© *The Journal on Information Technology in Healthcare* 2006; **4**(4): 222–230

Automatic Creation of Decision Support Systems: Application and Results in the Cardiovascular Diseases Domain

Markos Tsipouras, Themis Exarchos, Costas Papaloukas*, Aris Bechlioulis, Anna Kotsia*, Theodora Nanou, Christos Bazios*, Yannis Antoniou*, Dimitrios Fotiadis, Aikaterinh Naka*, Lampros Michalis**

Unit of Medical Technology and Intelligent Information Systems, Department of Computer Science, and *Department of Cardiology, University of Ioannina, Greece.

ABSTRACT

In this paper we present the Decision Support Framework (DSF) of the NOESIS platform. This is a web-based, dynamic knowledge management system that aims to assist healthcare professionals in making the best possible decisions for the prevention, diagnosis and treatment of patients with cardiovascular diseases. The core of NOESIS is a set of Decision Support Systems (DSS), that are automatically generated from the Decision Support Framework. The system was initially employed to generate DSS, for the following four cardiovascular sub-domains: (a) Detection of myocardial ischaemia (b) Detection of arrhythmic episodes (c) Diagnosis of coronary artery disease (CAD), (d) Prediction of clinical restenosis in patients undergoing angioplasty. The system was tested using data from international databases and from patients treated in the Department of Cardiology at the University of Ioannina. The results demonstrated that the system was highly accurate in detecting myocardial ischemic episodes and arrhythmias (sensitivity and specificity ranging from 81.79% to 99.94%), but less accurate for predicting coronary artery disease and clinical restenosis (sensitivity and specificity ranging from 40.00% to 83.33%). The study has demonstrated the potential of the system to support healthcare professionals in medical decision-making.

INTRODUCTION

Cardiovascular diseases (CVD) are the leading cause of death in many countries worldwide. According to the World Health Organization (WHO), CVD is the cause of death of 16.6 million people around the globe each year. The multifaceted nature of the disease, combined with a wide variety of treatments and outcomes,

Correspondence and reprint requests: Markos Tsipouras, Unit of Medical Technology and Intelligent Information Systems, Department of Computer Science, University of Ioannina. Email: markos@cs.uoi.gr.

and complex relationships with other diseases e.g. diabetes, have made diagnosis and optimal treatment of cardiovascular diseases a problem for all but experienced cardiologists.

NOESIS is a dynamic knowledge management system that aims to address this problem by providing "knowledge and decision support" to clinicians, and enabling a smooth transition from established medical knowledge to personal judgment[1]. It does this by simulating the probabilistic reasoning that health professionals implicitly perform, by producing qualitative and quantitative models representing deterministic causal relationships and probabilistic associations between symptoms and diseases. To achieve this the system retrieves information from heterogenous distributed medical media databases that mediate medical decisions and prognosis, and processes these through a decision support framework (DSF) to produce information that clinicians can use to aid them in patient diagnosis and management. The system is web-based so that it can be accessed by clinicians at anytime.

The core of the system is a set of Decision Support Systems (DSSs), that are automatically generated applying the Decision Support Framework (DSF) methodology to several different cardiovascular sub-domains. In order to use the DSF in a specific domain, an initial annotated dataset is required. This dataset is obtained from several resources, distributed over the web. The quality of the diagnosis, produced by the DSS, is analogous to the quality of this dataset. The DSF has been employed to generate four DSSs, for the following cardiovascular sub-domains:

- Detection of myocardial ischaemia
- Detection of episodic arrhythmias
- Diagnosis of coronary artery disease (CAD)
- Prediction of clinical restenosis in patients undergoing angioplasty

MATERIALS AND METHODS

The DSF is a three stage methodology (Figure 1) involving [2]:

- Creation of a rule-based classifier using medical knowledge or data mining techniques
- Development of a fuzzy logic model
- Optimisation of the fuzzy logic model's parameters

In the first stage, a set of crisp rules is generated. This can be performed by two different approaches. In one the rules are provided by domain experts (knowledge-based approach). In the other they are automatically generated from an annotated database using either association rule mining, or decision tree induction. In either case, a set of rules is created, in the form of a collection of "if … then". In the second stage, the crisp set of rules is transformed to a fuzzy set of rules using a membership function instead of the crisp ones and S and T norms instead of the binary *AND* and *OR* operators. For simplicity the technical aspects and mathemati-

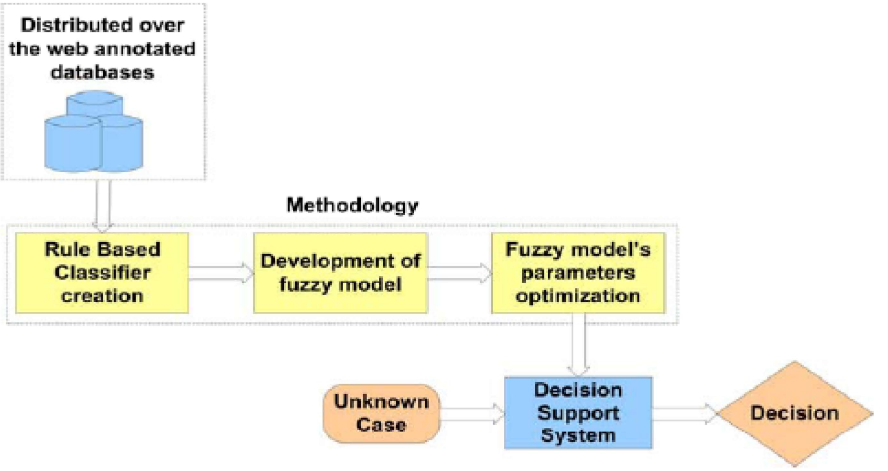

Figure 1. *The three-stage methodology used to create a Decision Support System*

cal calculations required for the fuzzy logic stages are not included in this paper. Details can be provided from the authors on request.

To apply the crisp and fuzzy logic rules, patient specific information is required for the DSS including:

- Patient symptoms
- Clinical findings
- Investigation results

RESULTS

Ischaemic Episode Detection

Myocardial ischemia is a condition where the blood supply to the heart is inadequate to meet the metabolic requirements of the heart muscle. As a result ionic shifts occur in cardiac myocytes leading to characteristic electrocardiogram (ECG) changes[3,4]. The above presented framework (using association rule mining for rule extraction) was evaluated for generating a DSS based on detecting ischaemic ECG changes. The dataset consists of 11 hours of two channel ECG recordings from the European Society of Cardiology (ESC) ST-T Database[5]. The above recordings were preprocessed in order to remove noise such as baseline wandering, alternating current (A/C) interference and movement artifacts. Five features were extracted from each cardiac beat:

- ST Segment Deviation
- ST Segment Slope
- ST Segment Area
- T Wave amplitude
- T Wave normal amplitude

Table 1. *Sensitivity and specificity results (%) for the ischaemic DSS*

	DSF (1st stage)	DSF
Sensitivity	86.37	87.80
Specificity	90.43	92.46

In addition, a sixth feature, the patient's age was employed.

The recordings resulted in 76,989 cardiac beats, from which 1,936 were used for the generation of the ischaemia DSS and 75,053 for testing it. The results, from the application of the framework employing only the first stage and then using all three stages, are presented in Table 1. After the detection of ischaemic beats, we followed our previously described approach to detect any ischaemic episodes[4].

Arrhythmic Episode Detection

An arrhythmia can be defined as an irregular single heartbeat (arrhythmic beat), or an irregular group of heartbeats (arrhythmic episode). Arrhythmias can occur in healthy hearts and be of minimal consequence, but may also indicate cardiac pathology and can cause serious problems such as stroke or cardiac arrest[6]. As a consequence automatic arrhythmia detection and classification is desirable. For the classification of cardiac arrhythmias, the initial set of crisp rules was given by medical experts and then the final two stages of the framework were applied. The only feature used was the tachogram, extracted from ECG recordings with QRS detection. A three *RR* interval sliding window was used to classify the middle *RR* interval into one of four categories:

(1) Ventricular flutter/fibrillation (VF)
(2) Premature ventricular contraction (PVC)
(3) Normal sinus rhythm (N)
(4) Second degree heart block (BII)

Records from the MIT-BIH[7] arrhythmia database were used to create and evaluate the arrhythmia DSS. The training dataset (*trainD*) was a randomly selected subset of *D*, that contained equal number of patterns from each class (250) whilst the test dataset (*testD*) consisted of the remaining patterns of *D*. 20 different pairs of *trainD* and *testD* were created. The mean values for sensitivity and specificity are presented

Table 2. *Sensitivity and specificity results (%) for the arrhythmic DSS*

	VF	PVC	N	BII
Sensitivity	99.05	81.79	95.62	98.98
Specificity	99.5	98.49	93.68	99.94

in Table 2. After the classification of the beats, we followed our previously described method, to classify the arrhythmic episodes[8].

Early Diagnosis of Coronary Artery Disease

Coronary artery disease (CAD) is the result of development of atherosclerotic plaques in the coronary arteries, resulting in narrowing of the coronary lumen and ultimately occlusion. It is the leading cause of death in western countries. The "gold standard" method for the diagnosis of CAD is coronary angiography (CA) but this is an invasive and costly procedure[9]. A non-invasive method that could accurately and reliably predict the presence of CAD would be of great clinical value. For the CAD diagnosis, the initial set of rules were generated from a decision tree. The dataset included 199 subjects suspected of having CAD and who underwent coronary angiography for the first time in the University Hospital of Ioannina. Patients with known CAD were excluded from the study. Eighty nine of the subjects had normal angiograms and in the remaining 110 subjects the presence of CAD was confirmed by two experts.

In order to characterise each subject, the 19 features shown in Table 3 were used. Family history was defined as the presence of CAD in a father or brother when aged 55 years or younger, or in a mother or sister at an age of less than 65 years. Hypertension was defined as systolic blood pressure (SBP) more than

Table 3. *Characteristics used in the Coronary artery disease DSS*

#	Feature	Units
1	Age	years
2	Sex	male(1), female(0)
3	Family History	yes(1), no(0)
4	Smoking	smoker (2), ex-smoker (1), non-smoker (0)
5	Diabetes	FBGC ≥ 126mg/dl (1) less (0)
6	Hypertension	DBP>90mmHg and/or SBP>140mmHg (1) less (0)
7	Hyperlipidemia	total cholesterol over 220mg/dl (1) less (0)
8	Creatinine	mg/dL
9	Glucose	mg/dL
10	Total Cholesterol	mg/dL
11	HDL	mg/dL
12	TRG	mg/dL
13	BMI	kg/ m2
14	Waist	cm
15	HR	bpm
16	SBP	mmHg
17	DBP	mmHg
18	PWVcf	m/sec
19	AIx	%

Key: ALx = Augmentation Index, BMI= Body mass index, bpm – beats per minute, DBP = Diastolic blood pressure, FBGC = Fasting blood glucose, HDL = High Density Lipoprotein, HR = heart rate, PWVcf = Carotid – Femoral Pulse Wave, SBP = Systolic blood pressure, TRG = Triglycerides

Table 4. *Sensitivity and specificity results (%) for the CAD DSS*

	DSF (1st stage)	DSF
Sensitivity	61.82	80.00
Specificity	53.93	65.17

140mmHg and/or diastolic blood pressure (DBP) more than 90mmHg or use of antihypertensive agents. Diabetes mellitus was defined as a fasting blood glucose concentration more than 126mg/dl or antihyperglycemic drug treatment. Current smoking was defined as having smoked a cigarette within the week preceding the angiogram. Hyperlipidemia was defined as total cholesterol over 220mg/dl or use of lipid-lowering agents (statins or fibrates). Body mass index (BMI) was calculated as weight (kg) divided by the height in metres squared (m^2). Carotid – Femoral Pulse Wave Velocity (PWVcf) and Augmentation index (AIx) were measured non-invasively using applanation tonometry, as indices of vascular stiffness. In order to confirm the presence or absence of CAD, coronary angiography was performed by the Judkins technique. All coronary angiograms were visually assessed by two experienced cardiologists to reach consensus agreement. Significant CAD was defined as at least one stenosis of 50% or greater diameter in at least one coronary artery vessel. The absence of CAD was defined as completely smooth epicardial coronary arteries.

The ten fold stratified cross validation method was used for evaluation. The procedure was applied to each fold, generating ten different crisp set of rules and fuzzy models. Both, the crisp set of rules and the final fuzzy model, have been evaluated in our dataset. Table 4 presents the average sensitivity and specificity. The overall accuracy of the crisp set of rules is 58%, while the average accuracy for the fuzzy DSS is 73%.

Prediction of Clinical Restenosis in Patients Undergoing Angioplasty

Percutaneous transluminal angioplasty at the site of a coronary stenosis is commonly used for the treatment of patients with single vessel and two vessel disease. Unfortunately restenosis occurs in a high percentage of patients within 6 months of the intervention. Restenosis is partially related to patient characteristics[10], and identifying patients at increased risk for restenosis can be potentially useful as these patients may be selected to have additional treatment, e.g. insertion of drug-eluting stents or intracoronary radiation. To try and predict clinical restenosis in patients undergoing angioplasty, the set of rules was generated from a decision tree. The dataset consisted of 679 patients that underwent angioplasty at the University Hospital of Ioannina.

In order to characterise the subjects, the 15 features shown in Table 5 were used. Family history, hypertension, diabetes mellitus, current smoking, and hyperlipidemia

Table 5. *Features used in the clinical restenosis DSS*

#	Feature	Units
1	Age	years
2	Sex	male(1), female(0)
3	Family History	yes(1), no(0)
4	Smoking	smoker (1), non-smoker (0)
5	Diabetes	FBGC ≥ 126mg/dl (1) else (0)
6	Hypertension	DBP>90mmHg and/or SBP>140mmHg (1) else (0)
7	Hyperlipidemia	total cholesterol over 220mg/dl (1) else (0)
8	CAD History	yes(1), no(0)
9	Prior PTCA	yes(1), no(0)
10	Prior CABG	yes(1), no(0)
11	Single Vessel Disease	yes(1), no(0)
12	Clinical Presentation	Unstable angina (1), Acute myocardial infarction (2), Stable angina (3)
13	Vessel Treated	Left anterior descending (1), Left circumflex (2), Right coronary artery (3), Left main (4), Bypass graft (5)
14	IIB/IIIA	yes(1), no(0)
15	Stent Type	Balloon (0), Stent (1)

Key: IIB/IIIA = Glycoprotein IIB/IIIA inhibitors (anti-platelet drugs), CABG = Coronary artery bypass grafting, CAD = Coronary artery disease, DBP = Diastolic blood pressure, FBGC = Fasting blood glucose level, PTCA = Percutaneous transluminal coronary angioplasty, SBP = Systolic blood pressure.

were defined as for the coronary artery disease DSS. Clinical presentation of CAD was classified as unstable angina, stable angina or acute myocardial infarction. The vessels treated with angioplasty were the right coronary artery, left main stem, left anterior descending artery, left circumflex artery and bypass grafts. The patients underwent either angioplasty with balloon alone or balloon followed by stenting with a non-coated metal stent. All patients were followed up for at least 12 months. The composite end point of the study was clinical restenosis manifested as cardiac death, a new non fatal myocardial infarction or a new revascularisation attempt of the stented vessel at less than 6 months after the initial angioplasty procedure.

All features that were used, except age, are binary or discrete values and for this reason, the last two stages of the framework did not show any improvement. In addition, due to the imbalanced class distribution of our dataset (clinical restenosis was observed in 158 out of 679 subjects), the notion of cost sensitive learning (CSL)

Table 6. *Sensitivity and specificity results (%) for the clinical restenosis DSS*

	DSF (1st stage)	DSF (1st stage) with CSL
Sensitivity	40.00	64.43
Specificity	83.33	60.21

was introduced during decision tree induction. Two thirds of the dataset were used for training and one third for testing. Table 6 presents the sensitivity and specificity with and without the use of CSL.

DISCUSSION

In the current study we introduced a novel methodology for automated generation of a DSS and its application to four cardiovascular sub-domains. The produced DSSs were evaluated using data taken from international databases and clinical data taken from patients treated at the University Hospital of Ionnia. The results have demonstrated the usefulness of the generated DSSs as well as confirmed the validity of the overall methodology. The arrhythmia and ischaemia DSSs demonstrated very high accuracy, with sensitivities and specificities ranging between 86 and 99 %. The highest accuracy was achieved with the web-based ECG application which was capable of diagnosing ventricular flutter/fibrillation and second degree heart block with a sensitivity and specificity of 99%.

The CAD DSS is a highly novel approach, with respect to both the technical and the medical aspects. Coronary angiography is the acknowledged “gold standard” technique for diagnosing CAD. However, since it is an invasive and costly procedure, it is not entirely suitable for mass population screening. The proposed DSS utilises relatively easily obtained patient data, e.g. pertinent patient history, routine blood tests and non-invasive assessment of vascular stiffness. The rule-based nature of the DSS makes the decision making process transparent. Furthermore, the use of coronary angiography for the initial annotation of the database is a great advantage with respect to the quality of our dataset.

The fourth DSS provides an integrated system for decision support in clinical restenosis after angioplasty. To our knowledge there are no methods in the literature for addressing this problem. This DSS system was the least accurate of the four and further work is needed to refine it. To do this it will be necessary to identify the most important factors that affect smooth muscle migration, proliferation and its modulation and implement these in the DSS.

The NOESIS platform will provide health professionals with a new powerful decision support framework, and a new, innovative way of searching information through a dynamic knowledge management system. This will reduce the time and effort spent searching for information with currently available tools. Because it is a web-based application, it is a time and place transparent innovative tool that can be used to support decision making for both primary care and hospital physicians in their daily clinical practice. Further exploitation might focus on the application of the framework to other medical domains. This can be easily implemented, due to the fully automated nature of the framework, with the definition of other findings and appropriate diagnoses, provided by experts.

ACKNOWLEDGEMENTS

This research was part funded by the program "Heraklitos" of the Operational Program for Education and Initial Vocational Training of the Hellenic Ministry of Education and by the European Commission as part of the NOESIS project (IST-2002-507960).

REFERENCES

1 NOESIS. http://www.noesis-eu.org/.

2 Exarchos TP, *et al*. A platform for wide scale integration and visual representation of medical intelligence in cardiology: the decision support framework. *Proc. 32nd CinC*, France, 2005, pp. 167–70.

3 Goldman MJ. *Principles of Clinical Electrocardiography*, 11th ed. CA: Lange Med. Pubs., 1982.

4 Papaloukas C. Fotiadis DI, Liavas AP, Likas A, Michalis LK. A knowledge-based technique for automated detection of ischemic episodes in long duration electrocardiograms. *Med. Biol. Eng. Comput*. 2001; **39**: 105–12.

5 *European Society of Cardiology European ST-T Database Directory*. Pisa: S.T.A.R. 1991.

6 Sandoe E, Sigurd B. *Arrhythmia: A Guide to Clinical Electrocardiology*. Bingen: Publishing Partners Verlags GmbH, 1991.

7 *MIT-BIH Arrhythmia Database* 3rd ed. Harvard-MIT Division of Health Sciences and Tech., 1997.

8 Tsipouras MG, Fotiadis DI, Sideris D. An arrhythmia classification system based on the RR interval signal. *Artif. Intel. Med*. 2005; **33**: 237–50.

9 Vogel RA. Coronary Risk factors, Endothelial function and arteriosclerosis: a review. *Clinical Cardiology* 1997; **20**: 426–32.

10. Welt FGP, Rogers C. Inflammation and restenosis in the stent era. *Arterioscler. Thromb. Vasc. Biol*. 2002; **22**: 1769–76.