

RESEARCH

Open Access



A novel ontology and machine learning driven hybrid cardiovascular clinical prognosis as a complex adaptive clinical system

Kamran Farooq^{1*}  and Amir Hussain^{2,3}

*Correspondence:
kfa@cs.stir.ac.uk

¹ Computing Science
and Mathematics Division,
University of Stirling, Stirling,
Scotland FK9 4LA, United
Kingdom

Full list of author information
is available at the end of the
article

Abstract

Purpose: This multidisciplinary industrial research project sets out to develop a hybrid clinical decision support mechanism (inspired by ontology and machine learning driven techniques) by combining evidence, extrapolated through legacy patient data to facilitate cardiovascular preventative care.

Methods: The proposed cardiovascular clinical decision support framework comprises of two novel key components:(1) Ontology driven clinical risk assessment and recommendation system (ODCRARS) (2) Machine learning driven prognostic system (MLDPS). State of the art machine learning and feature selection methods are utilised for the prognostic modelling purposes. The ODCRARS is a knowledge-based system which is based on clinical expert's knowledge, encoded in the form of clinical rules engine to carry out cardiac risk assessment for various cardiovascular diseases. The MLDPS is a non knowledge-based/data driven system which is developed using state of the art machine learning and feature selection techniques applied on real patient datasets. Clinical case studies in the RACPC, heart disease and breast cancer domains are considered for the development and clinical validation purposes. For the purpose of this paper, clinical case study in the RACPC/chest pain domain will be discussed in detail from the development and validation perspective.

Results: The proposed clinical decision support framework is validated through clinical case studies in the cardiovascular domain. This paper demonstrates an effective cardiovascular decision support mechanism for handling inaccuracies in the clinical risk assessment of chest pain patients and help clinicians effectively distinguish acute angina/cardiac chest pain patients from those with other causes of chest pain.

Conclusion: The new clinical models, having been evaluated in clinical practice, resulted in very good predictive power, demonstrating general performance improvement over benchmark multivariate statistical classifiers. Various chest pain risk assessment prototypes have been developed and deployed online for further clinical trials.

Keywords: Clinical decision support framework, Cardiovascular decision support framework, Hybrid clinical decision support framework

Introduction

The adoption of clinical decision support systems (CDSSs) in the diagnosis and administration of major chronic diseases e.g. (Dementia Lindgren 2011), cancer, diabetes (OConnor et al. 2011), hypertension (Luitjes et al. 2010) and heart disease (DeBusk et al. 2010) have made significant contributions in improving the clinical outcomes at primary and secondary care healthcare organisations all over the world. CDSS have also made it possible for system developers and knowledge engineers to collate and construct domain expert knowledge for the purpose of clinical risk assessment and screening by clinicians (Khong and Ren 2011).

Clinical decision support systems are being extensively deployed in healthcare settings all over the world. Modern clinical decision support systems are increasingly dissimilar to each other, despite following the same generic architecture which defines a typical CDSS (Burstein et al. 2011). These clinical decision support systems incorporate a variety of innovative techniques to perform various key operations which include clinical knowledge dissemination and collecting patient's medical history for effective clinical decision making. These systems aim to provide clinical decision support and automatic personalised clinical advice through inference capabilities (Mohiuddin 2011). They also help to streamline clinical workflows through integration with electronic healthcare records for patient clinical history collection, diagnosis, inference and training.

Clinical decision support operations are an integral part of modern healthcare management systems. They assist clinicians, patients and healthcare stakeholders by providing expert clinical knowledge and patient-centric information (Classen et al. 2011). The information provided by these intelligent clinical systems is used for clinical decision making in order to improve the effectiveness and quality of healthcare. Automated cardiovascular decision support systems are now being deployed in hospitals and primary care organizations in order to meet the ever growing clinical needs of prognosis in the areas of cardiovascular disease and coronary heart disease. Computerized decision support strategies have already been implemented successfully in several areas of cardiovascular care (Kuperman et al. 2007). These applications are being used as part of the extension of clinical informatics infrastructure in the UK and US. These systems are also being used in both primary and secondary care settings for providing efficient healthcare delivery to its patients. In order to capitalise on the benefits provided by cardiovascular decision support systems, a strong foundation in evidence-based medicine and well-established clinical practice guidelines (CPGs) have to be considered to ensure clinical governance in the next generation clinical systems.

Background

Ontology driven clinical decision support frameworks

An ontology is an explicit specification of a conceptualization. The term is borrowed from philosophy, where an ontology is a systematic account of existence. For AI systems, what "exists" is that which can be represented. When the knowledge of a domain is represented in a declarative formalism, the set of objects that can be represented is called the universe of discourse. This set of objects, and the describable relationships among them, are reflected in the representational vocabulary with which a knowledge-based program represents knowledge. Thus, in the context of AI, we can describe the ontology

of a program by defining a set of representational terms. In such an ontology, definitions associate the names of entities in the universe of discourse (e.g., classes, relations, functions, or other objects) with human-readable text describing what the names mean, and formal axioms that constrain the interpretation and well-formed use of these terms. Formally, an ontology is the statement of a logical theory (Gruber 1993). Ontologies are often equated with taxonomic hierarchies of classes, but class definitions, and the subsumption relation, but ontologies need not be limited to these forms. Ontologies are also not limited to conservative definitions, that is, definitions in the traditional logic sense that only introduce terminology and do not add any knowledge about the world (Herbert and Enderton 1972).

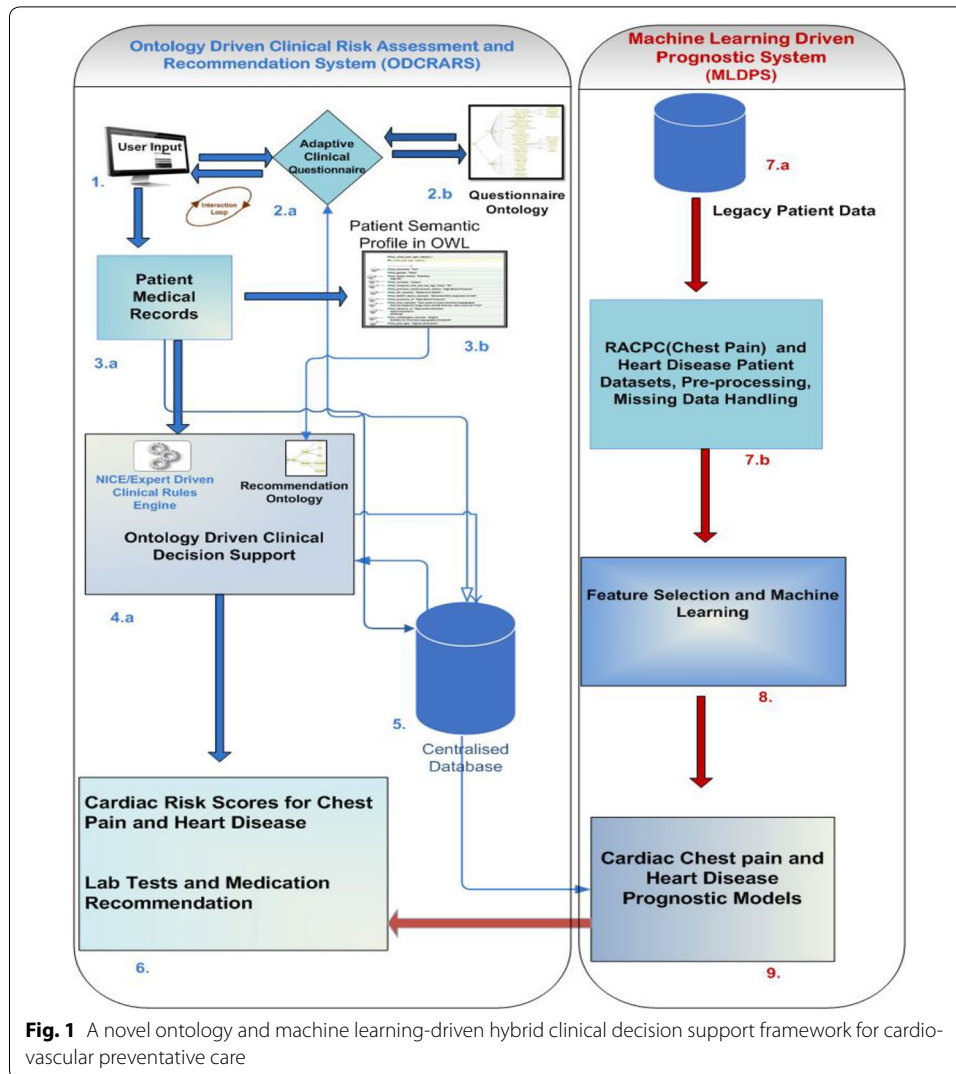
The Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) is an onto-logical resource specifically developed some thirty years ago with a view to standardize healthcare systems. SNOMED CT and with UMLS are clinical thesauruses, aiming to resolve documentation standardization issues in clinical systems. These are large scale medical taxonomies which have been exploited in modern clinical systems showing significant good results in the targeted clinical systems. In Mortensen et al. (2014) it shows that the clinicians using healthcare systems equipped with SNOMED outperformed clinicians using conventional systems without SNOMED CT capabilities.

Machine learning driven cardiovascular decision support systems

Machine learning refers to a type of artificial intelligence algorithm designed to identify patterns in input data, such as patient characteristics, in order to perform complex classification tasks. Machine learning based clinical decision support systems can avoid the bottleneck of knowledge acquisition because knowledge is directly learned through the clinical data. In addition, ML-based clinical decision support systems are able to give recommendations that are generated by non-linear forms of knowledge, and are easily maintainable by simply adding new cases (Chi 2009).

In Nahar et al. (2013), a number of computational intelligence techniques were utilised in the detection of heart disease as a preventative measure. A comparative analysis of six well-known machine learning classifiers was carried out using the Cleveland heart disease dataset. Authors introduced medical knowledge driven feature selection (MFS) and it was compared against the state of the art feature selection algorithms. Their experimental results showed that machine learning classification combined with MFS significantly improved the performance of binary classification. MFS feature selection technique was combined with computerised feature selection process to further refine classification accuracies obtained in previous iterations. MFS combined with Naive Bayes and Sequential minimal optimisation (SMO for training of support vector machine) provided the best classification accuracies and TP (true positive) and F-measure resulted in a higher performance as compare to experimental setups based on state of the art feature selection techniques combined with machine learning classifiers.

We proposed an ontology and machine learning driven hybrid clinical decision support framework for cardiovascular preventative care as shown in Fig. 1. The development of the machine learning driven prognostic system (MLDPS) was carried out in close collaboration with clinical experts. The rapid access chest pain clinic's case study was identified by the consultant cardiologist from Raigmore Hospital in Inverness, UK.



The key objective of the RACPC clinical case study was to help improve the diagnostic and performance capabilities of the RACPC. The heart disease clinical case study was carried out in collaboration with general medical practitioners from UK in order to develop a preventative care mechanism for patients who are at risk of developing heart disease.

The ODCRARS is a knowledge-based system which is based on clinical expert’s knowledge, encoded in the form of clinical rules (utilised by the clinical rules engine) to carry out cardiac risk assessment for various cardiovascular diseases. The MLDPS is a non knowledge-based/data driven prognostic system which is developed by applying machine learning and feature selection techniques on legacy patient datasets. This approach eliminates the need for writing clinical rules thereby reducing dependency on clinical experts to encode their advice in the clinical decision making. Non-knowledge based clinical decision support systems are utilised in providing point-of-care clinical decision making and implementation of such solutions facilitate development of cost effective solutions with improvement in the quality of care provided.

The rest of this paper will be in sections: In “[Background](#)” section, we provide a detailed description of the novel machine learning driven prognostic system based on the chest pain clinical case study and the complete development life cycle followed by validation results. At the end we conclude our findings and provide future directions of our research.

Methods

MLDPS development based on rapid access chest pain clinic’s clinical case study

An iterative development process, based on machine learning and feature selection has been utilised in the development of machine learning driven prognostic models. The MLDPS’s development process is general enough to handle a variety of healthcare datasets which will enable researchers to develop cost effective and evidence based clinical decision support systems. For the purpose of this paper, development and validation of the MLDPS based on the chest pain clinical case study will be discussed in detail. The key stages of the prognostic model development process are shown in Fig. 2. The general description of each stage is as follows:

Results and discussion

The consultant cardiologist from Raigmore Hospital specified a revised clinical requirement to break original patient dataset down into clinical risk factors and lab test results and create two new study groups. The key clinical objective of introducing this demarcation amongst clinical risk factors and lab results was to evaluate the impact of classification results using these two new datasets. So two new study cohorts were created for this purpose as shown in Table 1, so that a comparison could be drawn among two study groups. Another clinical requirement was to compare the clinical effectiveness of two models separately and to classify chest pain patients (predicting risk of cardiac or non

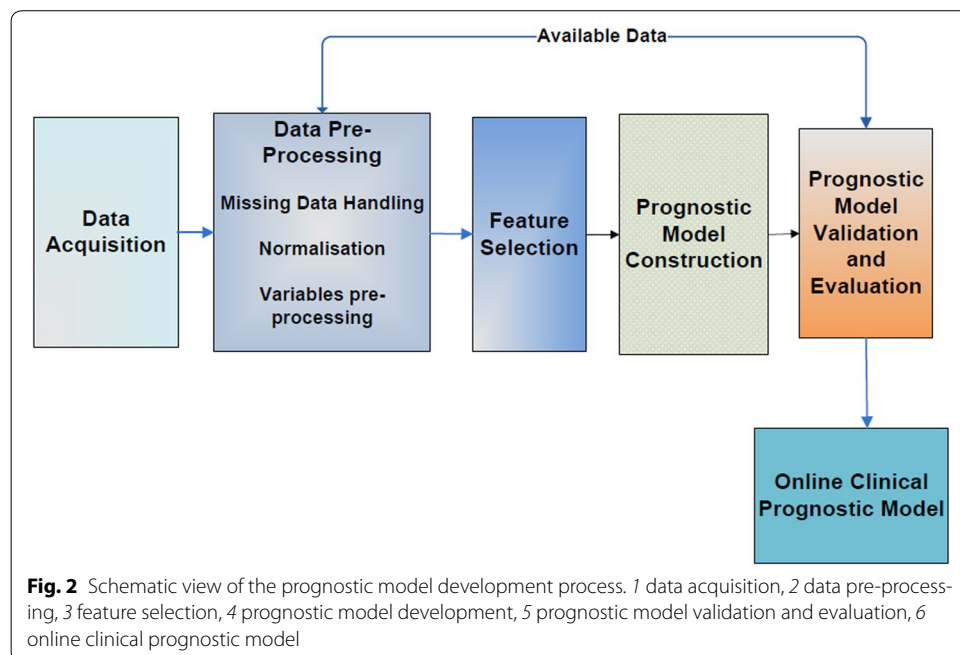


Table 1 Clinical risk factors and test results in two study groups

	Study group 1 Risk factors	Study group 2 Lab test results
1	Smoker	Pathway
2	No of cigarettes	Initial assessment
3	Number of years smoking	ETT result
4	Age	CT result
5	Sex	MPS result
6	Diabetes type	Angio result
7	Hypertension	
8	Raised cholesterol	

cardiac chest pain) purely on the basis of the risk factors and test results information independently.

For the comparative analysis, the original patient dataset was distributed into two study sets as follows:

A detailed comparative analysis of some of the most sophisticated machine learning classifiers combined with state of the art feature selection techniques were utilised for data classification purposes. Experimental setups comprises of the logistic regression (LR), decision tree (DT) and support vector machine (SVM) classifiers combined with forward selection (FS), backward selection (BS), sequential forward floating selection (SFFS), P value feature selection, minimum redundancy and maximum relevance feature selection (mRMR) techniques were utilised. The expert driven (ED) feature selection i.e. pre-selected clinical variables by the clinical domain expert is compared with the state of the art feature selection techniques.

Study group 1: clinical risk factors

In the study group 1, patient demographics including clinical risk factors are included for the comparative analysis purpose. In the first stage, state of the art machine learning classifiers and feature selection techniques are utilised. The experimental setups used for this purpose are shown in the Table 2. Candidate clinical variables preselected by the clinical domain expert were classified using the LR, DT and SVM classifiers and results were compared with the state of the art feature selection methods as shown in our experimental setups. The purpose of expert-driven (ED) data classification was to develop a baseline model using the LR classifier.

As it can be seen in Table 2, the LR based classification setups combined with backward feature selection method (smoker, number of years smoking, age, diabetes type and raised cholesterol) were able to classify the RACPC patient dataset with a classification accuracy of 68.99 %. Also, it is interesting to find out that the DT combined with BS feature selection method classified the patient dataset with a classification accuracy of 65.05 % using just one feature, which is patient’s age. The SVM combined with FS, classified the patient dataset with a classification accuracy of 70.07 % using patient’s age, sex and hypertension. In the case of SVM (linear kernel function), similar clinical variables were picked up by the BS wrapping technique.

Table 2 Study group 1 (risk factors)- feature selection

	Experimental setup	Selected features	Accuracy
1	LR + FS	4, 5, 6, 2, 1, 3	68.45
2	LR + BS	1, 3, 4, 5, 6, 8	68.99
3	LR + ED	All	66.12
4	LR + SFFS	4, 5, 6	67.92
5	LR + P-value	4, 5, 7, 8, 6, 3, 1, 2	66.12
6	LR + mRMR	4, 5, 7, 6, 8, 3, 1, 2	66.12
7	DT + FS	4, 7, 8, 6, 2	65.41
8	DT + BS	4	65.05
9	DT + ED	All	62.36
10	DT + SFFS	4	65.05
11	DT + P value	4, 5, 7, 8, 6, 3, 1, 2	62.36
12	DT + mRMR	4, 5, 7, 6, 8, 3, 1, 2	62.36
14	SVM + FS	4, 5, 1	70.07
15	SVM + BS	4, 5, 7	69.71
16	SVM + ED	All	68.45
17	SVM + SFFS	4, 5, 1	70.07
18	SVM + P value	4, 5, 7, 8, 6, 3, 1, 2	68.45
19	SVM + mRMR	4, 5, 7, 6, 8, 3, 1, 2	68.45

SFFS, is classed as a refined forward selection method, is also utilised in all of our clinical case studies. Results of SFFS combined with LR, DT and SVM, were compared with the BS, FS, P value and mRMR methods to analyse its effectiveness. The results of SVM + SFFS with a more transparent logistic regression based model combined with BS, demonstrate that using three clinical variables, patient’s cardiac chest pain can be distinguished (whether it is cardiac or non-cardiac). So performance complexity trade-offs can be considered if the clinical support decision function requires higher degree of accuracy by comprising on transparency of a clinical prognostic model.

Evaluation

After extracting features and identifying those with most discriminative power for each classifier, *k*-fold cross validation, leave-one-out validation (LOOCV) is performed in order to assess the performance of these classifiers. The experimental results reported in confusion matrices show that the LR + BS, DT + FS and SVM + SFFS are the best classification setups given the imbalanced nature of the patient dataset. Because our two classes (cardiac and non cardiac) are not equally distributed, different evaluation measurements are reported, namely weighted accuracy, unweighted accuracy, precision, recall, F-measure and Matthew’s correlation are reported in Table 4. The confusion matrices for LR, DT and SVM based classification setups and weighted classification accuracies are reported in Tables 3, 5 and 6. True positive (TP), false negative (FN), false positive (FP), true negative (TN) rates are provided for the actual and predicted outputs (classification outputs).

In order to quantify performances of the best classification setups, the Receiver Operating Characteristic (ROC) curves are used as shown in Fig. 3 (evaluating the underlying area), which compare the specificity and sensitivity of experimental setups. In clinical domain, ROC curve analysis is used to determine the cut off value for a clinical test.

Table 3 The confusion matrix of LR and feature selection based classification setups, study group 1

Predicted output												
Actual	LR+FS		LR+BS		LR+ED		LR+SFFS		LR+P		LR+mRMR	
A	197	87	193	91	188	96	194	90	188	96	188	96
B	89	185	82	192	93	181	89	185	93	181	93	181
Accuracy	68.45		68.99		66.12		67.92		66.12		66.12	

Table 4 Experiment results in terms of different evaluation measurements

	LR + BS (%)	DT + FS (%)	SVM + SFFS (%)
Weighted accuracy	68.99	65.41	70.07
Unweighted accuracy	69.01	65.38	70.18
Precision	67.96	66.90	63.73
Recall	70.18	65.74	73.88
Fmeasure	69.05	66.32	68.43
Matthew's correlation	38.03	30.78	40.67

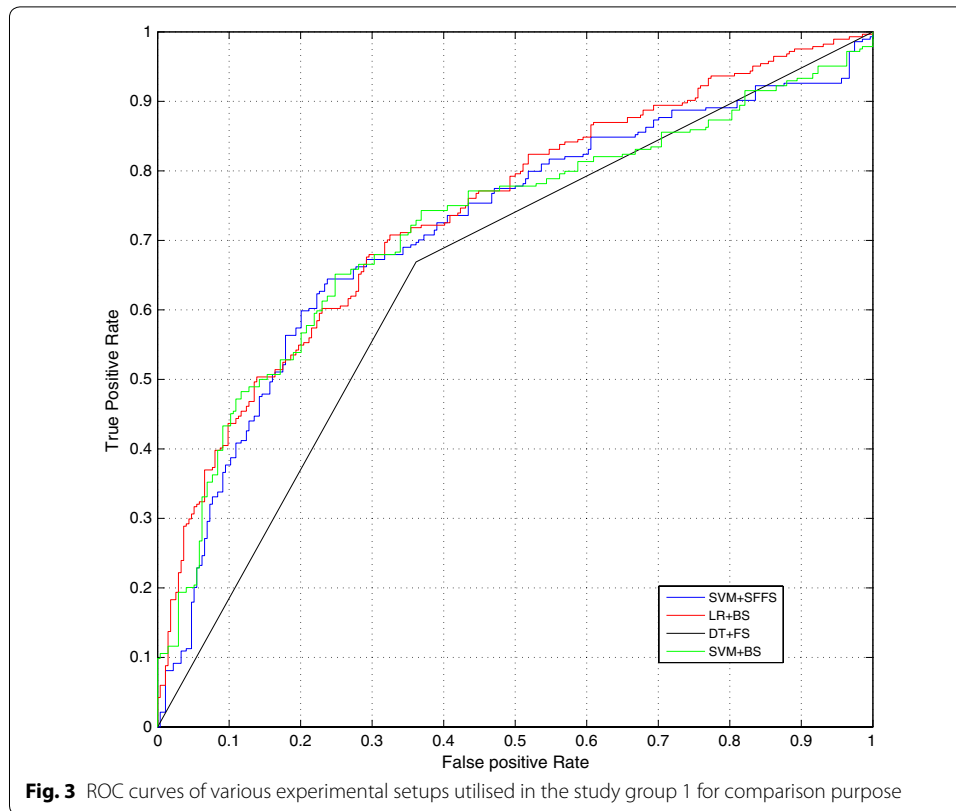
Table 5 Confusion matrix of DT and feature selection based classification setups, study group 1

Predicted output												
	DT + FS		DT + BS		DT + ED		DT + SFFS		DT + P		DT + mRMR	
Actual												
A	190	94	170	114	169	115	170	114	169	115	169	115
B	99	175	81	193	95	179	81	193	95	179	95	179
Accuracy	65.41		65.14		62.3656		65.05		62.36		62.36	

Table 6 Confusion matrix of SVM and feature selection based classification setups, study group 1

Predicted output												
	SVM + FS		SVM + BS		SVM + ED		SVM + SFFS		SVM + P		SVM + mRMR	
Actual												
A	181	103	183	101	179	105	181	103	179	105	179	105
B	64	210	68	206	71	203	64	210	71	203	71	203
Accuracy	70.07		69.71		68.45		70.07		68.45		64.45	

The ROC curve is a graph of sensitivity (y-axis) vs. 1- specificity (x-axis). Maximizing sensitivity corresponds to some large y value on the ROC curve. Maximizing specificity corresponds to a small x value on the ROC curve. Thus a good first choice for a test cut-off value is that value which corresponds to a point on the ROC curve nearest to the



upper left corner of the ROC graph. This is not always true however. For example, in the cardiac risk assessment it is important not to miss detecting a patient with cardiac chest pain therefore it is more important to maximize sensitivity (minimize false negatives) than to maximize specificity. In this case the optimal cut-off point on the ROC curve will move from the vicinity of the upper left corner over toward the upper right corner.

Performance evaluation of experimental setups

In addition to the ROC curve analysis which is used to evaluate the performance of best classification setups. A one way ANOVA (analysis of variance) is also employed to compare means of classification accuracies obtained in three experimental setups to establish whether the difference in classification accuracies within groups and among other classifiers is significant or they are statistically equal. Table 7 shows detailed analysis of the one-way ANOVA test which is performed using LR, DT and SVM experimental setups.

In the summary section, it shows the average classification accuracies of the LR,DT and SVM classification groups.

For the single factor Anova test, the null hypothesis is defined as follows:

$H_0 : \mu_1 = \mu_2 = \mu_3$ (the means are all equal, hence the difference in means in all of three experimental setups are all the same)

$H_1 : \text{At least two of the means are different}$

$\alpha = 0.05$

In the ANOVA section in Table 7, sum of squares (SS), degree of freedom (df) and mean square values are provided. As it can be seen that the F statistic value (28.34) is

Table 7 One-way ANOVA test for the performance evaluation of LR, DT and SVM based classification setups

Anova: single factor						
Summary						
Groups	Count	Sum	Average	Variance		
Logistic regression	6	403.72	67.28	1.7478		
Decision tree	6	382.59	63.765	2.38611		
Support vector machine	6	415.2	69.2	0.69228		
ANOVA						
Source of variation	SS	df	MS	F	P value	F crit
Between Groups	91.20	2	45.60	28.34	8.02793E-06	3.68
Within Groups	24.13	15	1.6087			
Total	115.3354944	17				

greater than the critical value of F (8.02). Also the P value is <0.05, so on this basis the null hypothesis is rejected and it is now established that the difference in the classification accuracies within groups and among other classifiers is statistically significant.

Study group 2: lab test results

In this study group, clinical variables representing various lab test results are included for the comparative analysis purpose. The statistical P values for the clinical variables involved in this study group are provided in Table 8. It shows that the “Pathway”, “Initial assessment”, “ETT” and “CT result” are the most significant clinical variables in the list. The state of the art feature selection and machine learning techniques are applied. Details of the LR, DT and SVM based machine learning setups are provided in the Table 9. As it can be seen, that 18 experimental setups are employed to classify the patient data in study group 2. An expert driven (pre-selection by clinical domain expert) feature selection and LR based baseline model was developed which was then compared with state of the art machine learning and feature selection techniques.

As it can be seen in the Table 9, “initial assessment” is a common clinical variable amongst the majority classification groups. It is interesting to notice that LR + FS and LR + SFFS based experimental setups attained the best classification accuracy using only one variable (initial assessment). The best classification setups are DT + FS, DT + BS, DT + SFFS. All of these setups handled the data sparsity issue with a classification accuracy of 82.97 %. “CT scan result” is also found to be common among the majority

Table 8 P values of the clinical variables (study group 2)

	Clinical variables Lab test results	P value
1	Pathway	1.93e-27
2	Initial assessment	1.48e-21
3	ETT result	0.04
4	CT result	0.05
5	MPS result	0.17
6	Angio result	0.9

Table 9 Feature selection results, study group 2 (test results)

	Experimental setup	Selected features	Accuracy (%)
1	LR + FS	2	69.89
2	LR + BS	1, 4, 5, 6	72.58
3	LR + ED	All	67.92
4	LR + SFFS	2	69.89
5	LR + P value	6, 2, 5, 1, 4, 3	67.92
6	LR + mRMR	2, 6, 1, 5, 4, 3	67.92
7	DT + FS	2, 6, 4, 3	82.97
8	DT + BS	2, 3, 4, 6	82.97
9	DT + ED	All	81.89
10	DT + SFFS	2, 6, 4, 3	82.97
11	DT + P value	6, 2, 5, 1, 4, 3	81.89
12	DT + mRMR	2, 6, 1, 5, 4, 3	81.89
14	SVM + FS	2, 3	70.96
15	SVM + BS	2, 4, 5	70.96
16	SVM + ED	All	68.63
17	SVM + SFFS	2, 3	70.96
18	SVM + P value	6, 2, 5, 1, 4, 3	68.63
19	SVM + mRMR	2, 6, 1, 5, 4, 3	68.63

classification groups. These findings corroborate the high statistical P values of “Initial assessment and CT scan result” and re-iterate their significance in the clinical decision making. The performance complexity trade-offs in this case could be considered to limit the amount of tests (by focussing on the most significant tests picked up in the classification setups), needed to diagnose a patient with cardiac chest pain.

Evaluation

After the feature extraction stage, a k -fold cross validation based leave-one-out validation (LOOCV) technique is used for performance evaluation of the classification methods. The confusion matrices of LR, DT and SVM combined with state of the art feature selection techniques are shown in Tables 5, 6 and 11.

The DT + FS, DT + SFFS, DT + BS and DT + mRMR classification groups are selected for analysis. In Table 10, different evaluation measurements are provided. As our two classes (cardiac and non cardiac) are not equally distributed which is why weighted accuracies and other measurements are reported. The confusion matrices of LR, DT and SVM based classification setups and weighted classification accuracies are

Table 10 Experiment results in terms of different evaluation measurements

	DT + FS (%)	DT + SFFS (%)	DT + BS (%)	DT + mRMR (%)
Weighted accuracy	82.97	81.89	82.97	82.97
Unweighted accuracy	83.09	81.98	83.09	83.09
Precision	76.41	77.46	76.41	76.41
Recall	88.57	85.60	88.57	88.57
Fmeasure	82.04	81.33	82.04	82.04
Matthew's correlation	66.68	64.15	66.68	66.68

Table 11 Confusion matrix obtained using LR based classification setups

Predicted output													
		LR + FS		LR + BS		LR + ED		LR + SFFS		LR + P		LR + mRMR	
Actual													
A		142	142	248	36	206	78	142	142	206	78	208	78
B		26	248	117	157	101	173	26	248	101	173	101	173
Accuracy (%)		69.89		72.58		67.92		69.89		67.92		67.92	

provided in Tables 11, 12 and 13. True positive (TP), false negative (FN), false positive (FP), true negative (TN) rates are provided for the actual and predicted outputs.

The receiver operating characteristic (ROC) curves are used to quantify performances of the best classification groups. In Fig. 4, performances of DT and LR based setups are plotted which compare the specificity and sensitivity in our experimental setups.

Performance evaluation of experimental setups

In addition to the ROC curve analysis, a one way ANOVA test is also utilised for the performance evaluation of the best classification groups. The one-way ANOVA test is used to compare means of classification accuracies obtained in three experimental setups. This test is used to ascertain whether the difference/improvement in classification accuracies within different classification groups and other classifiers (across different classification methods) is significant or they all are equal.

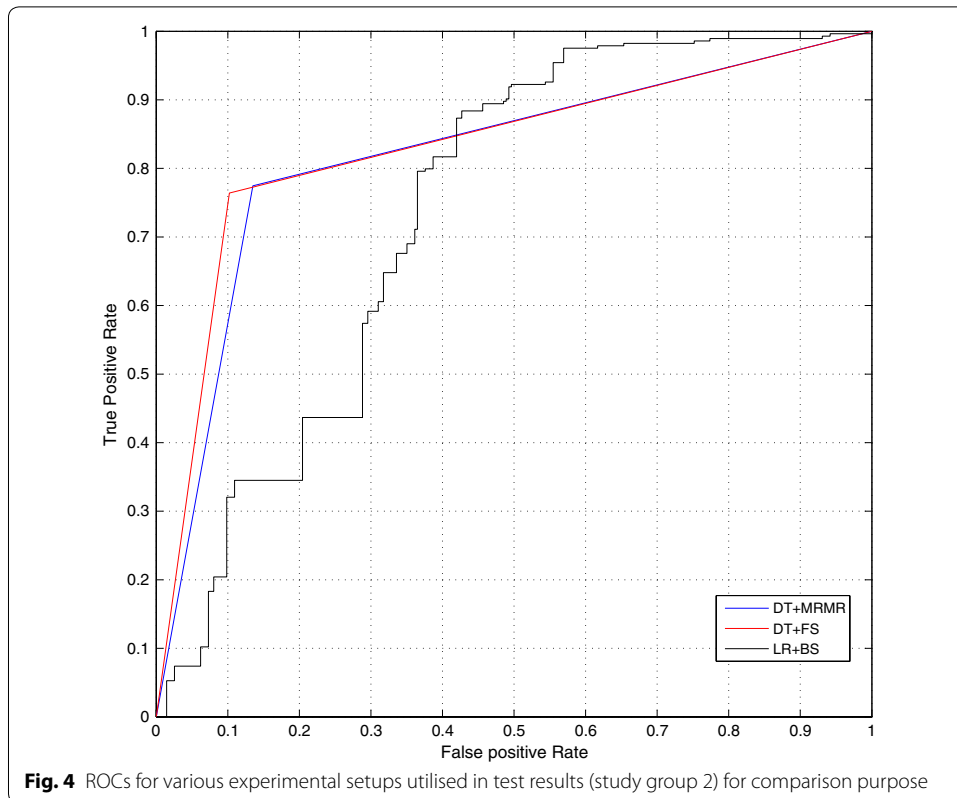
Table 14 provides detailed analysis of the one-way ANOVA. In the summary section, the average classification accuracies are calculated based on LR, DT and SVM classification setups.

Table 12 Confusion matrix obtained using DT based classification setups

Predicted													
		DT + FS		DT + BS		DT + ED		DT + SFFS		DT + P		DT + mRMR	
Actual													
A		217	67	217	67	220	64	217	67	220	64	220	64
B		28	246	28	246	37	237	28	246	37	237	37	237
Accuracy (%)		82.97		82.97		81.89		82.97		81.89		81.89	

Table 13 Confusion matrix obtained using SVM based classification setups

Predicted output													
		SVM + FS		SVM + BS		SVM + ED		SVM + SFFS		SVM + P value		SVM + mRMR	
Actual													
A		142	142	142	142	214	70	142	142	214	70	214	70
B		20	254	20	254	105	169	20	254	105	169	105	169
Accuracy (%)		70.96		70.96		68.63		70.96		68.63		68.63	



For the single factor ANOVA test, the null hypothesis is declared as follows:

$H_0 : \mu_1 = \mu_2 = \mu_3$ (the means are all equal, hence the difference in means in all of three experimental setups are all the same)

H_1 : At least two of the means are different

$\alpha = 0.05$

In the ANOVA section in Table 14, sum of squares (SS), degree of freedom (df) and mean square values are provided. As it can be seen that the F statistic value (183.50) is greater than the critical value of F (3.682). Also the P value is <0.05, so on this basis the null hypothesis is rejected and it is now established that the difference in the

Table 14 One-way ANOVA Test for the performance evaluation of LR, DT and SVM based classification setups (study group 2- test results)

Anova: single factor						
Summary						
Groups	Count	Sum	Average	Variance		
Logistic regression	6	416.12	69.35	3.4301		
Decision tree	6	494.58	82.43	0.34992		
Support vector machine	6	418.77	69.795	1.62867		
ANOVA						
Source of variation	SS	df	MS	F	P value	F crit
Between groups	661.6750111	2	330.83	183.50	2.8522E-11	3.682
Within groups	27.04368333	15	1.802912222			
Total	688.7186944	17				

classification accuracies within groups and among other classifiers (across LR, DT and SVM classification groups) is statistically significant.

Implementation of online clinical prognostic models

In the RACPC clinical case study, three datasets are utilised for the development of machine learning prognostic models for Raigmore Hospital’s RACPC clinicians. The results obtained through three patient datasets were analysed by the consultant cardiologist from Raigmore Hospital. It was decided to develop online cardiac chest pain prognostic models based on LR based classification setups which are shown in Table 15. The cardiac chest pain prognostic model has been developed using the first patient dataset containing both patient demographics and lab test results information. This was selected by the clinical domain experts for further development. Two expert driven RACPC cardiac chest pain prognostic models have also been developed and deployed online for clinical validation.

Logistic regression-based cardiac chest prognostic models have been developed and deployed online for the initial clinical validation by the consultant cardiologist from Raigmore hospital. Clinical questionnaires are encoded in HTML; logistic regression model is programmed in PHP, which generates an HTML page after data is collected from an HTML input form. The probability of cardiac chest pain risk score is calculated when user presses the “calculate score” button.

The machine learning cardiac chest pain prognostic model is intended to be used by RACPC clinicians. The user is asked to provide patient demographics information and details of CT, ETT and MPS lab test results. The cardiac chest pain risk score is calculated using the formula as shown below:

$$SCORE = 100.(1 + e^{-M})^{-1}$$

where

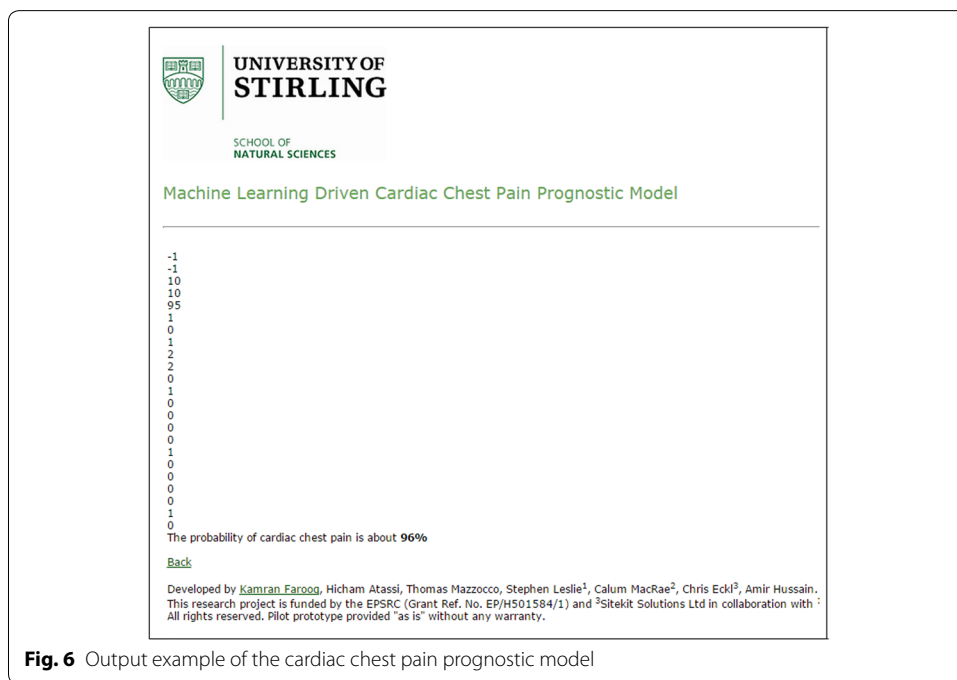
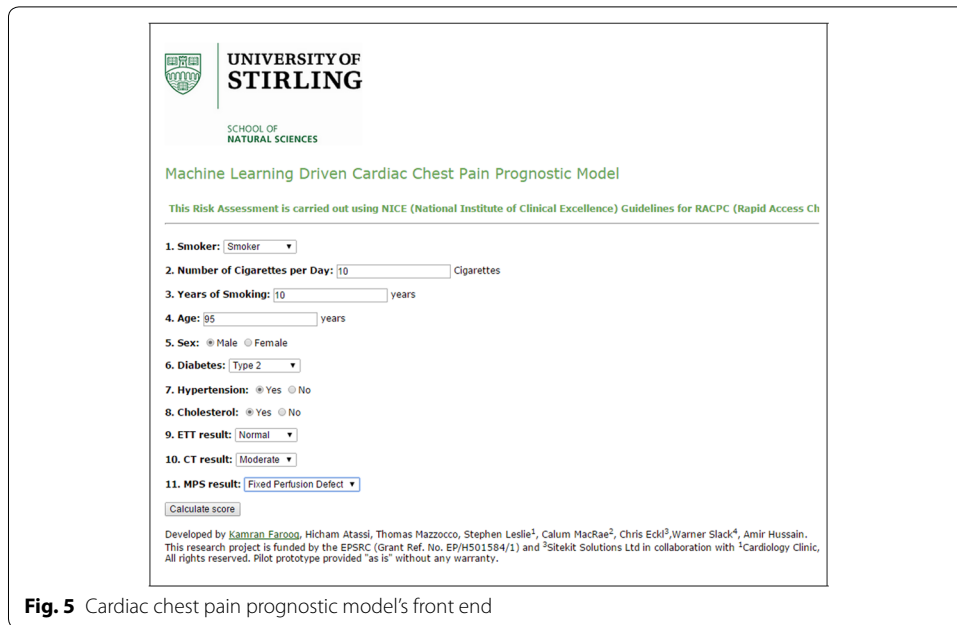
M co-efficients of each clinical variable used in the model.

The logistic regression model calculates the probability of cardiac chest pain using series of inputs as shown in Fig. 6.

The initial cardiac chest pain prognostic model as in Fig. 5 was validated by clinical domain expert from Raigmore Hospital. In the developed cardiac chest pain prognostic

Table 15 Classification setups considered for the development of machine learning driven cardiac chest pain prognostic model

Best classification setups		
Risk factors and test results		
Experimental setups	Selected features	Weighted classification Accuracy (%)
LR + FS	INA, AGE, ANG, SEX, MPS, YOS, NOC, HPT, PWY, ETT, CT, SMR	74.68
LR + BS	SMR, YOS, AGE, PWY, SEX, HPT, INA, CT, MPS, ANG	74.68
DT + SFFS	ANG, INA, CTT, ETT	78.63
DT + FS	ANG, INA, CT, ETT, DAB, SEX	77.84
SVM + FS	ANG, INA, CT, SEX, ETT, PWY, AGE, MPS, CHL, YOS	78.16
SVM + BS	YOS, AGE, PWY, SEX, HPT, CHL, INA, CT, MPS, ANG	78.32



model, we first determined the optimal number of variables, after applying *k*-fold cross-validation strategy, followed by development of prognostic model keeping in view clinical requirements of RACPC. The developed model calculates probability of cardiac chest pain. Two additional cardiac chest pain prognostic models have also been developed as per the clinical needs of Raigmore hospital's RACPC. In the second cardiac chest pain prognostic model, it was suggested to include additional two clinical variables, "Initial

assessment” and “Angio result”. LR classifier is used in the development of these expert driven prognostic models shown in Fig. 7.

Figure 8 shows the third cardiac chest pain prognostic model which is developed to calculate cardiac chest pain risk score using minimal set of variables. This cardiac chest pain prognostic model provides a cost effective cardiac chest pain risk assessment mechanism by using patient demographics and minimal lab test results, thereby reducing cost and dependency on CT scan and initial assessment procedures.

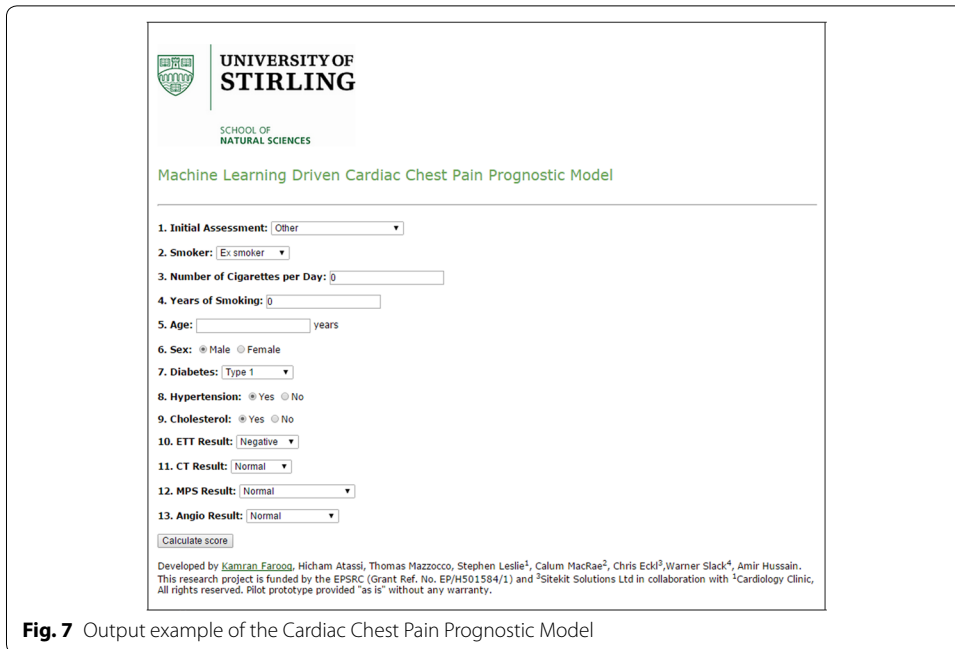


Fig. 7 Output example of the Cardiac Chest Pain Prognostic Model

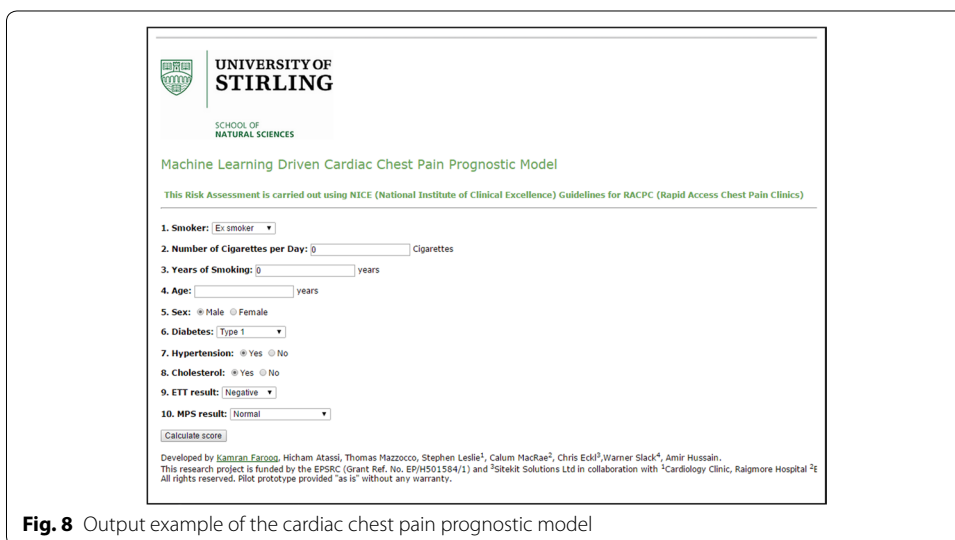


Fig. 8 Output example of the cardiac chest pain prognostic model

Validation of the machine learning driven system (MLDPS) and ontology driven clinical risk assessment and recommendation system (ODCRARS)

Clinical validation of the MLDPS involved testing of the web based prognostic models for cardiac chest pain, heart disease and breast cancer. Breast cancer prognostic models are not part of the ODCRARS and validation of these clinical models was carried out by an oncologist from the Beatson cancer centre in Glasgow. The cardiac chest pain and heart disease prognostic models were validated by a consultant cardiologist and a general medical practitioner from UK.

The machine learning driven cardiac chest pain prognostic model was developed under the supervision of a consultant cardiologist from Raigmore Hospital. This clinical model is developed using clinical features extracted in the RACPC clinical case study. The model was tested using clinical use cases for non-cardiac and known cardiac chest pain patients for clinical validation and sanity checking purposes.

The patient data was generated using the ODCRARS’s web front end. Patient demographics and past medical history were collated during patient’s review of the system which has been conducted using the patient’s interface. The patient data required for the cardiac chest pain risk score calculation was populated through the ODCRARS. As it can be seen in Fig. 9, system calculates cardiac risk scores for the selected patient for various cardiovascular diseases. The outcome risk scores over 4 and 10 year period, calculated using Framingham Heart Study (FHS) are provided in the doctor’s module.

The ODCRARS provides dedicated graphical user interface for the clinicians and patients to record their interactions with the system. Cardiologist using the doctor’s interface reviews patient data which was provided during the patient interview, conducted through an ontology driven intelligent context-aware information collection component. After reviewing patient’s summary data, the clinician carries out clinical risk assessment by clicking on the “Risk assessment” button. System brings up information on the front end as shown in Fig. 10, which shows details of cardiovascular risk assessment carried out

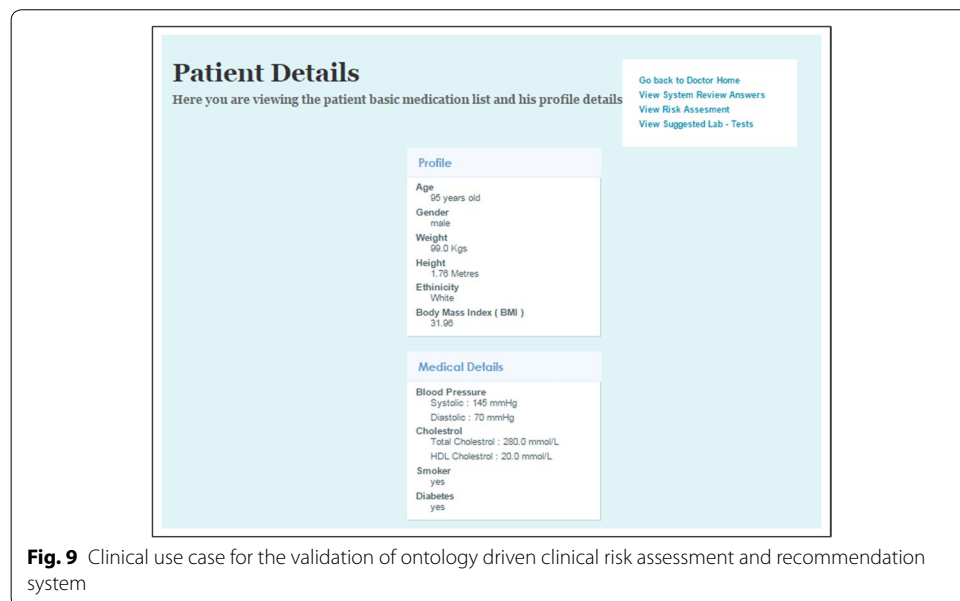


Fig. 9 Clinical use case for the validation of ontology driven clinical risk assessment and recommendation system

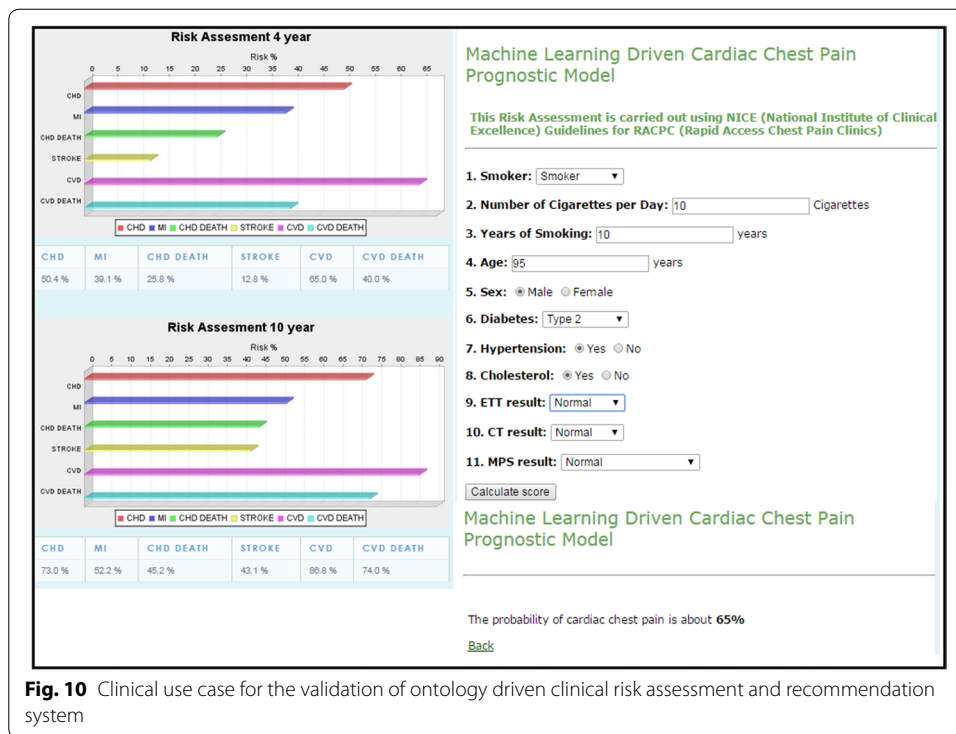
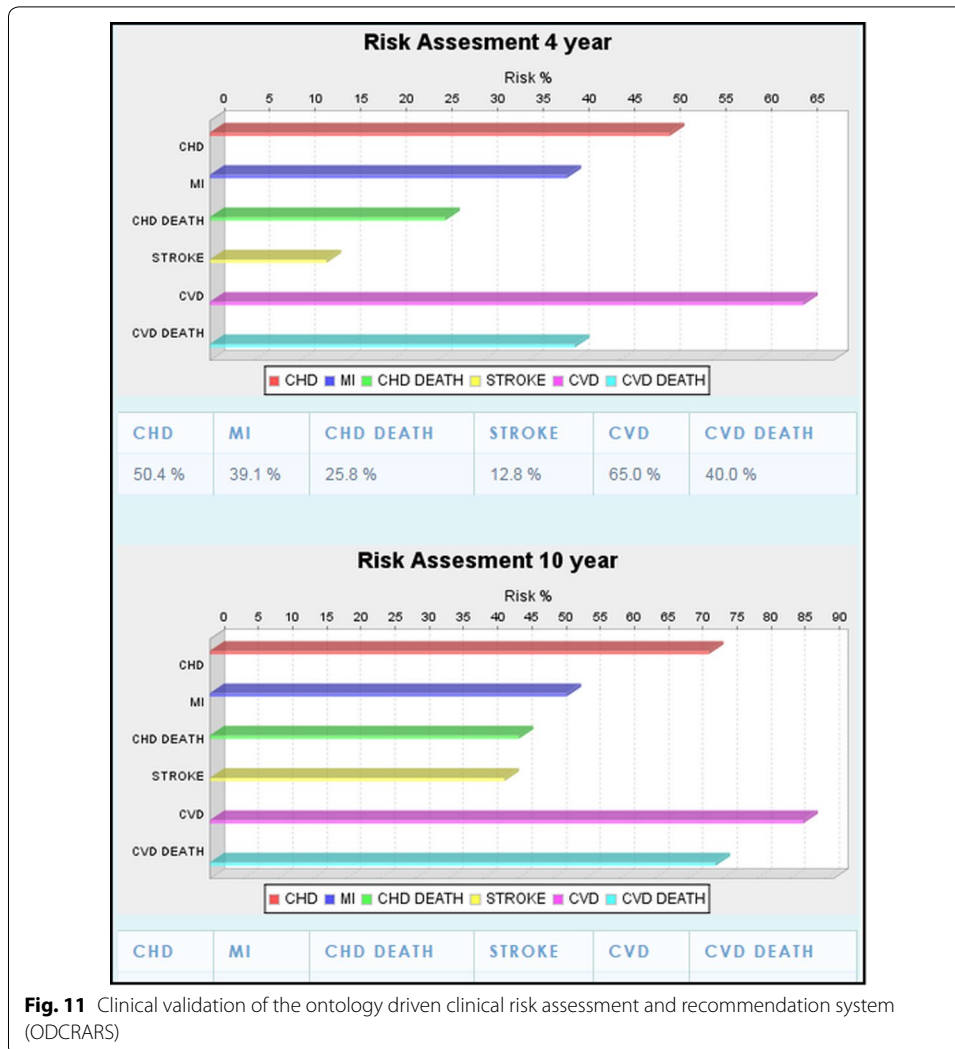


Fig. 10 Clinical use case for the validation of ontology driven clinical risk assessment and recommendation system

through ODCRARS. System provides details of cardiac risk scores for CHD, MI, CHD Death and Stroke conditions as shown in Fig. 11. It also brings up patient demographics information as shown in the Fig. 9, this information was provided during the patient registration procedure. The cardiologist also carries out cardiac chest pain risk assessment by clicking on the “Calculate score” button. The machine learning driven cardiac chest pain prognostic model calculates the cardiac chest pain risk score which is shown in Fig. 10. The ODCRARS provides a complete cardiac risk assessment profile for the patient selected by the clinician. In the “Risk assessment” module, cardiologist launches the machine learning driven heart disease prognostic model by clicking on the “heart disease prognostic model” link to verify information populated on the screen. Clinician then clicks on the “Calculate button to generate the heart disease risk score as shown in Fig. 12.

Clinical validation of the machine learning driven cardiac chest pain and heart disease prognostic models was carried out in a limited case study by a general medical practitioner from Edinburgh, Scotland. The focus of this clinical case study was to detect high-risk patients with ischaemic heart disease by carrying out cardiac risk assessment of patients using the machine learning driven prognostic models incorporated in the ‘Risk assessment module of the ODCRARS. Clinical trials were conducted using the in-house patient data to assess clinical prototypes suitability for general medical practitioners.

The ODCRARS, especially machine learning driven cardiac chest pain and heart disease prognostic models were presented at various e-health workshops and symposiums. The look and feel of these clinical prototypes was refined to incorporate users’ feedback, and adherence to usability guidelines for web browsers and mobile phone users. Also, clinical prototypes were demonstrated in an invited speaker talk at the Beth Israel Deaconess Medical Centre of Harvard Medical School.



Conclusions

We have demonstrated the design, development and validation of the machine learning driven prognostic system (MLDPS). It has been brought to light as a result of this clinical case study that we do not need all of the expensive lab tests to figure out if patient is presenting cardiac chest pain related symptoms. The Initial Assessment, CT and ET lab test results could provide the much needed clinical decision support to clinicians to reach patient diagnosis. We demonstrated a novel Machine Learning driven Prognostic system which was developed to help clinicians automatically distinguish cardiac chest patients from others with non-cardiac chest pain.

We have demonstrated clinical effectiveness of our proposed clinical decision support framework through clinical case studies in the cardiovascular domain. The proposed ontology and machine learning driven hybrid clinical decision support framework exploits functionality provided by each of its key components. Moreover, it brings/integrate them together in an intelligent manner to deliver a cost effective, holistic and efficient cardiovascular clinical risk assessment mechanism. The proposed clinical decision

Machine Learning Driven Cardiac Chest Pain Prognostic Model

This Risk Assessment is carried out using NICE (National Institute of Clinical Excellence) Guidelines for RACPC (Rapid Access Chest Pain Clinics)

1. Smoker:
2. Number of Cigarettes per Day: Cigarettes
3. Years of Smoking: years
4. Age: years
5. Sex: Male Female
6. Diabetes:
7. Hypertension: Yes No
8. Cholesterol: Yes No
9. ETT result:
10. CT result:
11. MPS result:

The probability of cardiac chest pain is about 65%

[Back](#)

Machine Learning Driven Heart Disease Prognostic Model

1. Age: Years
2. Sex: Male Female
3. Chest Pain Type:
4. Exercise Induced Angina: Yes No
5. Resting BP (in mm Hg on admission to the hospital):
6. Serum Cholesterol in mmol/L: (If Unknown leave this field empty)
7. (Fasting Blood Sugar > 120 mg/dl) ? Yes No
8. Resting Electrocardiography Results 0 1 2
 - 1. 0: Normal
 - 2. 1: Having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05
 - 3. 2: Showing probable or definite left ventricular hypertrophy by Estes' criteria
9. ST Depression Induced by Exercise Relative to Rest :
10. ST Segment:
11. Number of Major Vessels Coloured by Fluoroscopy: 0 1 2 3
12. Thallium Treadmill Stress Test: Maximum Heart Rate Achieved:
13. Thallium Heart Scans:

The Probability of Heart Disease is about 67%

[Back](#)

Fig. 12 Cardiac chest pain risk score calculation as part of the integrated ODCRARS

support framework could also be utilised in the clinical risk assessment of other chronic illnesses. We have also explained the functionality of a comparative machine learning and feature selection techniques, used in the development of the prognostic system. The MLDPS is validated by clinical domain experts in the RACPC, heart disease and breast cancer domains.

Our proposed MLDPS provides prognostic models for the RACPC clinicians to distinguish cardiac chest pain patients from those with non-cardiac symptoms. Our proposed clinical decision support framework provides a foundation for future clinical decision support systems to follow a multi-layered clinical decision support framework approach by learning from evidence-based/data driven legacy clinical data. Learning from legacy clinical data activity, provides an opportunity to reverse engineer existing clinical workflows, in order to remove redundant clinical pathways thereby providing clinicians recommendations/suggestions to refine clinical workflows.

The proposed clinical decision support framework utilises clinical expert’s knowledge, which is encoded in the form of clinical rules for clinical recommendation purposes. Also, it makes use of clinical rules (encoded in the form of look-up tables, statistical equations) provided in the Framingham Heart Study (FHS) for the cardiac risk score calculation for various cardiovascular diseases.

Authors' contributions

KF carried out the study, developed the proposed novel ontology and machine learning driven hybrid clinical decision support framework, worked on the clinical case studies under the close supervision of consultant cardiologists from UK and US hospitals and drafted this manuscript. AH actively participated during all research progress in the capacity as project's principal supervisor. Both authors read and approved the final manuscript.

Author details

¹ Computing Science and Mathematics Division, University of Stirling, Stirling, Scotland FK9 4LA, United Kingdom. ² Computing Science and Mathematics Division, University of Stirling, Stirling, Scotland FK9 4LA, United Kingdom. ³ Anhui University, Hefei, China.

Acknowledgements

This research project is funded by the EPSRC (Grant Ref.no. EP/H501584/1) and Sitekit Solutions Ltd. We would like to thank Professor Stephen Leslie, consultant cardiologist from Raigmore Hospital in Inverness, Scotland, UK for providing the required domain expertises as well as facilitating us to utilise the RACPC patient data for the development and validation of the proposed machine learning driven prognostic system. This research is also supported by The Royal Society of Edinburgh (RSE) and The National Natural Science Foundation of China (NNSFC) under the RSE-NNSFC joint project (2012-2015) [Grant number 61211130309] with Anhui University, China, and the Sino-UK Higher Education Research Partnership for Ph.D. Studies joint-project (2013-2015) funded by the British Council China and The China Scholarship Council (CSC).

Competing interests

The authors declare that they have no competing interests.

Received: 26 September 2015 Accepted: 28 June 2016

Published online: 12 July 2016

References

- Burstein F, Brezillon P, Zaslavsky A (2011) Supporting real time decision-making. Springer, New York
- Chi CL (2009). Theses and dissertations. p. 283
- Classen DC, Phansalkar S, Bates DW (2011) Critical drug-drug interactions for use in electronic health records systems with computerized physician order entry: review of leading approaches. *J Patient Saf* 7(2):61
- DeBusk RF, Houston-Miller N, Raby L (2010) *J Am Coll Cardiol* 55(10):A132
- Gruber T (1993) What is an ontology
- Herbert BE, Enderton A (1972) Mathematical introduction to logic. Academic press, Cambridge
- Khong P, Ren R (2011) Identification and control. *Int J Model* 12(1):133
- Kuperman GJ, Bobb A, Payne TH, Avery AJ, Gandhi TK, Burns G, Classen DC, Bates DW (2007) Using the wisdom of the crowds to find critical errors in biomedical ontologies: a study of SNOMED CT. *Journal of the American Medical Informatics Association* 14(1):29
- Lindgren H (2011) Artificial intelligence in medicine. Springer, New York, pp 129–138
- Luitjes SH, Wouters MG, Franx A, Scheepers HC, Coupé VM, Wollersheim H, Steegers EA, Heringa MP, Hermens RP, van Tulder MW (2010) Study protocol: Cost effectiveness of two strategies to implement the NVOG guidelines on hypertension in pregnancy: An innovative strategy including a computerised decision support system compared to a common strategy of professional audit and feedback, a randomized controlled trial. *Implement Sci* 5:68
- Mohiuddin SG (2011). Enabling health, independence and wellbeing for patients with bipolar disorder through personalised ambient monitoring. Ph.D. thesis, University of Southampton
- Mortensen JM, Minty EP, Januszky M, Sweeney TE, Rector AL, Noy NF, Musen MA (2014) Using the wisdom of the crowds to find critical errors in biomedical ontologies: a study of SNOMED CT. *J Am Med Inform Assoc* 22(3):640–648
- Nahar J, Imam T, Tickle KS, Chen YPP (2013) Computational intelligence for heart disease diagnosis: A medical knowledge driven approach. *Expert Syst Appl* 40(1):96
- O'Connor PJ, Sperl-Hillen JM, Rush WA, Johnson PE, Amundson GH, Asche SE, Ekstrom HL, Gilmer TP (2011) *Ann Family Med* 9(1):12

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
