stephen Leslie : Cardiologist, Raigmore Hospital

Hicham Atassi : Brno University of Technology

Dr John Moore : MIT New Media Lab

COSIPRA Collaborators

Prof. Francesco Piazza

University of Ancona, Italy

Prof. Zdenek Smékal

BUT, Department of Telecommunications, Czech Republic

Prof. Anna Esposito

International Insstitute of Advanced Scientific Studies, Italy

Dr. James Bron, Dr. Andy Shinn

Stirling Institute of Aquaculture

Prof. Mohamed Chetouani

Marie-Curie University, Paris, France

Prof. Yousef Al-Otaibi

King Saud University, Saudi Arabia

Prof. Kevin Gurney

University of Sheffield, UK

Prof. Kevin Gurney

University of Sheffield, UK

Prof. Mike Grimble

Industrial Control Centre, University of Strathclyde

Prof. Warner Slack

Harvard Medical School

Dr. Vassilis Cutsuridis

Boston University, USA

Prof. Igor Aleksander

Imperial College, London

Naeem ur Rahman

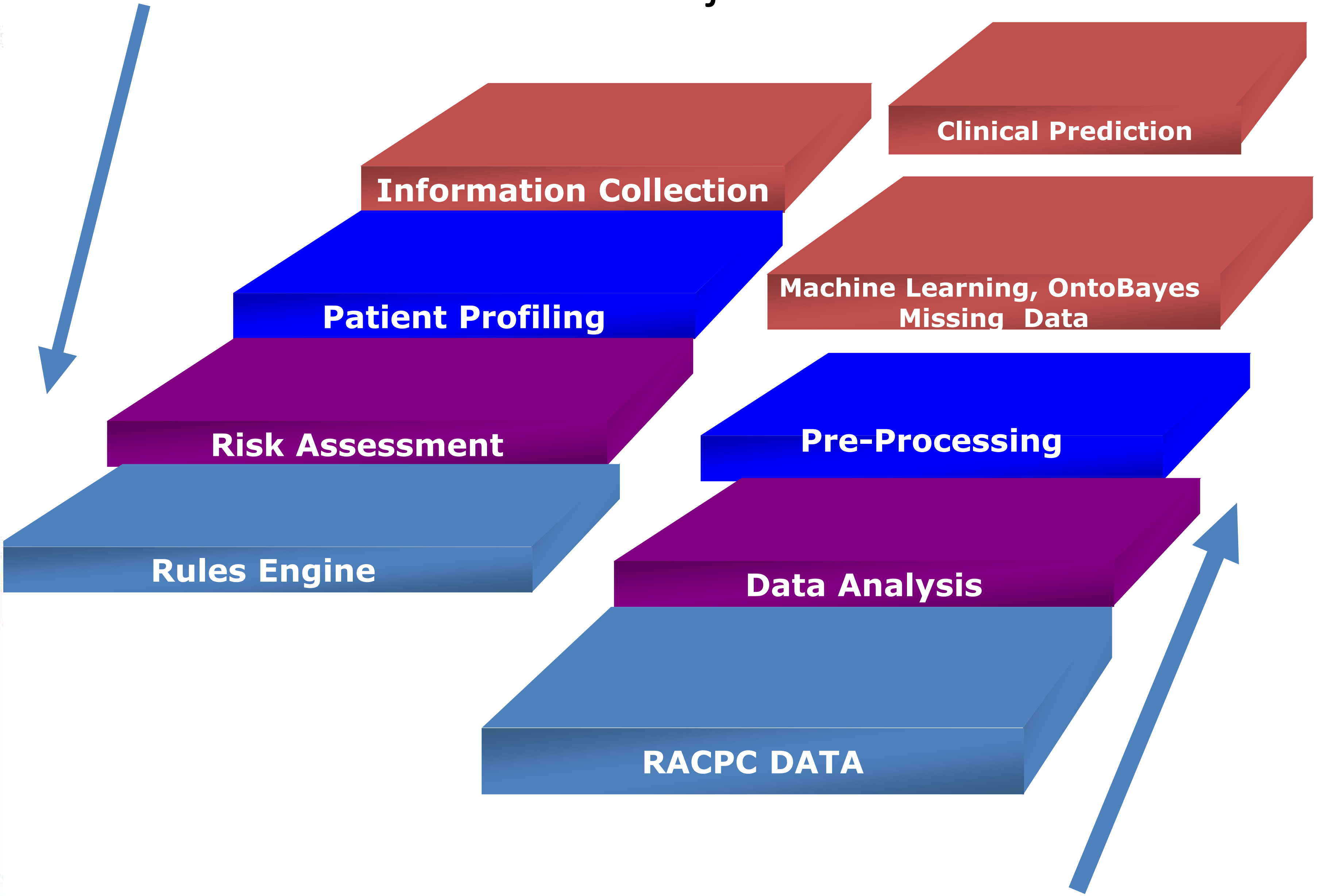
United Technologies (Unitec) LLC, UAE

Prof. Alexander Gelbukh

National Polytechnic Institute, Mexico

Contents

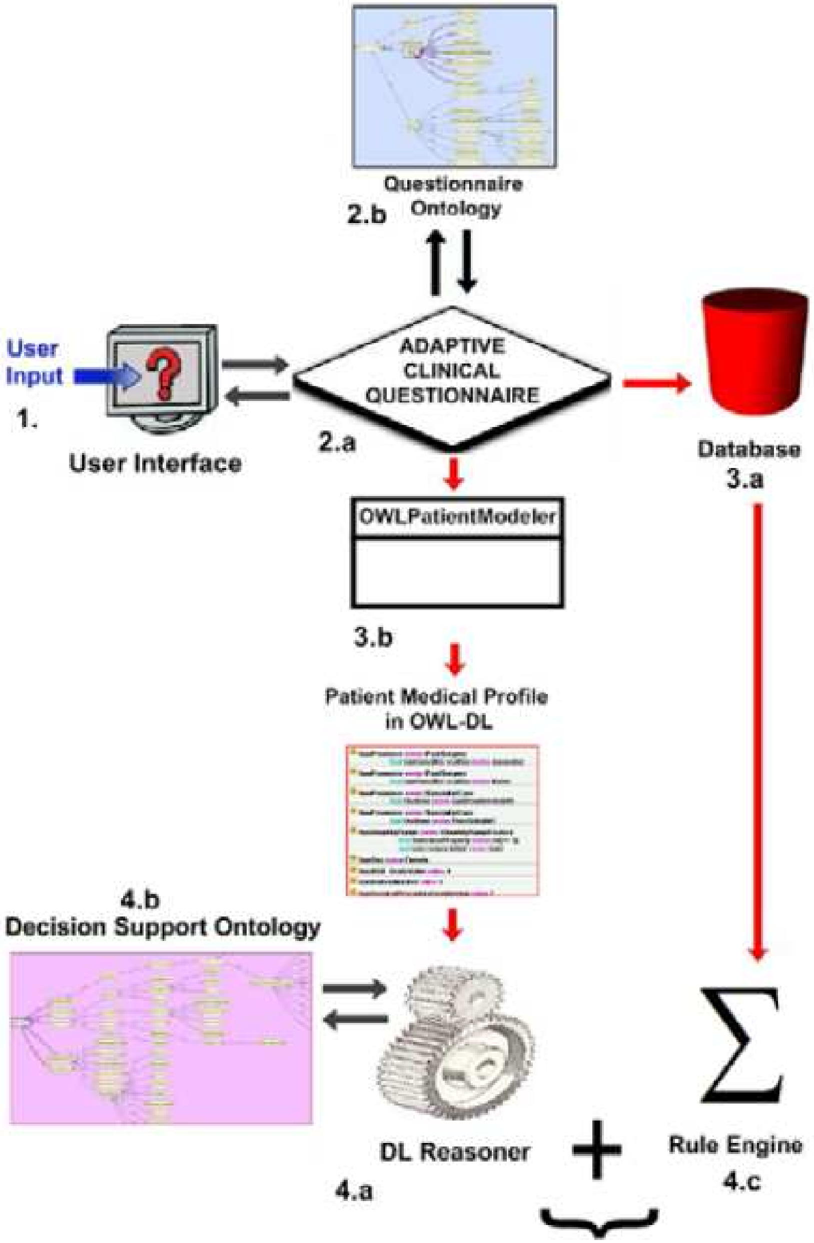
Research Objectives



Ontology Driven Prospective Cardiovascular Decision Support Framework

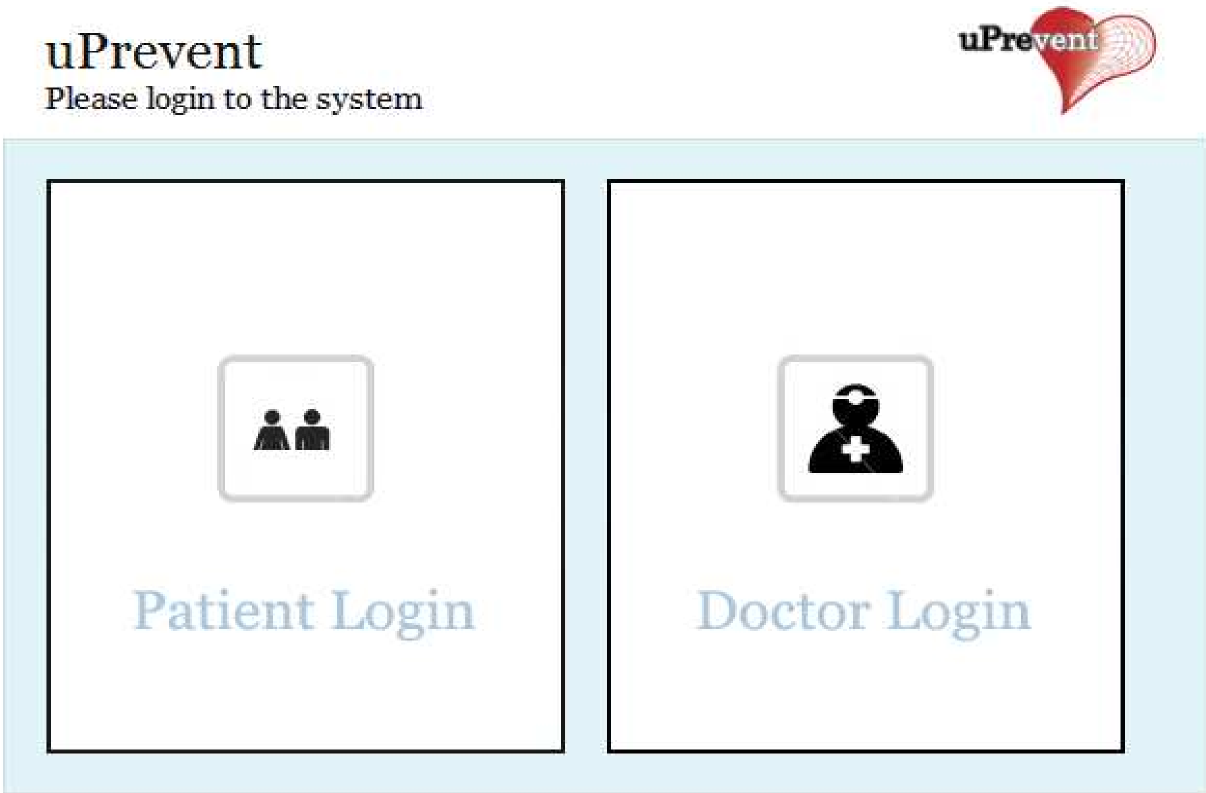
Key components of OntoCDSS

- Context sensitive Interviewing/screening questionnaires
- Patient Semantic Profiling (Electronic Patient Summary/Records
- Risk assessment of patients using domain ontologies based on Clinical Practice Guidelines
- Rules Engine (Heart Risk calculation)



Ontology Driven Prospective Clinical Decision Support System

1. Prospective Cardiovascular Decision Support System



Hello doctor

Here you will be able to able to register a patient and view their progress

Add a patient

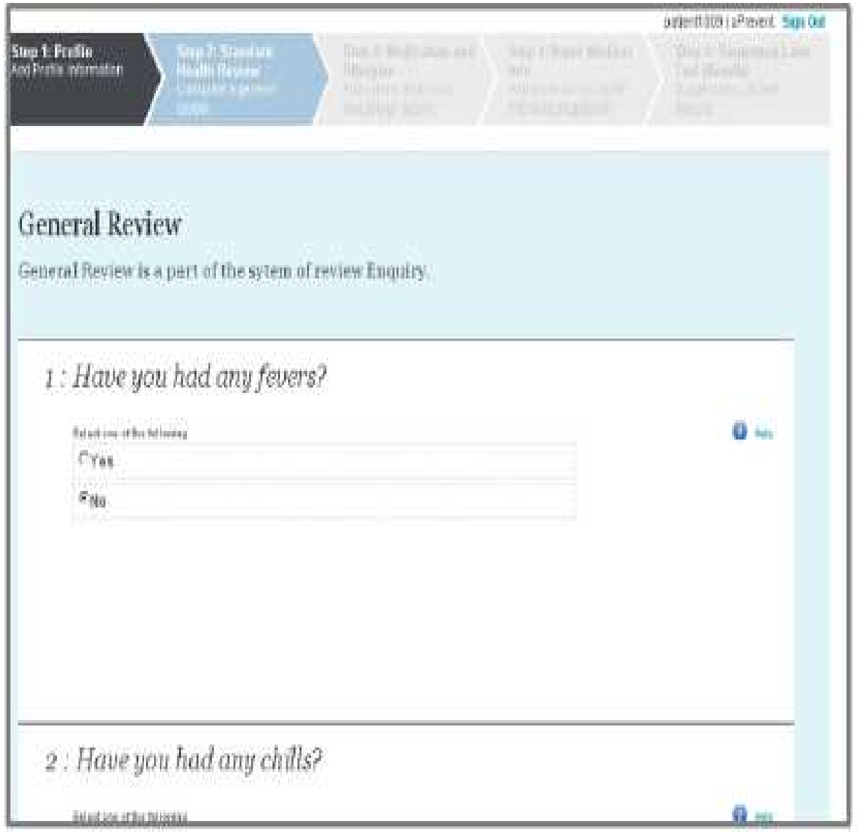
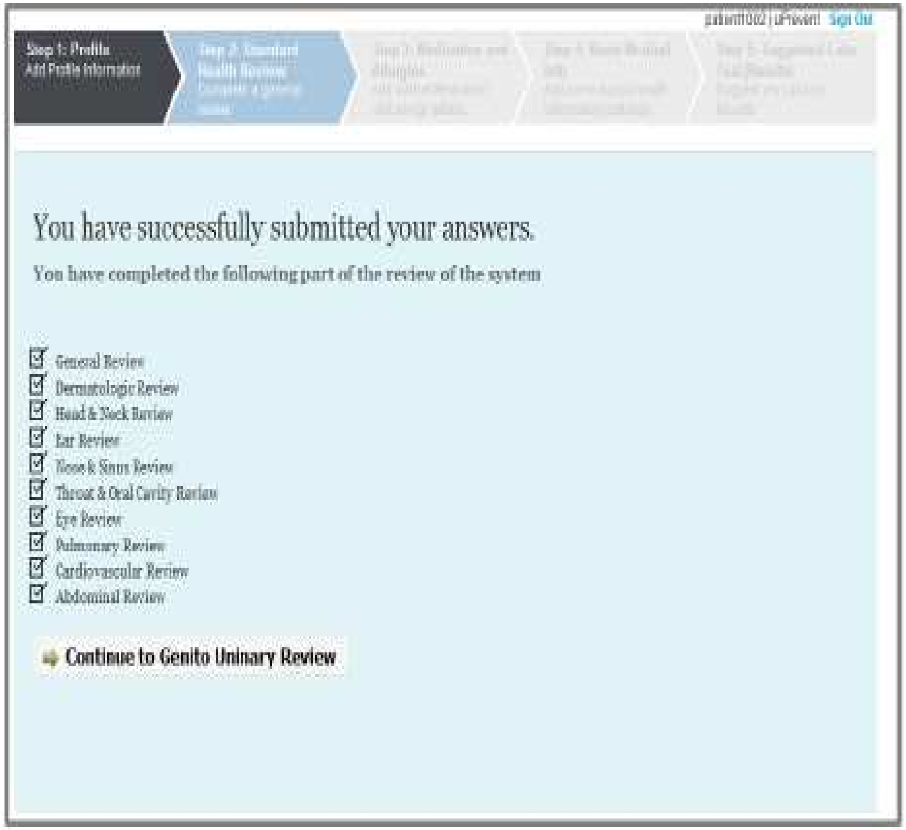
Search for a patient

Search for patients in the database and monitor his progress.

Search | View all

Search results

Patient	Profile	Review	Medication	Medical Details
CRDam				
Duncan	✓	✓		
erik	✓	✓	✓	
jacob	✓	✓	✓	✓
JLRoose	✓		✓	✓
jo-anne	✓	✓	✓	
kfarooq	✓	✓	✓	
marion0215				
patient-test01				
patient01s				
patient0210	✓	✓	✓	✓
patient0211	✓	✓	✓	✓
patient0212				


Sign Out
Step 1: Profile
Add Profile information
Step 2: Standard Health Review
General Review
General Review is a part of the sytem of review Enquiry.
1 : Have you had any fevers?
Yes
No
2 : Have you had any chills?

Sign Out
Step 1: Profile
Add Profile Information
Step 2: Standard Health Review
You have successfully submitted your answers.
You have completed the following part of the review of the system
General Review
Dermatologic Review
Head & Neck Review
Ear Review
Nose & Sinus Review
Throat & Oral Cavity Review
Eye Review
Pulmonary Review
Cardiovascular Review
Abdominal Review
Continue to Genito Uninary Review

Patient Details
Here you are viewing the patient basic medication list and his profile details
Go back to Doctor Home
View System Review Answers
View Risk Assesment
View Suggested Lab - Tests
Medication Review
Prescribed Medication List
MEDICATION
DOSAGE
HOW OFTEN
None
Medication Bought off Counter
MEDICATION
DOSAGE
HOW OFTEN
Vitamin C
100mg
Everday
Allergies to Medications
MEDICATION
ALLERGY
SYMPTOMS
Cod Liver Oil
Rash
Profile
Age
42 years old
Gender
female
Weight
Height
Ethnicity
White
Body Mass Index (BMI)
Medical Details
Blood Pressure
Cholestrol
Smoker
yes
Diabetes
no

Risk Assesment of the patient
The patient's estimated risk for a cardiovascular event occuring over a given time period are mentioned here.
Go Back to Patient Details
View Risk Assesment
View Suggested Lab - Tests
Risk Assesment 4 year
Risk %
CHD
MI
CHD DEATH
STROKE
CVD
CVD DEATH
CHD : Global Risk Score
Age Score
Total Cholestrol Score
Hdl Score
Sytolic Bp Score
Diabetes Score
Smoker Score
Total Score
CHD : Relative & Absolute Risk
Relative Risk
below-average
Absolute Risk
Risk Assesment 10 year

has_chest_pain_type patient_1

is_chest_pain_type patient_1

Data property assertions

1 has_forename "Tim"

has_gender "Male"

2 has_family_history "Diabetes
High BP"

3 has_surname "James"

4 has_Temporal_Unit_and_has_Age_Value "75"

has_previous_cardiovascular_history "High Blood Pressure"

5 has_GP_outcome "Referral to RACPC "

has_RACPC_Nurse_outcome "Abnormal ECG suspected of CAD"

2 has_presence_of "High Blood Pressure"

5 has_final_outcome "that cosist of some Coronary angiography
that has temporal range some (month that has value some int >=7)))"

2 has_absence_of "Myocardial Infarction
Hypercholesterol
Smoking"

4 has_cardiologist_outcome "Angina
Suitable for Coronary angiography treatment"

4 has_pain_type "Typical chest pain"

Semantic Patient Profile generated in OWL

Why Ontologies ?

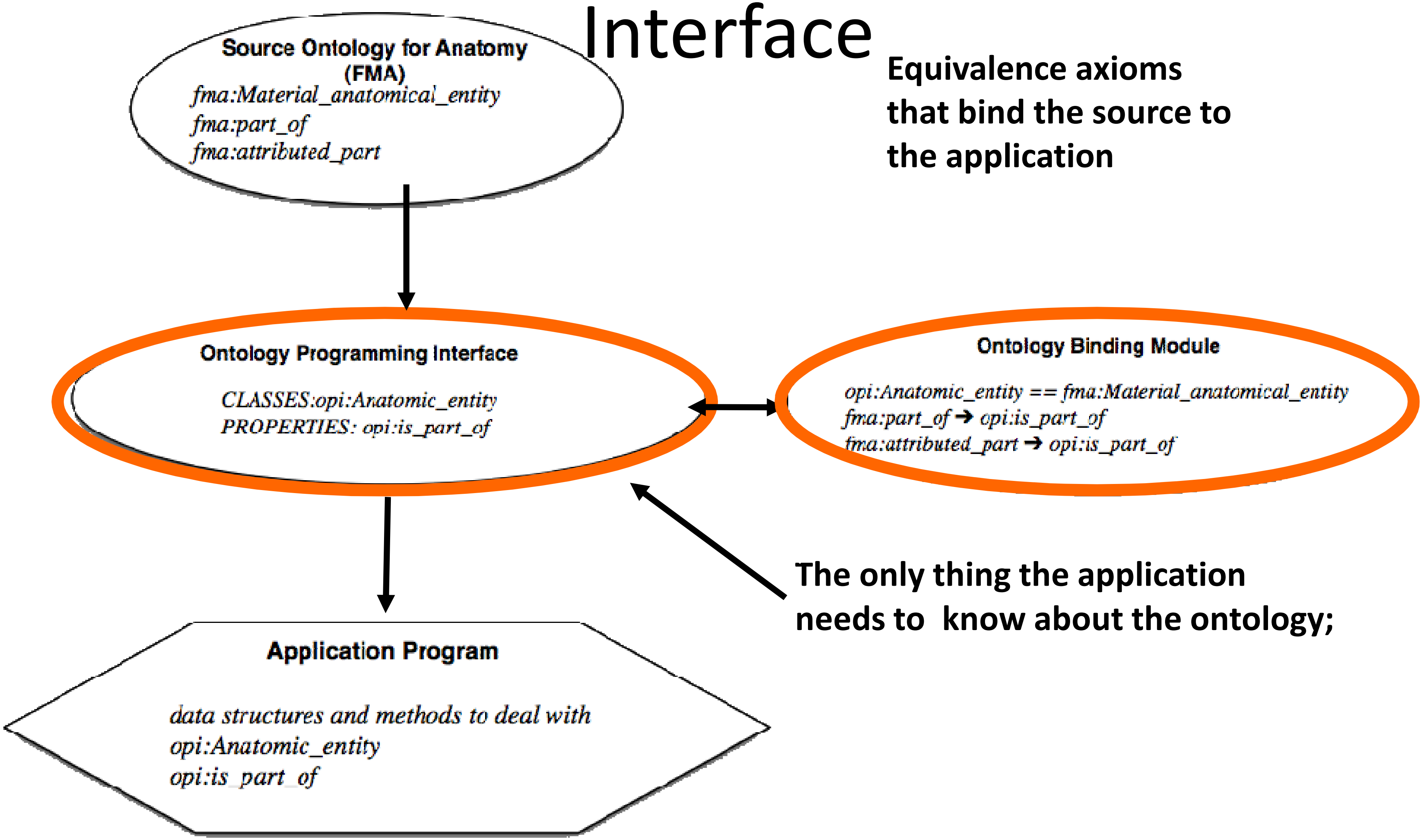
❖ Ontologies are used to capture knowledge about some domain of interest.

❖ An ontology describes the concepts (through **Classes** in the domain and also the relationships (**by defining Object and Data properties**) that hold between those concepts.

❖ *OWL DL* supports those users who want the maximum expressiveness without losing computational completeness (all entailments are guaranteed to be computed) and decidability (all computations will finish in finite time) of reasoning systems.

❖ Different ontology languages provide different facilities. The most recent development in standard ontology languages is OWL from the World Wide Web Consortium (W3C).

Fitting with Software Engineering Practice: e.g. Ontology Programming Interface

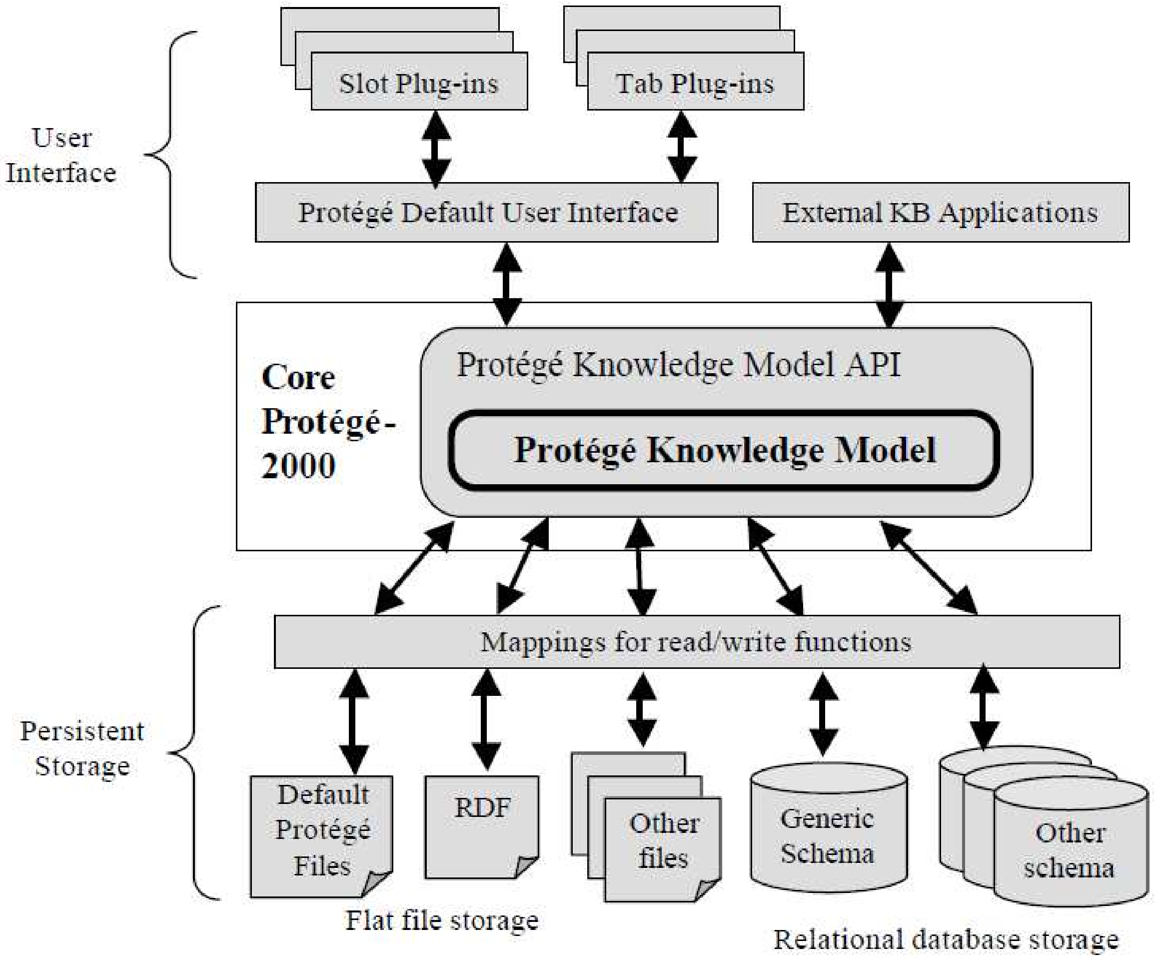


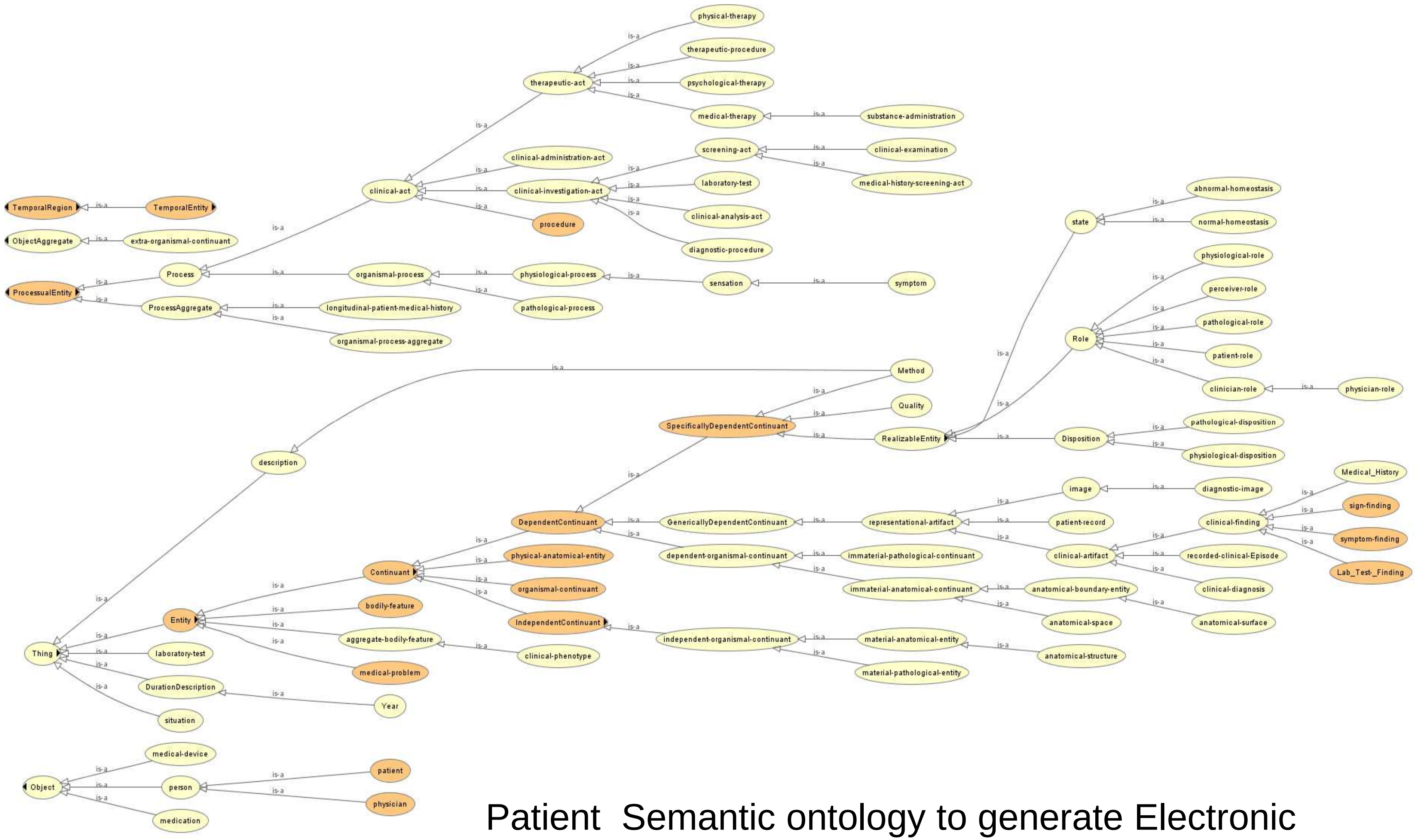
Protege

Built by Mark Musen (Stanford Center for Biomedical Research) in 1987

Protégé is neither an expert system itself nor a program that builds expert systems directly; instead, Protégé is a tool that helps users build *other tools that are custom tailored to assist with knowledge* acquisition for expert systems in specific application areas. (Musen, 1989a, p. 2)

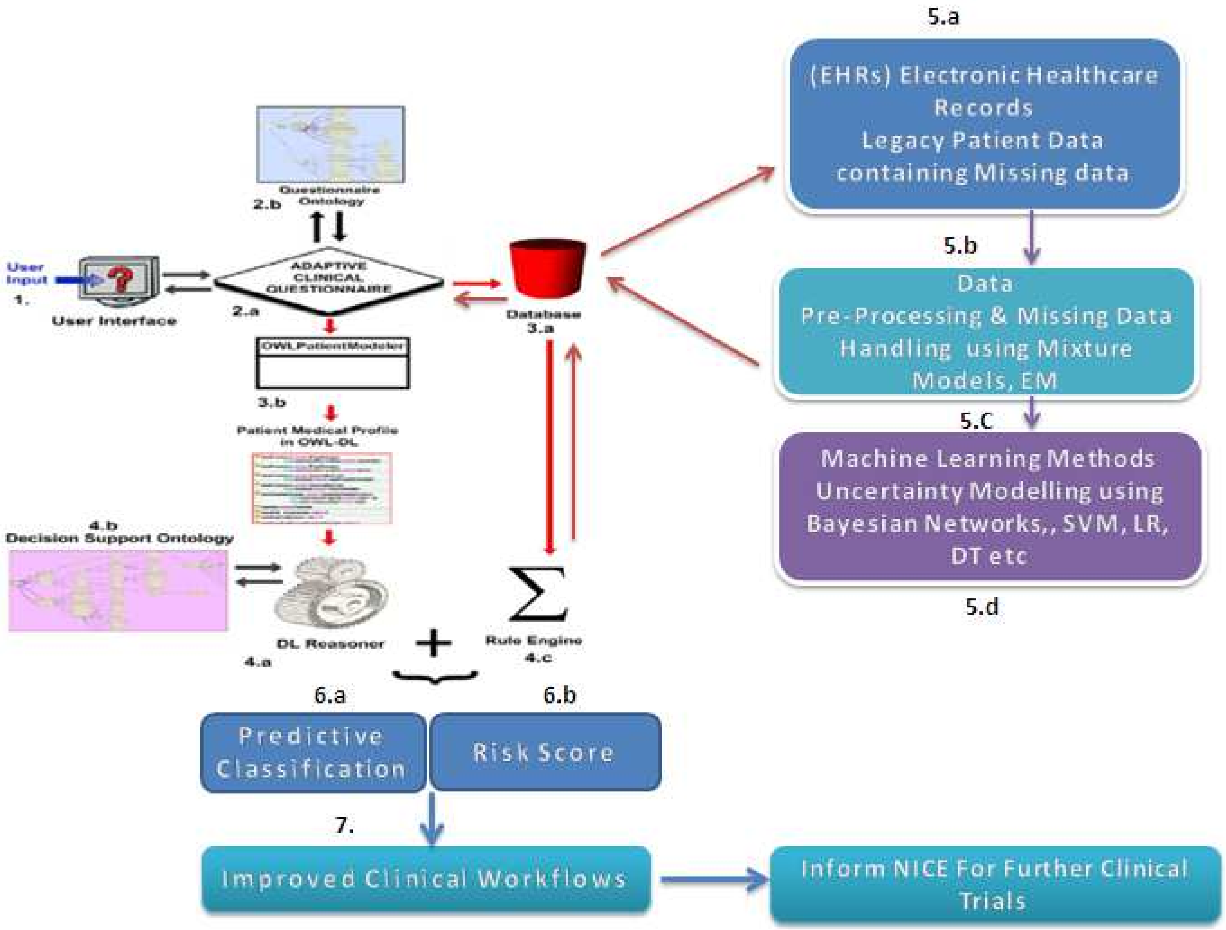
The Protégé- Architecture.





Patient Semantic ontology to generate Electronic Records

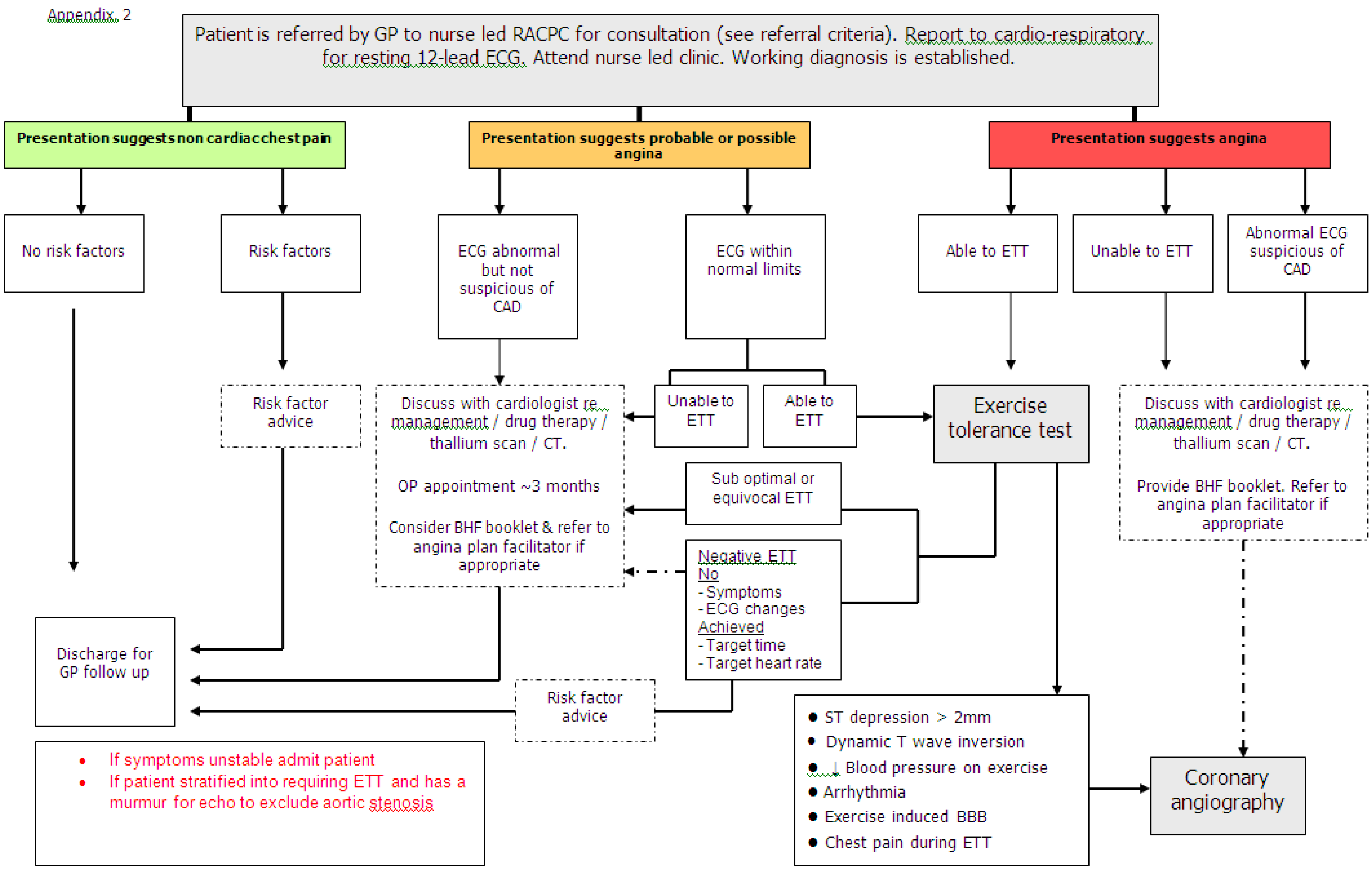
Learning from Retrospective Clinical data for making Effective Prospective Clinical Decision Support System



2. Learning from missing clinical data

- Raigmore Hospital RACPC case study
- 634 patients, 17 Binary variables containing missing data
- Expectation- Maximisation
- Mixture probabilistic model

Rapid Access Chest Pain Clinic Clinical workflow



Sample of Missing Data extracted from RACPC Dataset

not_coror	ATTENDED Y/N	NON CARDIAC Y/N	ETT Y/N	ANGINA Y/N	**FOR ANGIO Y/N**	PERFUSIO	CT	ADMIT
No	1	0	0	0	0	0	0	0
No	1	0	0	1	0	0	1	0
No	1	0	0		1	0	0	0
No	1	0	0	1	1	0	0	0
No	1	1	0	0	0	1	0	0
No	1		0	1	0	0	1	0
No	1	1	0	0	0	0	0	0
No	1	0	0		1	0	0	0
No	1	0	0	1	1	0	0	1
No	1		0	1	0	0		
No	1	0	0		1	0		
Yes	1	0	0	1	0	0		
No	1	0		1	0	0	1	0
No	1	0	0		1	0	0	0
No	1		0	1	1	0	0	0
No	1	0	0	1	1	0		0
No	1		0	1	1	0	0	0
No	1	1	0	0	0	0	0	0
No	1	0		1	1	0	0	0
No	1	0	0	1	1	0	0	0
No	1	0	0		0	1	0	0
Yes	1	1	0	0	0	0	0	0
No	1	1	1	0	0	0	1	0

Key Components of the Statistical Mixture Modeling and EM based Approach

1

joint distribution of the variables using the training samples.

Regarded the class label as categorical feature

2

Test Sample = ***worked out*** *likelihood to estimate the missing values.*

3

17 Binary Variables Model containing a mixture of 17 ***Bernoulli variables and a categorical variable.***

Key Components of the Statistical Mixture Modeling and EM based Approach

4

Given any sample X, the likelihood P (x,y) for each class y = 1,2,,,,5 is calculated and then the sample assigned with the label corresponding to the maximum likelihood.

5

Expectation Maximisation Technique to maximise the likelihood by iterating E and M steps.

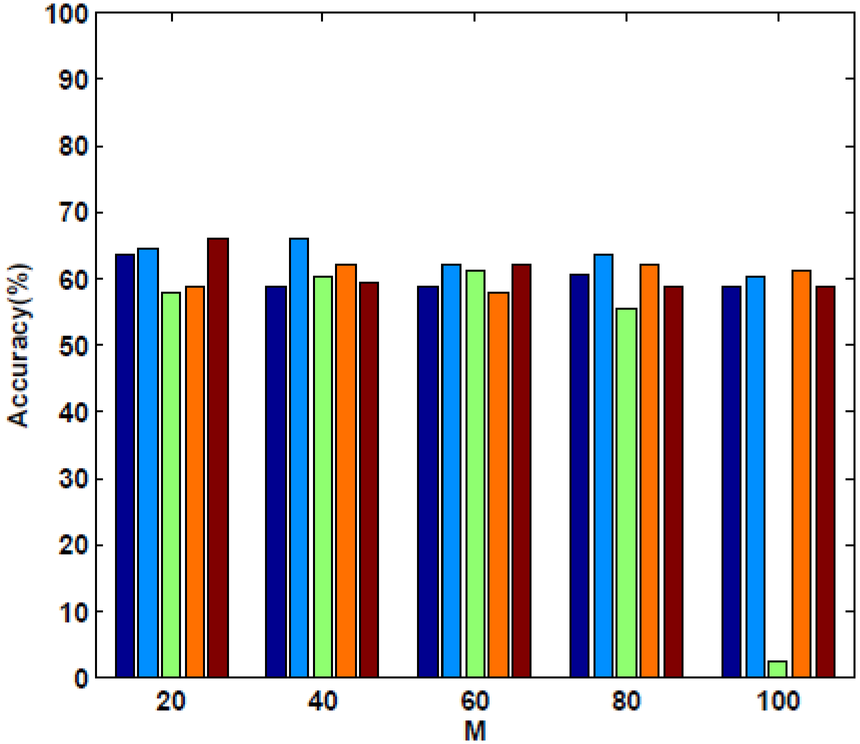
6

Data Classification using SVM Kernel Functions

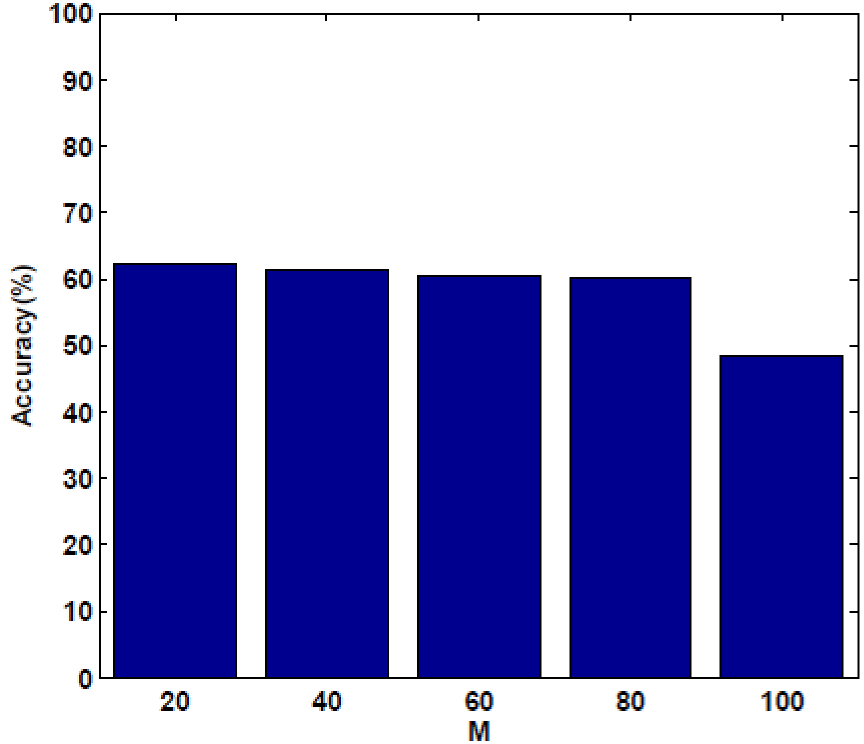
Linear, Polynomial, RBF and Sigmoid Functions

Learning from missing clinical data

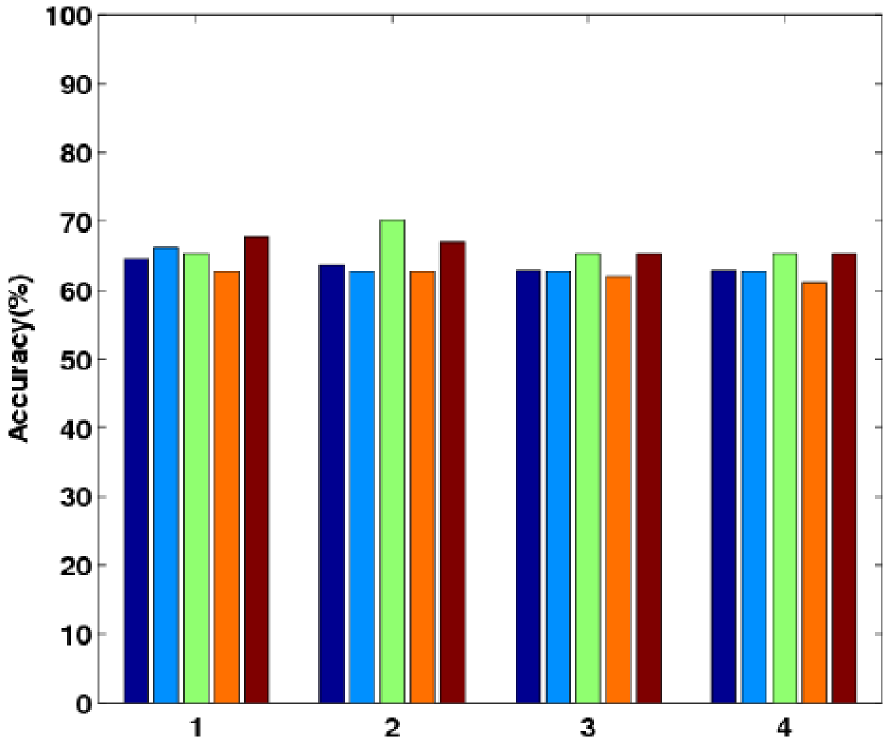
M



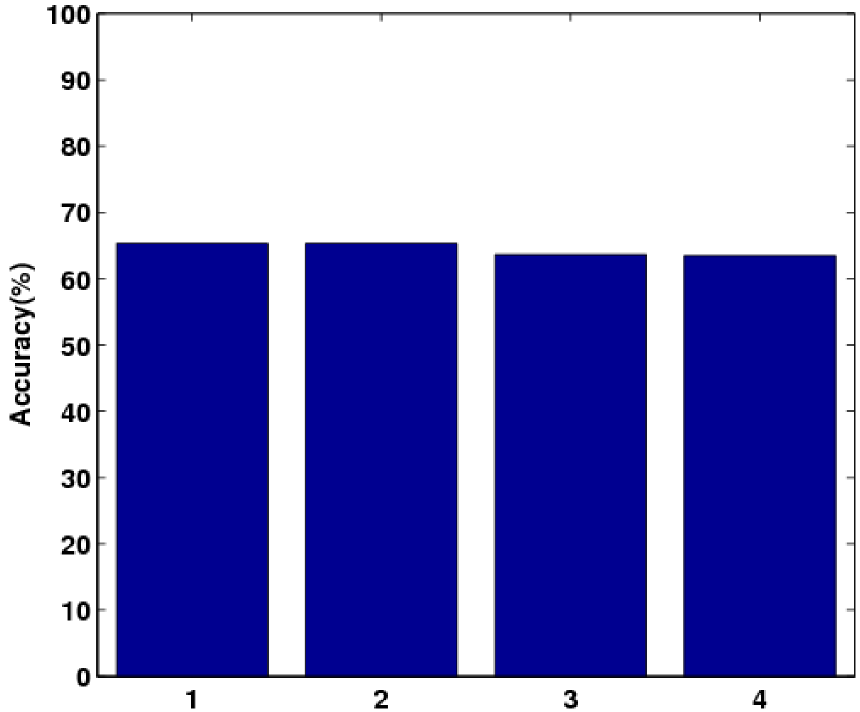
5 randomly selected datasets in which 4 datasets were used for training and 1 for testing (for each M)



Experimental results showing average accuracies of different number of mixture density models

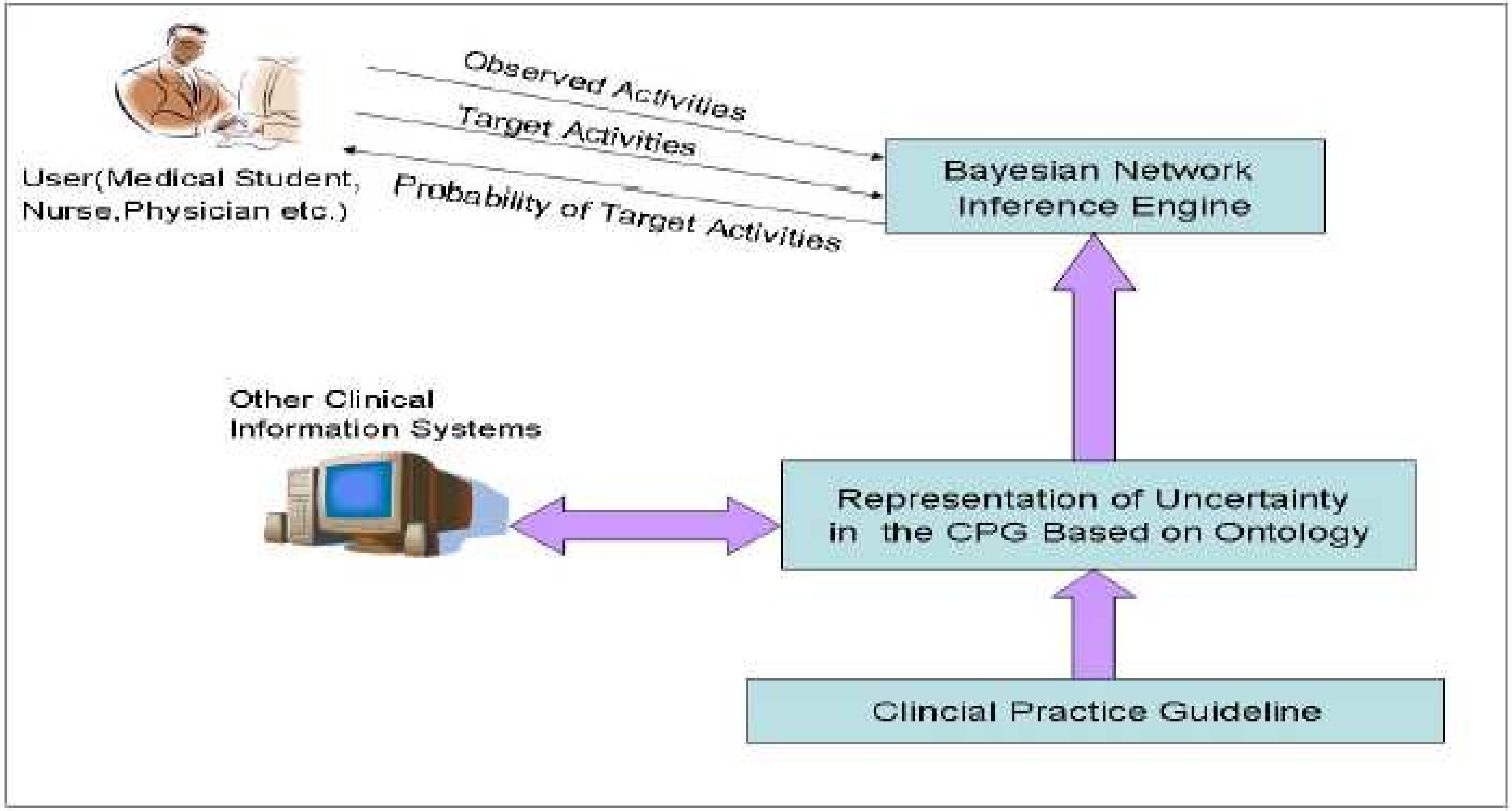


Accuracies obtained using 5 randomly selected datasets in which 4 datasets were used for training and 1 for testing for each different type of kernel function



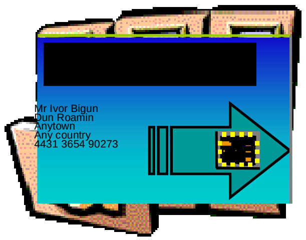
Experimental results showing average accuracies of different types of kernel functions including: 1- Linear, 2- Polynomial, 3- Radial Basis Function and 4- Sigmoid Function

3. Bayesian Network for Uncertainty Modelling in Clinical Guidelines



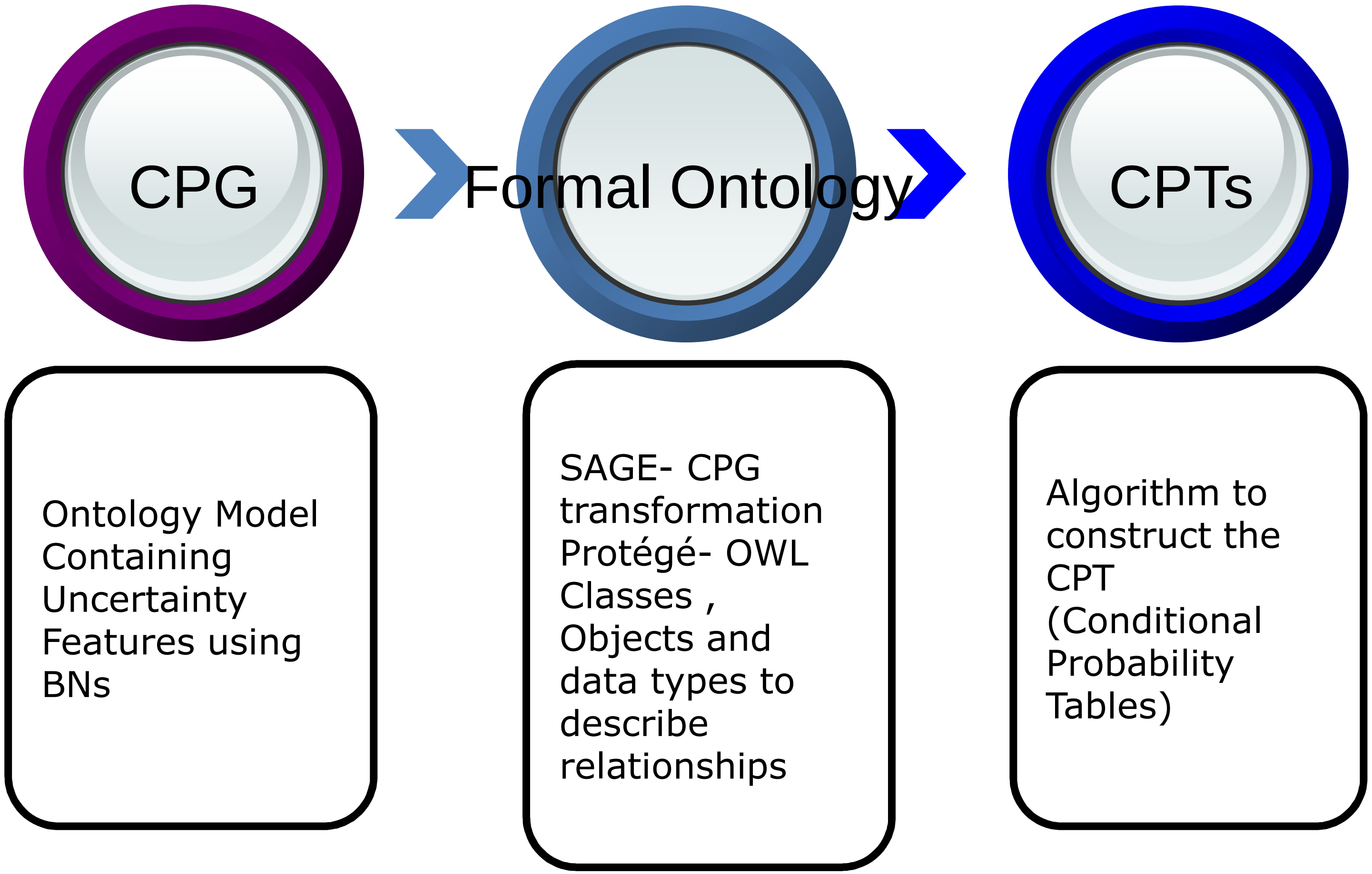
Why Clinical Practice Guidelines are essential ?

Clinical Guidelines ← Data Entry ← Clinical Record ← Clinical research & Decision Support ← Best Practice



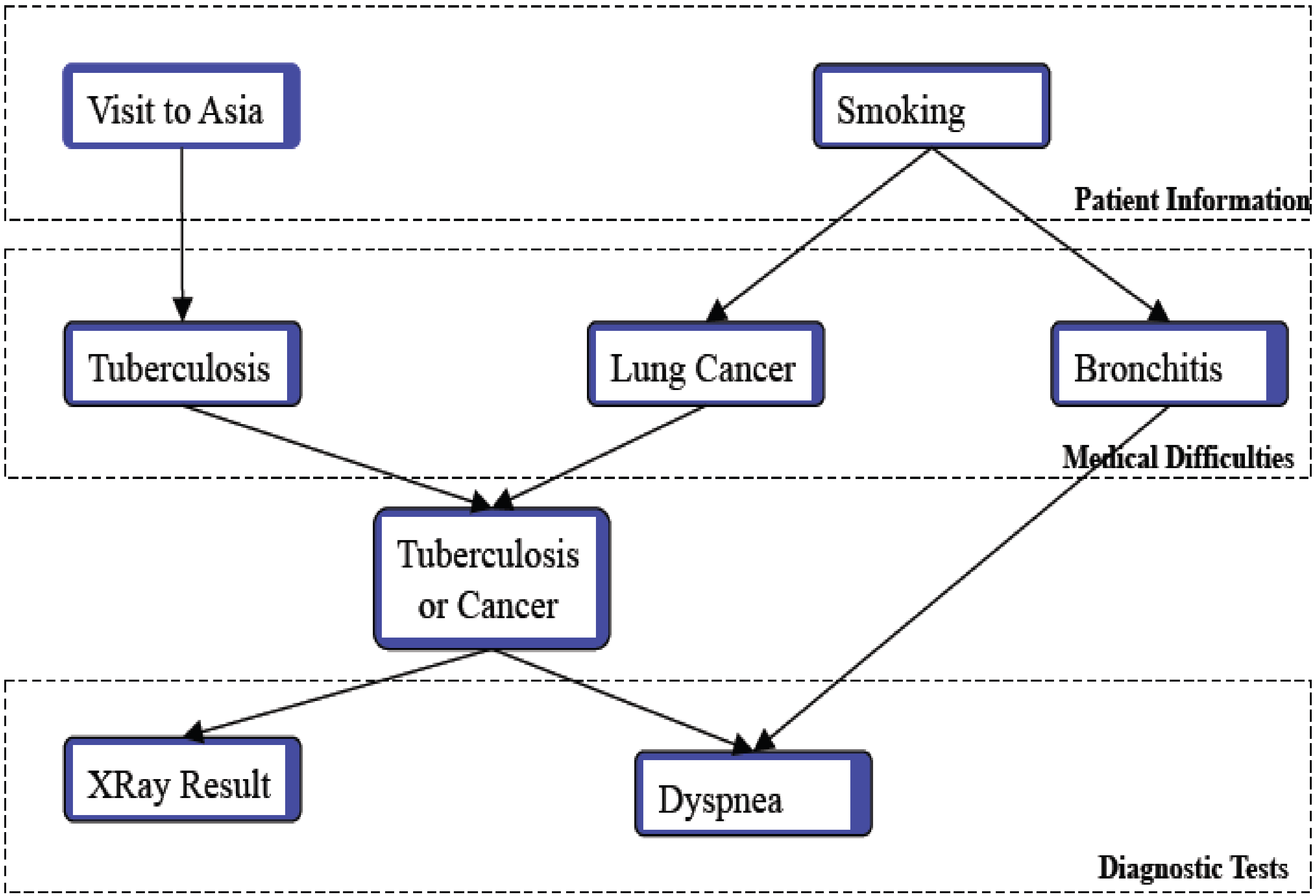
GALEN
Ontologies & Description logics,
SNOMED CT
→ Data Entry → Electronic Health Records → Clinical research Decision Support & Knowledge Presentation → **Best Practice**

Bayesian Network for Uncertainty Modelling in CPGs

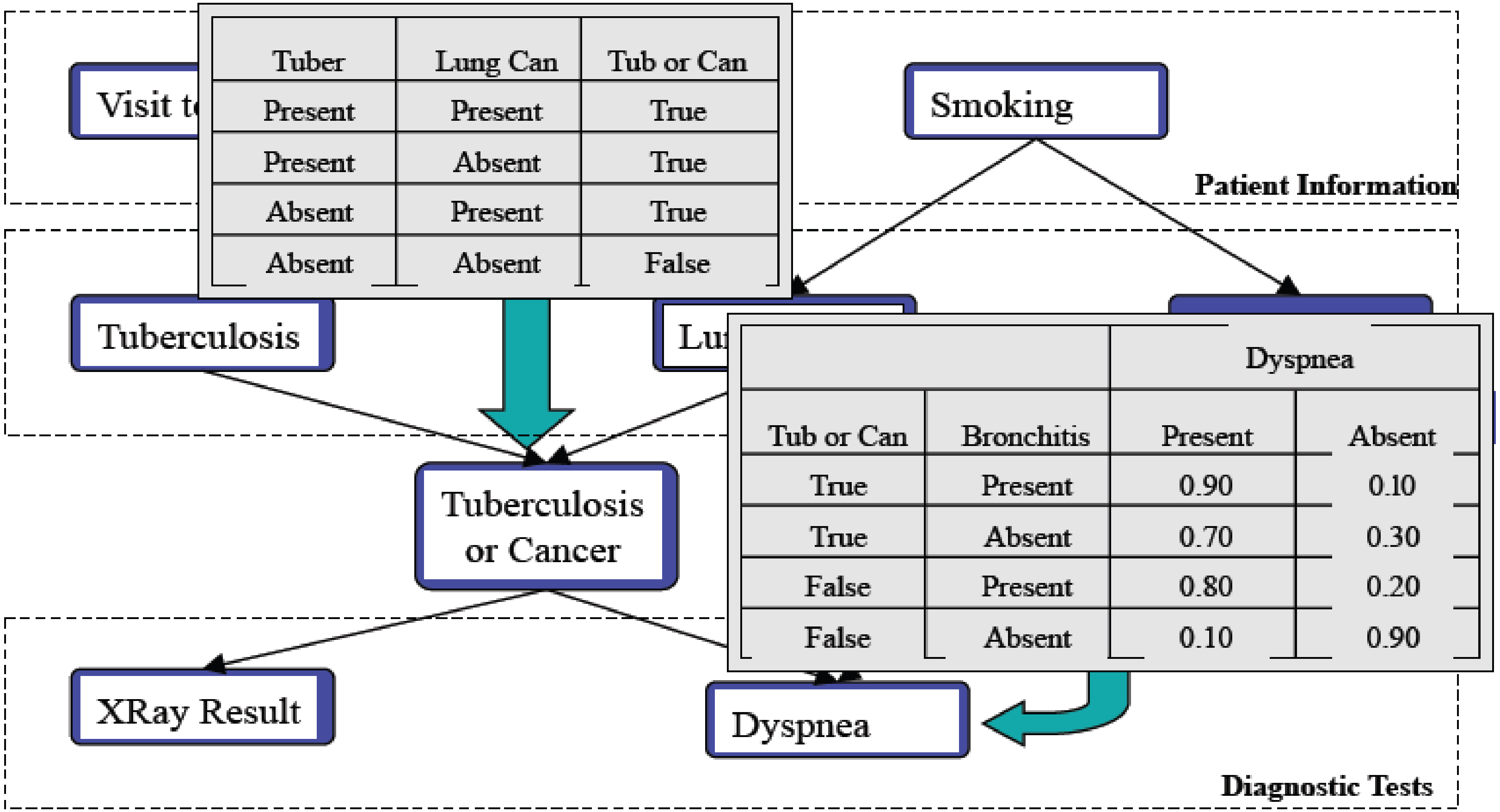


Development Stages

Example from Medical Diagnostics



Example from Medical Diagnostics

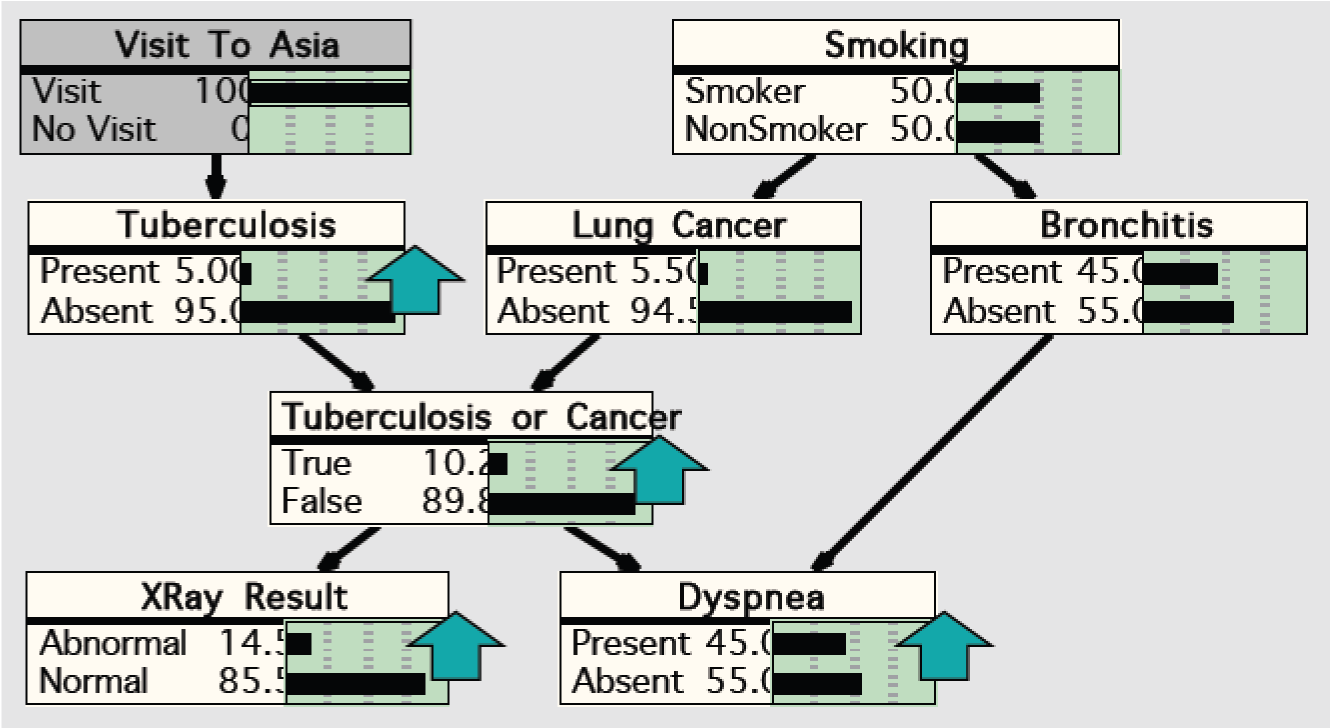


Tuber	Lung Can	Tub or Can
Present	Present	True
Present	Absent	True
Absent	Present	True
Absent	Absent	False

		Dyspnea	
Tub or Can	Bronchitis	Present	Absent
True	Present	0.90	0.10
True	Absent	0.70	0.30
False	Present	0.80	0.20
False	Absent	0.10	0.90

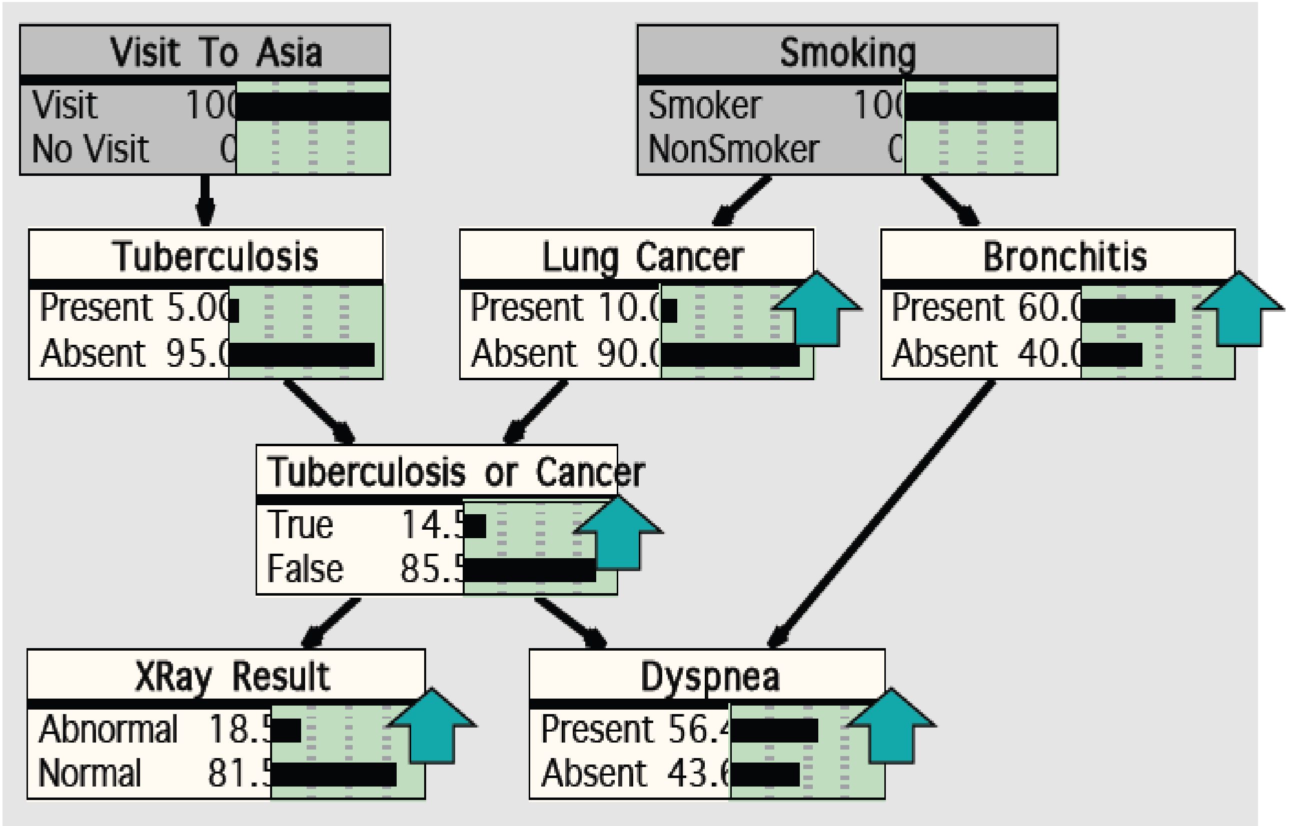
- Relationship knowledge is modeled by deterministic functions, logic and conditional probability distributions

Example from Medical Diagnostics

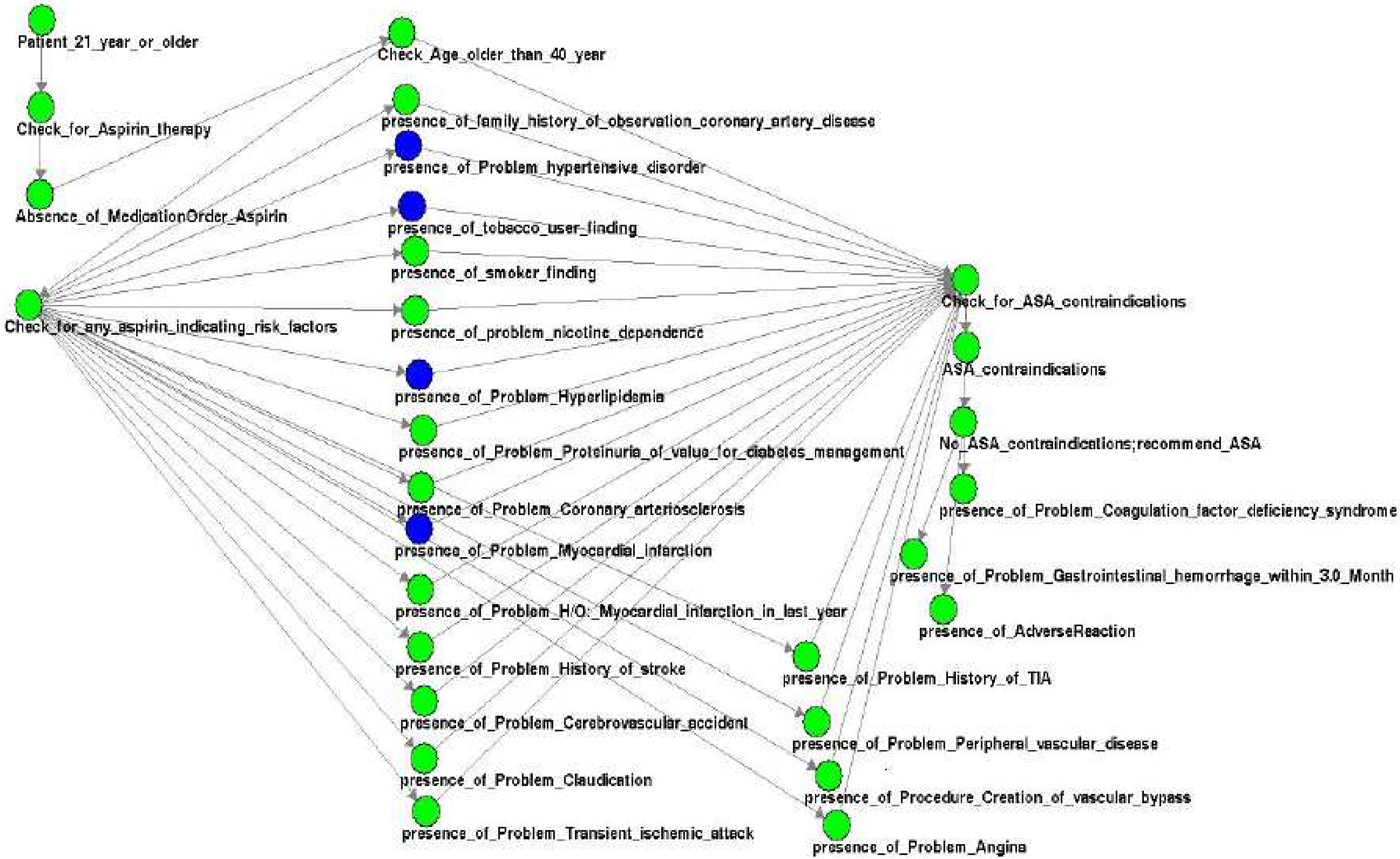


- As a finding is entered, the propagation algorithm updates the beliefs attached to each relevant node in the network
- Interviewing the patient produces the information that "Visit to Asia" is "Visit"
- This finding propagates through the network and the belief functions of several nodes are updated

Example from Medical Diagnostics



- Further interviewing of the patient produces the finding “Smoking” is “Smoker”
- This information propagates through the network



Ontology based Bayesian Network to check patient’s suitability for Angiography treatment

4. Predictive Analysis

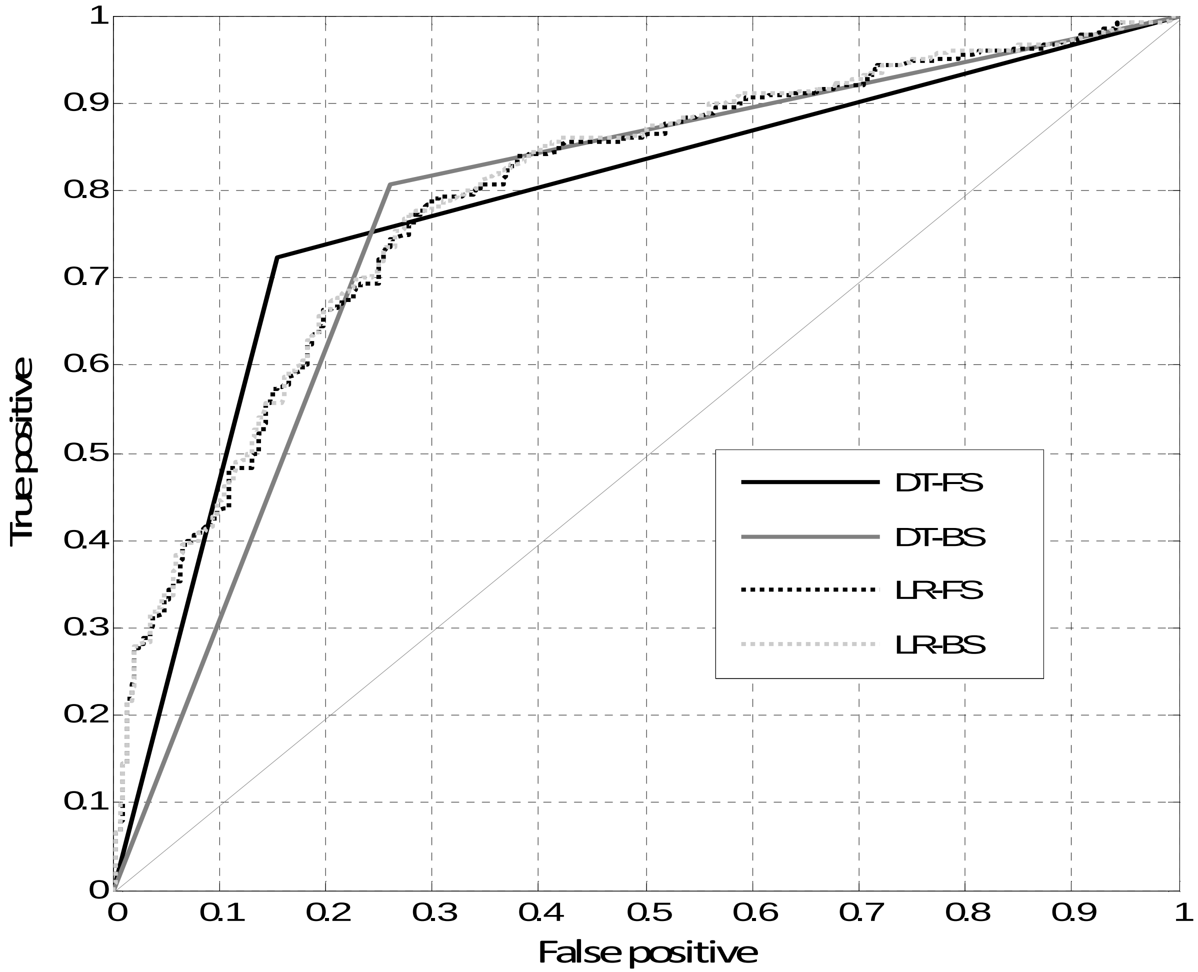
Clinical Predictors for the Cardiac/ Chest Pain Risk Assessment

Features/Risk Factors		**Targets/Final Diagnosis**	
	Acronym		**Number of Patients**
Smoker	SMR	Acute Coronary Syndrome	9
Number of Cigarettes	NOC	Angina	274
Number of Years Smoked	YOS	Arrhythmia	11
Age	AGE	Declined Investigation	4
Pathway	PWY	GI Pain	39
Sex	SEX	Heart Failure	2
Diabetes Type	DAB	Syndrome X	5
Hypertension	HPT	Valve disease	3
Raised Cholesterol	CHL	Myocarditis	1
Initial Assessment	INA	Non Cardiac Symptoms	284
ETT Result	ETT		
CT Result	CTT	**Total patients**	**632**
MPS Result	MPS		
Anigo Result	ANG		

Classification results in terms of several evaluations

Method	Unweighted accuracy	Weighted accuracy	Precision	Recall	F-measure	Matthew's Correlation
DT-FS	77.8481	78.4604	72.4138	85.1351	78.2609	0.5674
DT-BS	77.6899	77.3454	80.7471	79.1549	79.9431	0.5483
LR-FS	74.6835	74.4212	77.0115	77.0115	77.0115	0.4884
LR-FS	74.6835	74.4536	76.7241	77.1676	76.9452	0.4888

ROC curves for different experimental setups



Weighted classification accuracies in each iteration

Iteration	FS-DT		BS-DT		FS-LR		BS-LR	
1	**ANG**	64.7867	MPS	76.0240	**INA**	66.0596	ETT	74.3423
2	**INA**	71.7298	NOC	76.5198	AGE	67.8100	CHL	74.2776
3	CT	77.3454	CHL	76.8395	**ANG**	71.9423	DAB	74.4212
4	ETT	78.4341	SMR	77.1127	SEX	72.6789	NOC	74.4536
5	DAB	78.4341	ETT	77.1592	MPS	73.3831	MPS	73.8931
6	SEX	78.4604	DAB	76.8719	YOS	74.0550	SMR	73.3042
7	HPT	77.5943	YOS	73.6421	NOC	73.9113	HPT	73.8141
8	CHL	76.9650	AGE	75.0000	HPT	73.9902	YOS	73.6705
9	MPS	74.2492	PWY	77.3069	PWY	74.3099	CT	72.7113
10	NOC	73.9619	SEX	76.6270	ETT	74.3099	PWY	72.6789
11	PWY	76.3761	HPT	77.3454	CT	74.3099	SEX	71.9423
12	SMR	75.3379	CT	71.7298	SMR	74.4212	**INA**	68.1743
13	AGE	75.1153	**INA**	64.7867	DAB	74.1339	**ANG**	62.0690
14	YOS	75.1153	**ANG**		CHL	74.1663		

Risk factors and two classes (weighted)

22	FS+SVM RBF	70.153
23	BS+ SVM RBF	69.7133
24	SFFS+ SVM RBF	70.0717
25	MRMR+ SVM RBF	64.8746, 66.8459, 69.7133, 67.3835, 68.4588, 69.5341, 68.8172
26	FQ+ SVM RBF	64.8746, 66.8459, 69.7133, 68.9964, 68.4588, 69.5341, 68.8172
27	Pval+ SVM RBF	
28	ALL+ SVM RBF	68.4588
29	FS+knn (3)	63.6201
30	BS+ knn (3)	65.233
31	SFFS+ knn (3)	65.233
32	MRMR+ knn (3)	56.6308, 60.2151, 56.4516, 58.9606, 63.9785, 63.9785, 61.6487,
33	FQ+ knn (3)	56.6308, 60.2151, 56.4516, 63.6201, 63.9785, 63.9785, 61.6487
34	Pval+ knn (3)	
35	ALL+ knn (3)	63.0824
36	FS+SVM Lin	68.4588
37	BS+ SVM Lin	68.9964
38	SFFS+ SVM Lin	67.9211
39	MRMR+ SVM Lin	65.233, 67.3835, 67.3835, 67.7419, 67.2043, 67.2043, 66.8459
40	FQ+ SVM Lin	65.233, 67.3835, 67.3835, 67.5627, 67.2043, 67.2043, 66.8459
41	Pval+ SVM Lin	
42	ALL+ SVM Lin	66.129

Conclusions

1. The aim of our research is to help improve the diagnostic capabilities and performance of RACPC, and to eliminate delay in the cardiovascular risk assessment of patients with chest pain effectively as well as distinguishing acute angina patients with other causes of chest pain. An AI inspired prospective clinical decision support framework has been developed for the primary and secondary care clinicians.

2. Electronic Healthcare Records/ patient summary/paperless system (by replacing paper based records) for the risk assessment of patients thus providing predictive analysis using the patient semantic profile developed in the doctor-patient consultation/initial stages

3. We have demonstrated through our case studies that we can provide efficient clinical making using our developed framework. We can predict different cardiovascular diseases, calculate patients cardiac risk score in terms of developing heart disease in the next ten years, prescribing STATIN and recommending necessary lab tests etc. We demonstrated this work successfully.

4. The developed CDSS framework is also capable of dealing with missing/impartial data as well as handling clinical uncertainties . A retrospective data analysis of the clinical studies evaluating 14 risk factors for chest pain patients was performed for the development of RACPC specific risk assessment models to distinguish between cardiac and non cardiac chest pain. This study cohort comprises of 632 patients suspected of cardiac chest pain, data were electronically recorded from August 2011 to May 2013 using distributed databases held at the Raigmore Hospital . The new predictive models have resulted in very good predictive power, demonstrating general performance improvement compared to a state-of-the-art prediction model.

Other CDSS- Related Work

Dementia diagnosis DSS

Used technique: Logistic regression

- gives for each record a probability of belonging to a class (binary classifier with a threshold)
- 70% of medical publications make use of it
  - diagnosis
  - prognosis
  - analysis of contributing factors
  - risk modelling
- easily implementable
- transparent model

Dementia diagnosis DSS

- Different sources of evidence are used in the dementia diagnosis process
- Timely diagnosis of dementia is a condition for improving dementia care
- GPs have a central role in the diagnosis process, but 50-80% of cases are missed
- There is a limited range of readily available diagnostic instruments

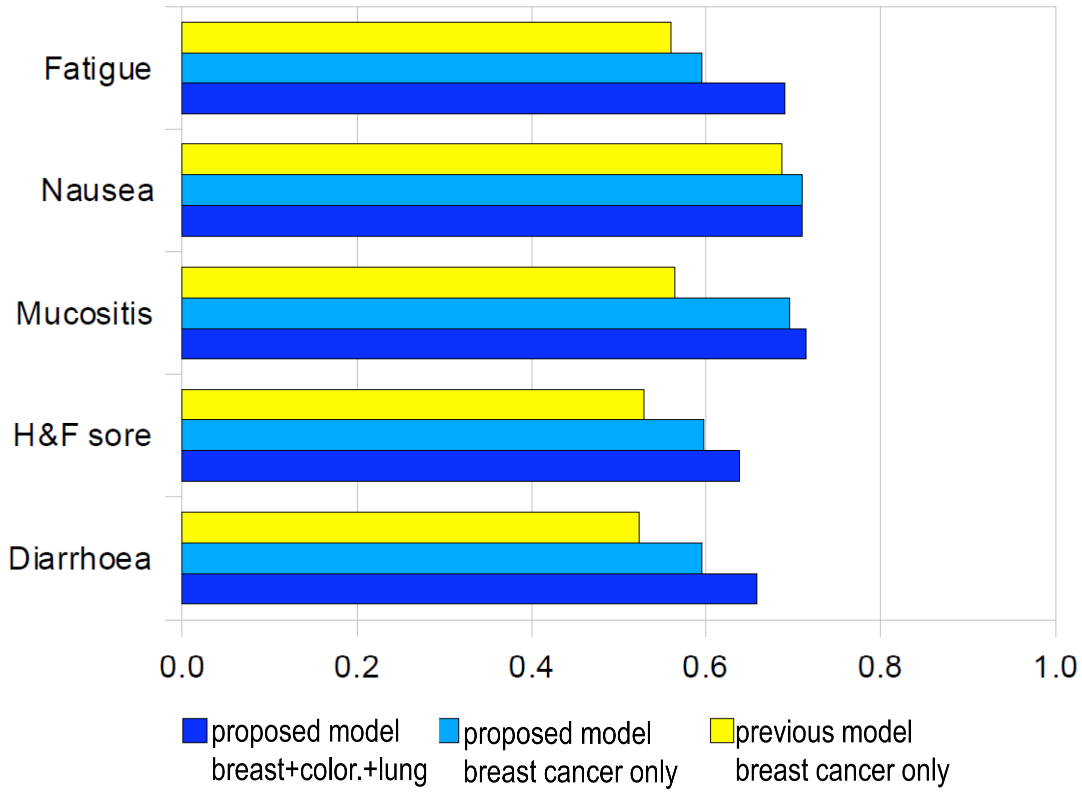
Dementia diagnosis DSS

Data-driven variables selection using an intelligent backward stepwise logistic regression approach

	Benchmark (expert-driven)	*Current model*	*Current model (w/o cross-val.)*
Accuracy	75.0%	90.2%	91.5%
AUC	0.764	0.879	0.905
R^2	-	-	0.365 ÷ 0.601

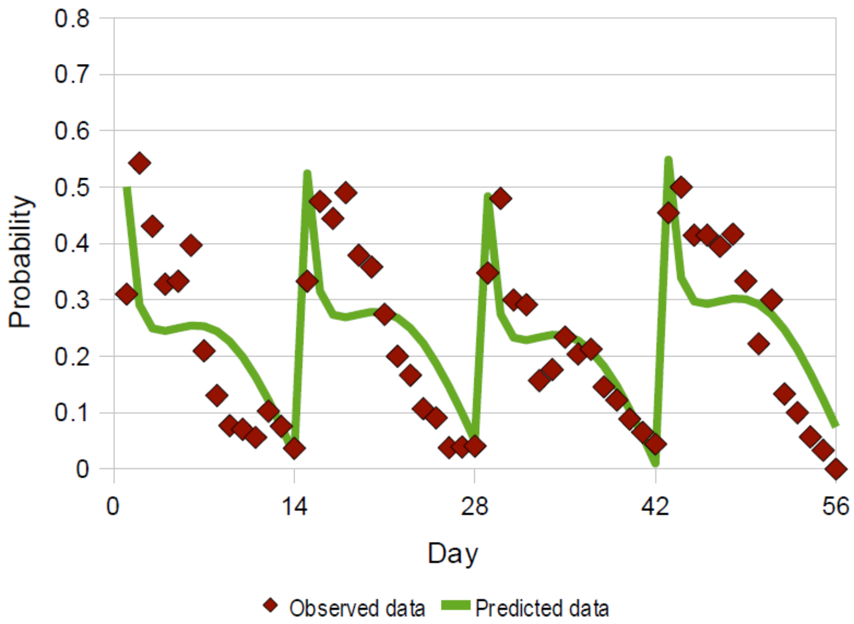
Chemotherapy side-effects modelling

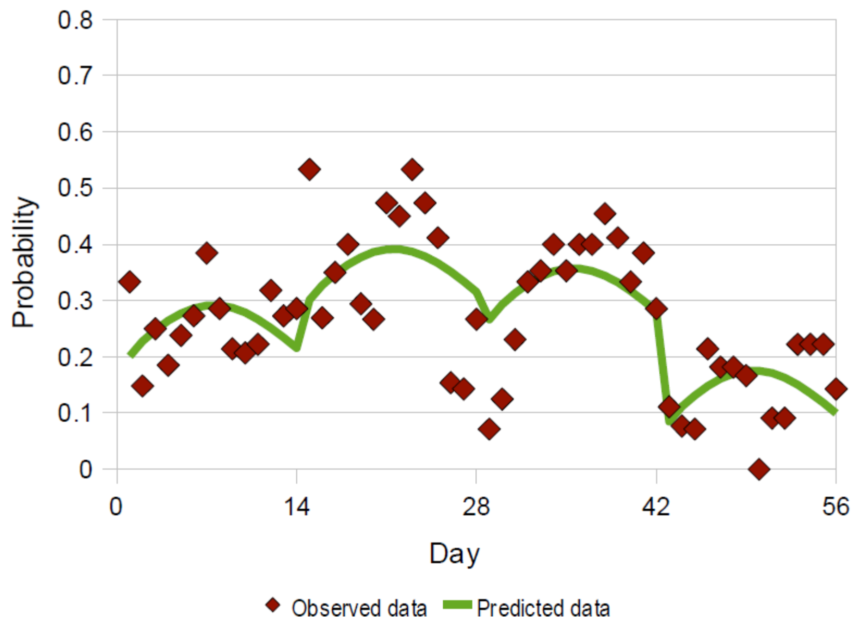
Model performance (AUC from ROC)



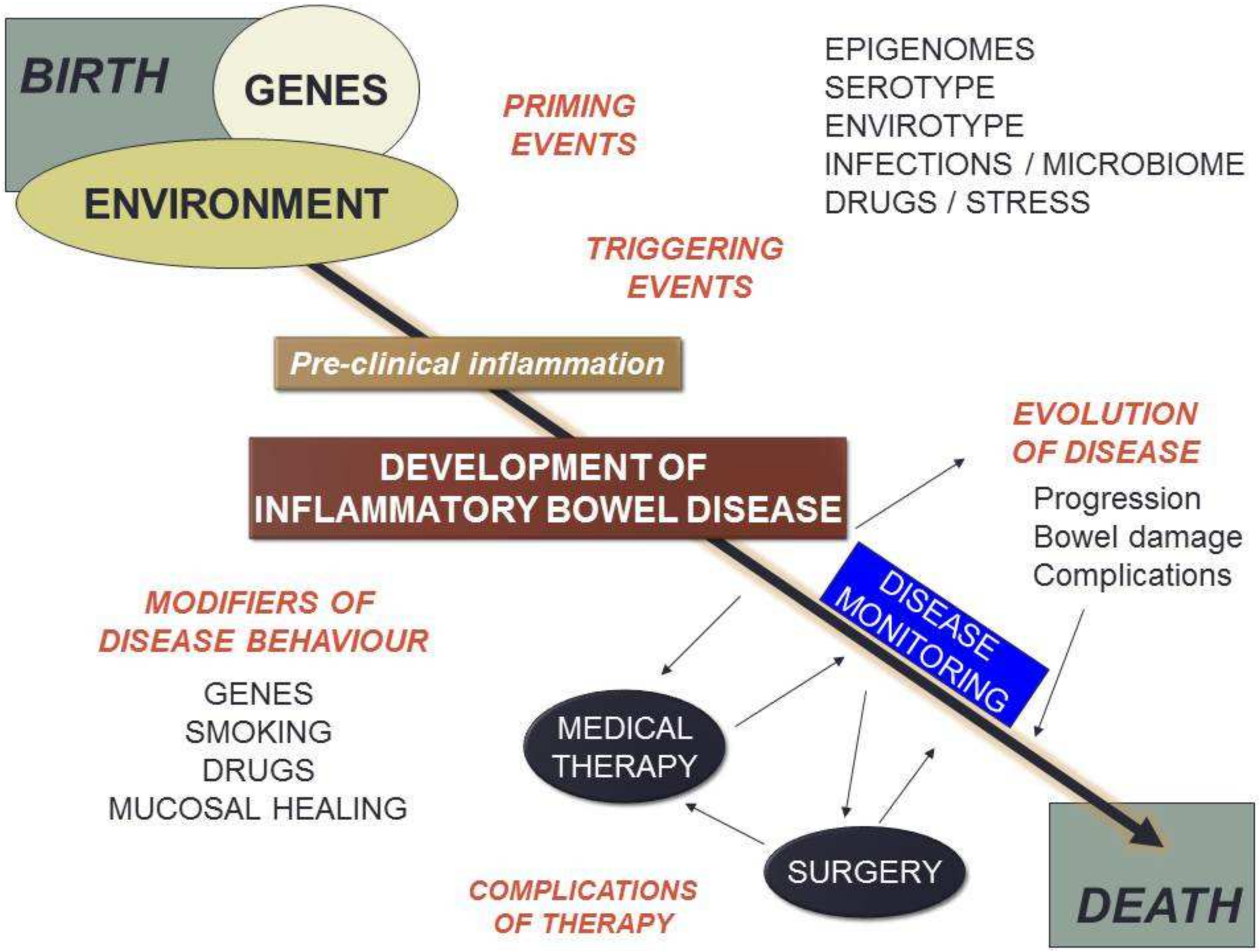
Chemotherapy side-effects modelling

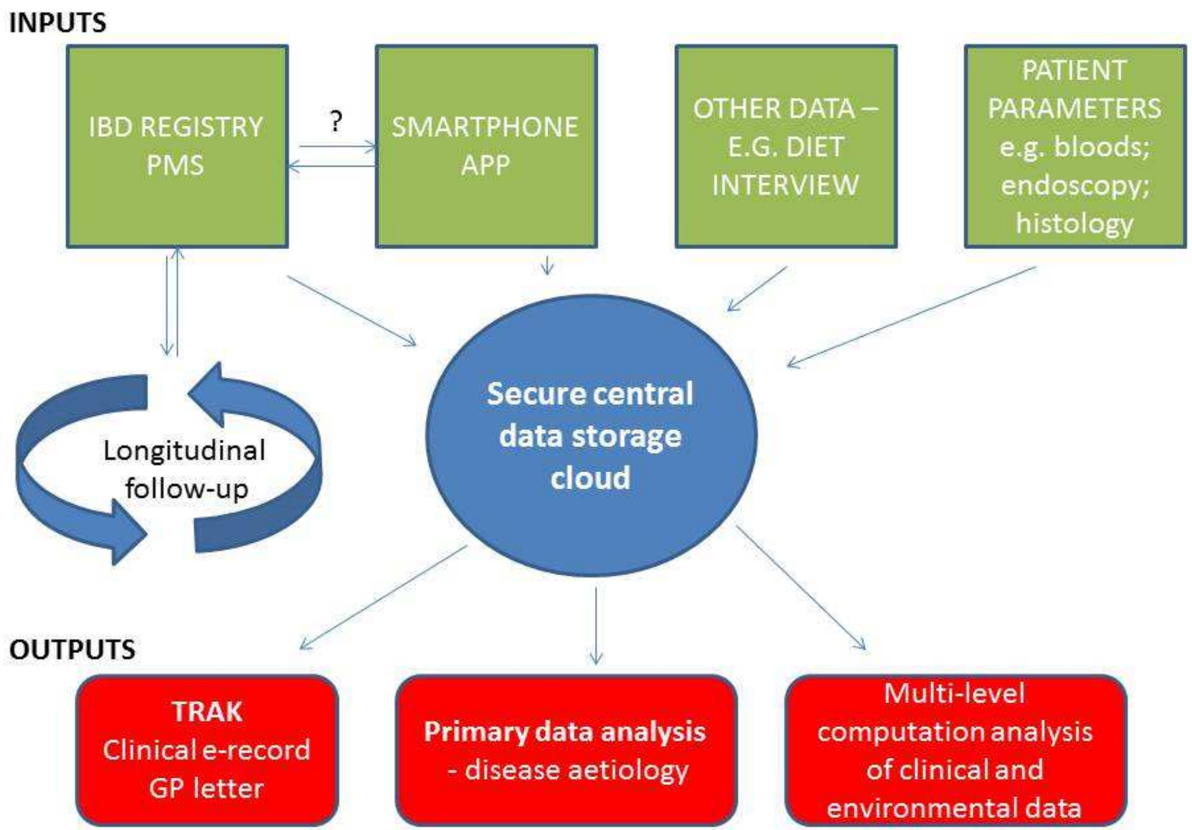
Examples of 'intelligent' mapping



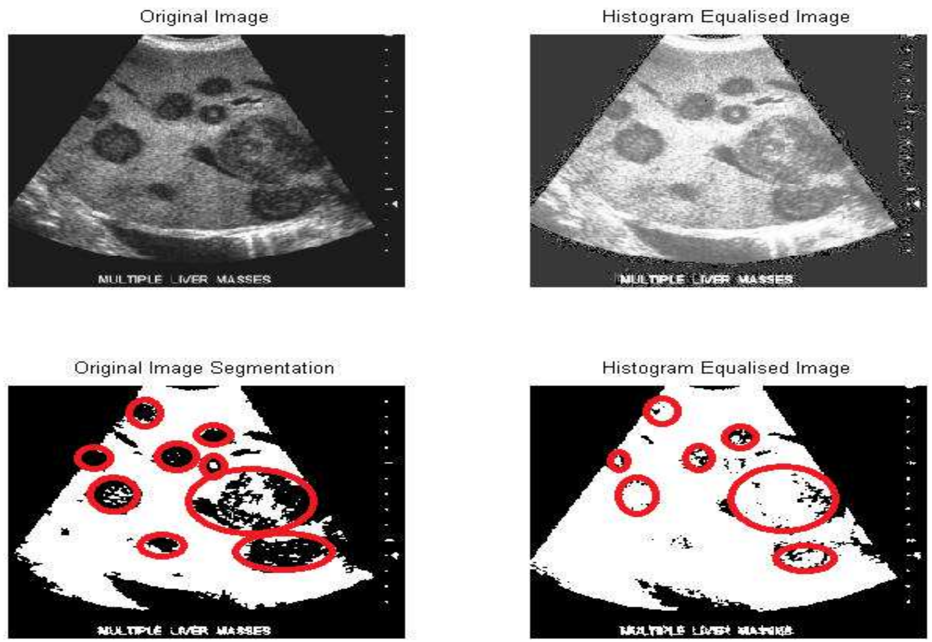


Defining the Inflammatory Bowel Disease Exposome : the effect of the environment and diet on disease aetio-pathogenesis across Scotland

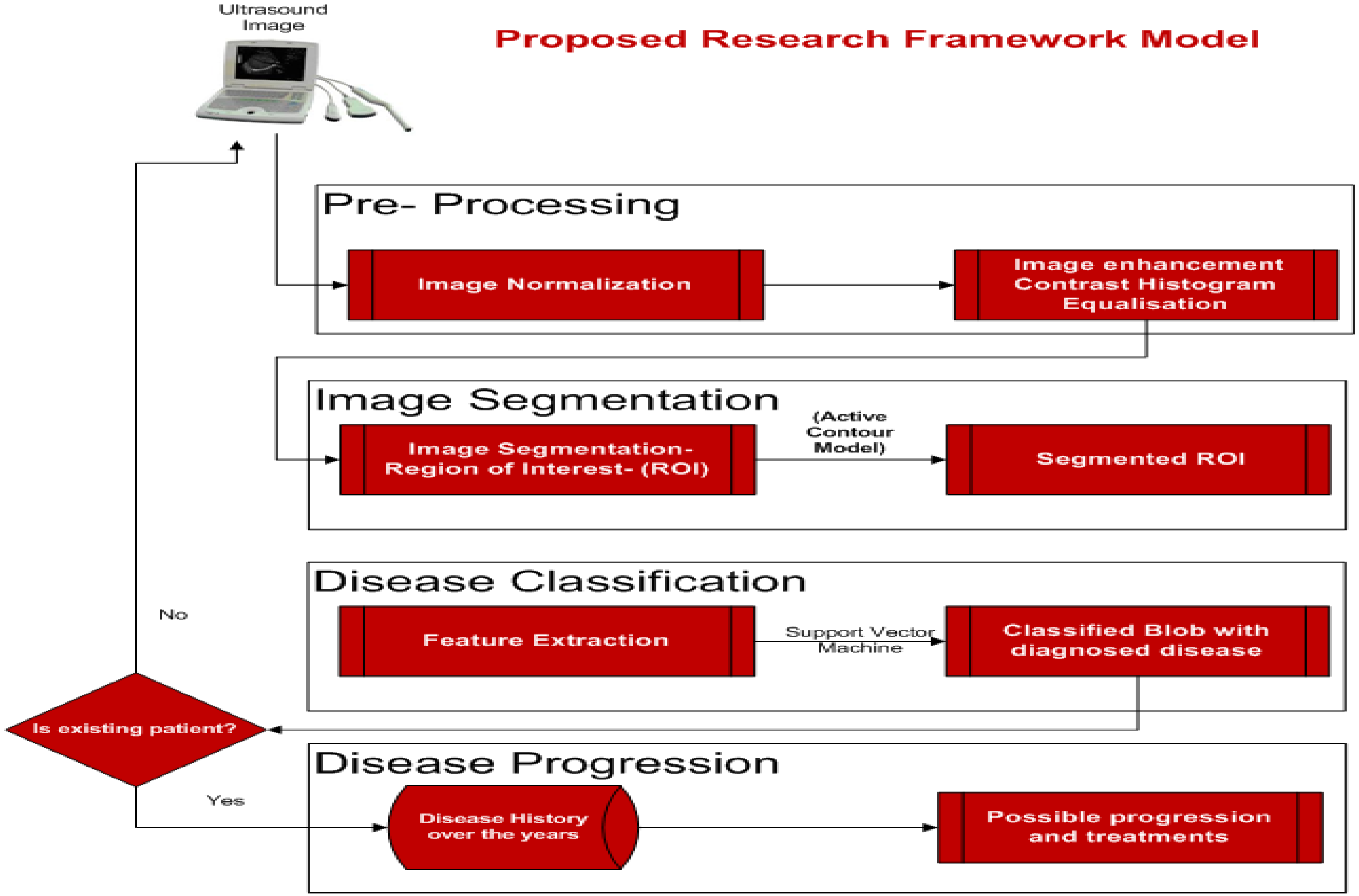



INPUTS
IBD REGISTRY PMS
?
SMARTPHONE APP
OTHER DATA – E.G. DIET INTERVIEW
PATIENT PARAMETERS e.g. bloods; endoscopy; histology
Longitudinal follow-up
Secure central data storage cloud
OUTPUTS
TRAK Clinical e-record GP letter
Primary data analysis - disease aetiology
Multi-level computation analysis of clinical and environmental data

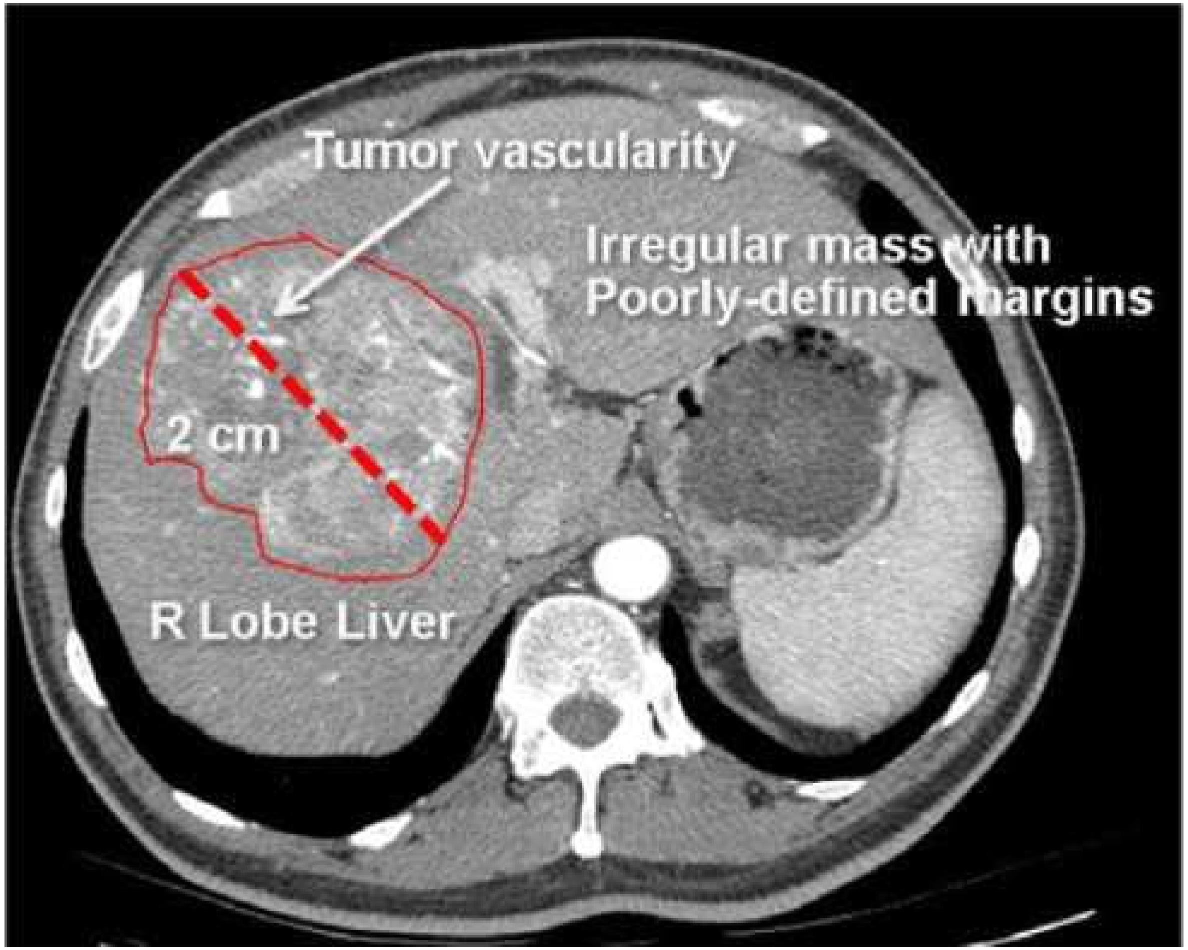
Liver Cancer Detection, Classification and Progression Prediction from Ultrasound Images

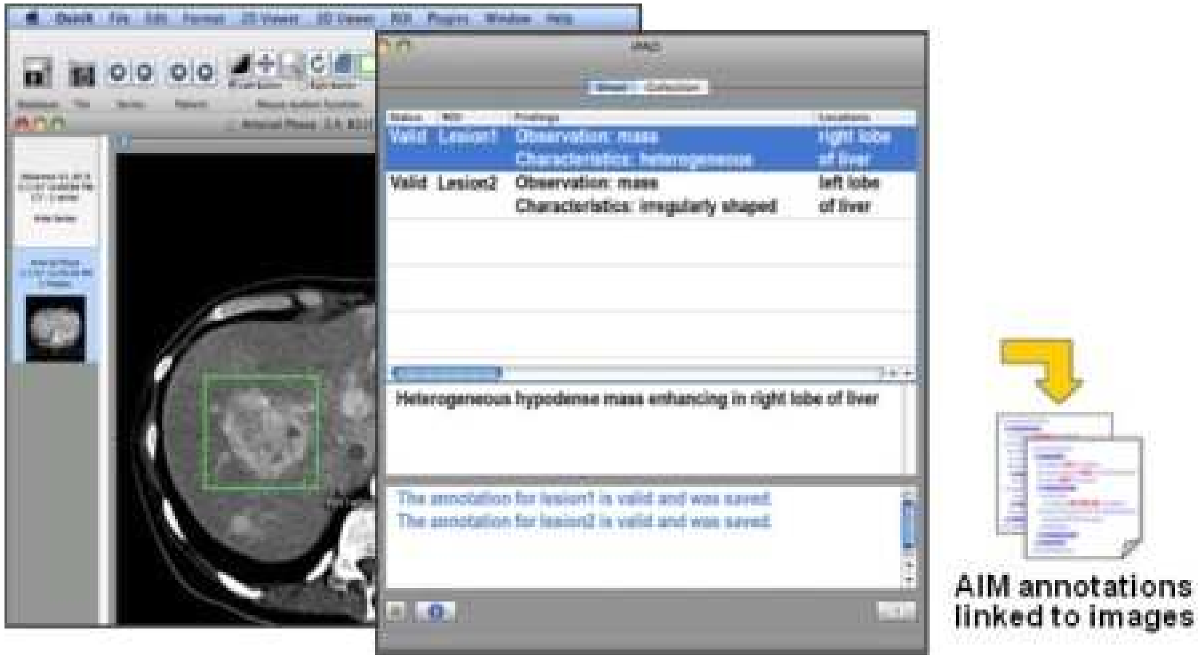


Proposed research framework model



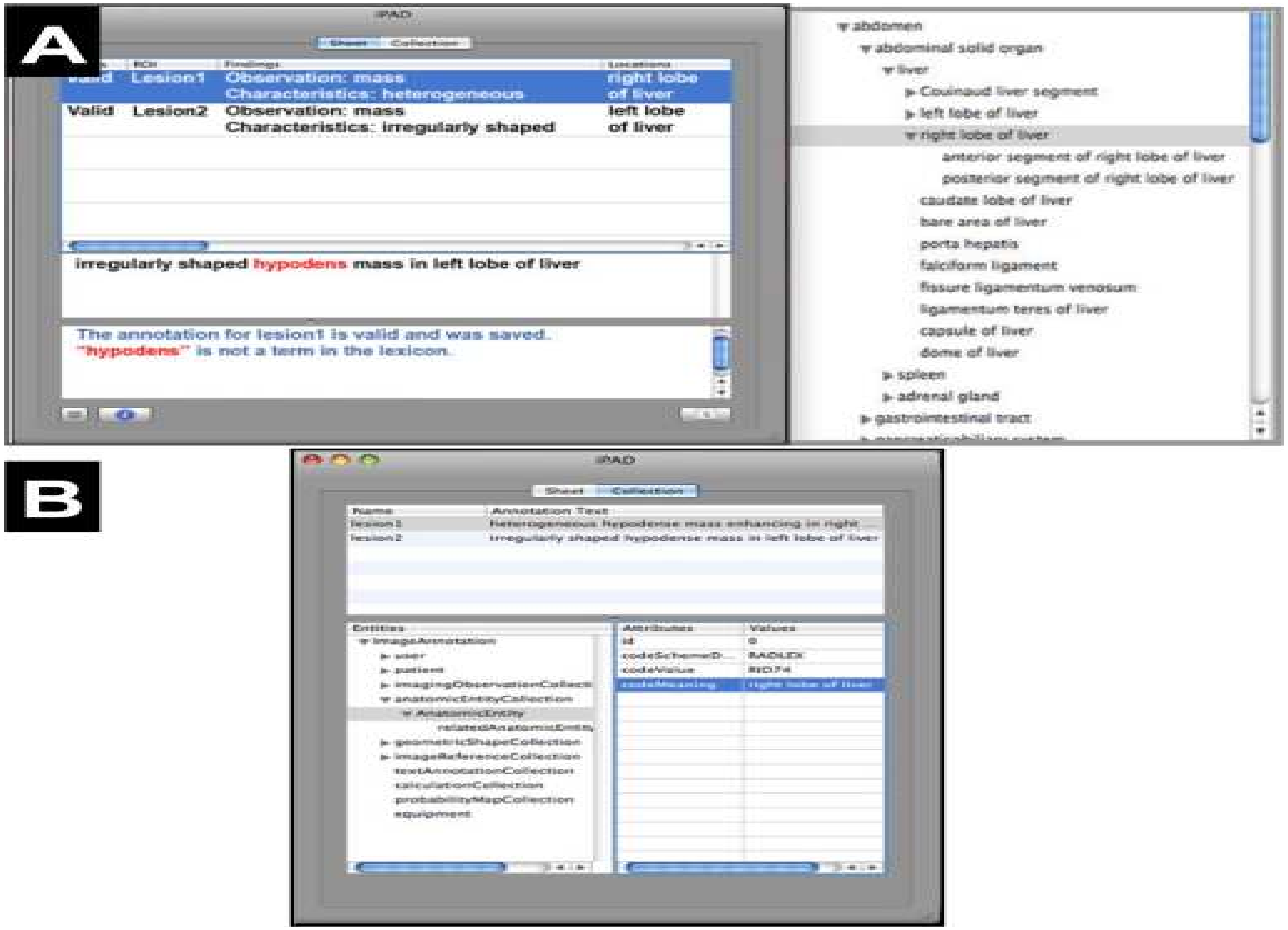
AAM and Ontology Driven Annotation





Valid Lesion1 Observation: mass Characteristics: heterogeneous right lobe of liver
Valid Lesion2 Observation: mass Characteristics: irregularly shaped left lobe of liver
Heterogeneous hypodense mass enhancing in right lobe of liver
The annotation for lesion1 is valid and was saved.
The annotation for lesion2 is valid and was saved.
AIM annotations linked to images

Image A



A
B
iPAD
Sheet
Collection
Valid Lesion1 Observation: mass Characteristics: heterogeneous right lobe of liver
Valid Lesion2 Observation: mass Characteristics: irregularly shaped left lobe of liver
irregularly shaped hypodens mass in left lobe of liver
The annotation for lesion1 is valid and was saved.
"hypodens" is not a term in the lexicon.
abdomen
abdominal solid organ
liver
Couinaud liver segment
left lobe of liver
right lobe of liver
anterior segment of right lobe of liver
posterior segment of right lobe of liver
caudate lobe of liver
bare area of liver
porta hepatis
falciform ligament
fissure ligamentum venosum
ligamentum teres of liver
capsule of liver
dome of liver
spleen
adrenal gland
gastrointestinal tract

Publications

Book Chapters

1. **Kamran Farooq, Amir Hussain, Stephen Leslie, Chris Eckl, Warner Slack: *Ontology-driven cardiovascular decision support system.* Pervasive Health 2011: 283-286**
2. **Kamran Farooq, Amir Hussain, Stephen Leslie, Chris Eckl, Calum MacRae, Warner Slack: *An Ontology Driven and Bayesian Network Based Cardiovascular Decision Support Framework*. BICS 2012: 31-41**
3. **Kamran Farooq, Amir Hussain, Stephen Leslie, Chris Eckl, Calum MacRae, Warner Slack: *Semantically Inspired Electronic Healthcare Records*. BICS 2012: 42-51**
4. **Kamran Farooq, Peipei Yang, Amir Hussain, Kaizhu Huang, Chris Eckl, Calum MacRae, Warner Slack: *Efficient Clinical Decision Making by learning from missing Clinical Data*. IEEE SSCI, Singapore 2013: p1024. (Nominated for the best paper award)**
5. **Kamran Farooq, Amir Hussain, Hicham Atassi, Stephen Leslie, Chris Eckl, Calum MacRae, Warner Slack- *A Novel Clinical Expert System for Chest Pain Risk Assessment*. BICS, Beijing, June 2013.**

Journals

1. **Kamran Farooq, Amir Hussain, Hicham Atassi, Stephen Leslie, Chris Eckl, Calum MacRae, Warner Slack - *Effective Clinical Decision Making using Ontology Driven and Machine Learning approach*, Elsevier International Journal of Expert Systems with Applications, *submitted , Aug 2013*.**

2. **Kamran Farooq, Amir Hussain, Stephen Leslie, Chris Eckl, Calum MacRae, Zeeshan Malik, Warner Slack-** ***Data Visualisation framework to improve clinical decision making for RACPC patients,*** **Journal of medical Internet Research*, submitted, September 2013*.**

3. **Kamran Farooq, Amir Hussain, Stephen Leslie, Chris Eckl, Chen Si-Bao, Calum MacRae, Warner Slack Robust** ***–Mixture Modelling with SMEM (Split and Merge EM) Algorithm based on real clinical data collected in the retrospective case study*****. The Journal of the Pattern Recognition,** ***In Preparation.***

4. **Kamran Farooq, Amir Hussain, Hicham Atassi, Zeeshan Malik, Stephen Leslie, Chris Eckl, Calum MacRae, Warner Slack -*Towards Learning from Retrospective Legacy Data for making Effective Prospective Clinical Decision Support Systems*, Springer's Cognitive Computation Journal, *Accepted, September 2013*.**

5. **Zeeshan Malik, Amir Hussain, Kamran Farooq-** ***Incremental Laplacian Eigenmap*****- IEEE Transactions on Neural Networks,** ***submitted , October 2013*****.**

Thank You

Any Questions?

For more details:

http://WWW. COSIPRA.STIR.AC.UK

http://link.springer.com/journal/12559

http://www.cs.stir.ac.uk/events/CICARE2014/