

## COPYRIGHT NOTICE

**UB ResearchOnline**  
**<http://researchonline.ballarat.edu.au>**

This is an Author's Submitted Manuscript of an article published in Stranieri, A., Abawajy, J., Kelarev, A., Huda, S., Chowdhury, Morshed., Jelinek, H. (2013) An approach for Ewing test selection to support the clinical assessment of cardiac autonomic neuropathy, *Artificial Intelligence in Medicine*, 58(3):185-193, available online at:  
<http://dx.doi.org/10.1016/j.artmed.2013.04.007>

Copyright © Elsevier

# An Approach for Ewing Test Selection to Support the Clinical Assessment of Cardiac Autonomic Neuropathy

Andrew Stranieri<sup>a</sup>, Jemal Abawajy<sup>b</sup>, Andrei Kelarev<sup>b</sup>, Shamsul Huda<sup>a</sup>, Morshed Chowdhury<sup>b</sup>, Herbert F. Jelinek<sup>c</sup>

<sup>a</sup>*CIAO, University of Ballarat, PO Box 663, Ballarat, Victoria 3353, Australia*

<sup>b</sup>*School of Information Technology, Deakin University, Victoria 3125, Australia*

<sup>c</sup>*Centre for Research in Complex Systems, School of Community Health, Charles Sturt University, PO Box 789, Albury, NSW 2640, Australia*

---

## Abstract

This paper addresses the problem of determining optimal sequences of tests for the clinical assessment of cardiac autonomic neuropathy (CAN). Specifically, we studied the Ewing battery consisting of the five most important tests in the diagnosis of CAN, and investigated an application of decision trees and the recently proposed optimal decision path finder (ODPF) procedure for indentifying optimal sequences of tests. We present the outcomes of an exhaustive evaluation of the performance of decision trees for all 32 subsets of the Ewing battery. These outcomes are required for practical determination of such sequences using the ODPF procedure for various cost-functions including the minimization of effects when any one or more of these tests cannot be completed because of the individual difficulties faced by each patient in performing the tests. We used several feature selection methods to find the best features that can increase the predictive accuracy in situations where one of the Ewing attributes is missing, and have prepared complete outcomes of all tests required for finding the best sequences of tests including the Ewing tests combined with additional features. We applied these outcomes to determine the best sequence of the five tests belonging to the Ewing battery for the clinical assessment of CAN and cost-function equal to the number of tests, as one of the appropriate cost-functions used in this setting.

*Keywords:* optimal sequence of tests, J48, cardiac autonomic neuropathy

*2010 MSC:* 68T05, 68T10

---

## 1. Introduction

The problem of finding an optimal sequence of tests in diverse knowledge domains has been addressed by various authors including Chi, Streen and Katz for a classification or a diagnosis [6], Thompson [46] for determining the sequence of tests that maximizes the predictive accuracy of a disease diagnosis, and Oddi and Cesta [38] for scheduling tasks to manage medical resources. Artificial Intelligence (AI) methods were applied to planning and scheduling of tests for a number of diseases by Houshyar and Khayyal [23], Marinagi et al. [36] and Spyropoulos [44]. Scheduling tests is a well known topic outside medicine including vehicle fault diagnosis considered by Bartels and Zimmermann [1], and other domains considered by Dy and Brodley [12] and Liberatore and Nydick [35].

In this paper, we use data mining methods to find optimal sequences of tests for the clinical risk assessment of cardiac autonomic neuropathy (CAN), prepare outcomes required for determination of optimal sequences for individual cost-functions, and find most effective additional tests. In addition, visual representations are proposed that may simplify the use of experimental results in practice.

Data mining, as a part of knowledge discovery from databases (KDD), can be used to identify novel, valid, useful, and ultimately understandable patterns in data and has been applied extensively to data from the medical domain by Bellazzi and Ferrazzi [2], Escalante et al. [13], Gagliardi [20], Kukar et al. [32], Liang and Zhang [34], Shouman et al. [43] and Van et al. [48].

CAN is a complication of diabetes that involves damage to the autonomic nerve fibres that innervate the heart and blood vessels. The resulting abnormalities in heart rate control and vascular dynamics is thought to account for many deaths [28]. The Ewing battery of tests [15, 17, 18] are used clinically to assess a patient's risk of CAN. There are five tests in the Ewing battery: changes in heart rate associated with 1) lying to standing, 2) deep breathing, 3) attempted exhalation against a closed airway (Valsalva manoeuvre) and changes in blood pressure associated with 4) hand grip and 5) lying to standing.

Ewing and Clarke [17] recommended that all five tests be performed in a preferred sequence however this takes time and is not possible for all patients. For instance, the hand grip test may not be performed due to arthritis. The lying to standing tests often cannot be performed due to mobility challenges and some patients have conditions where forceful breathing is contra-indicated. Further, clinicians sometimes have an idiosyncratic preference for one test or sequence of tests over another, [16], [21]. Although the time to perform all five tests, at around 20-30 minutes, is not long, this is sometimes difficult to achieve in the context of a

busy practice. These issues result in CAN risk assessments being made in practice on the basis of only a subset of Ewing tests.

Chi, Streen and Katz [6] proposed and investigated an algorithm for determining optimal sequences of tests briefly illustrated in Figure 1, the Optimal Decision Path Finder (ODPF) procedure. The ODPF uses a pre-specified threshold of certainty required for the diagnosis of a disease. The first test selected is identified as the one that is most likely to lead to a threshold crossing for a diagnosis. The next test selected depends on the result of the first test. For instance, if the first test involves blood pressure which is found to be 160/90, then the second test is one that is most likely to cross the disease classification threshold given that blood pressure is 160/90.

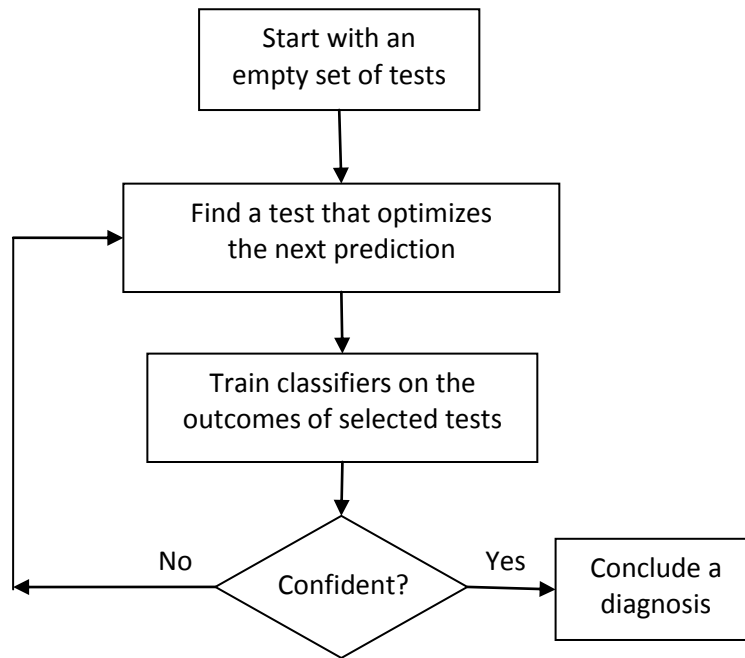


Figure 1: Optimal Decision Path Finder (ODPF) procedure

Chi, Streen and Katz [6] use a lazy support vector based classifier to determine the minimum set of features to classify an instance into disease or non-disease given the results of previous tests. The lazy support vectors are repeatedly used in an algorithm that involves calibrating the support vector machine's binary output into one that involves a 10 category certainty assessment. The algorithm takes into account the cost of each test which further complicates the application and requires that the algorithm be implemented and embedded into a decision support

system for the clinician to use during a consultation. However, this reduces a clinician's autonomy, and mandates a reliance on technological systems that may not be available in all clinical settings.

This paper explores the use of decision trees to infer optimal sequences of tests given that some tests may not be able to be performed at all. The main claim made here is that decision trees, one of the most important algorithms for clinical applications, can be effectively applied to determine the optimal boundary values, and sequence of tests. This is because decision tree outcomes are easy to generate, interpret and understand. In addition, a relatively simple visual presentation can present the sequence in a manner that can be used in practice.

In this paper, following Shouman et al. [43] and Van et al. [48], we report the performance of decision trees as one of the main data mining algorithms used in clinical applications involving heart disease. Decision tree induction implementations, advanced by Quinlan [42], are readily available. The J48 classifier implemented in Weka [22] was used in this study; however decision tree induction algorithms are also available in Rattle [49, 50] and many other sources.

Lamb et al. [33] address the issue of selecting an optimal sequence of tests to predict the risk of falls for elderly women. They use classification tree algorithms to generate a decision tree that depicts a minimum number of tests or questions to make an accurate prediction of a falls risk. Zuzek et al. [52] note that the problem of identifying an optimal sequence of tests in order to reach a diagnosis at minimum cost has been studied by numerous authors in relation to mechanical or electrical fault diagnosis.

The current paper presents outcomes of our evaluation of the performance of the decision trees for all subsets of the Ewing battery. We use these outcomes to determine the best sequence of Ewing tests for the clinical assessment of CAN. Although we assume that each Ewing test is equally costly, the determination of the optimal sequence of tests using an individual cost-function for each test can be carried out using our tables. The main benefit inherent in the use of decision trees for the identification of an optimal sequence of tests is that the decision trees are simply generated, and easily understood by clinicians.

The paper is organized as follows. The next section elaborates on the dataset and pre-processing deployed for this study. Following that, Section 3 describes measures of performance used in our experiments to evaluate decision trees. Section 4 contains the experimental results and discussions. A summary of conclusions is presented in Section 5.

## 2. Cardiac Autonomic Neuropathy

The dataset derives from the DiScRi project by Jelinek et al. [28], a diabetes complications screening programme in Australia where members of the general public participate in a comprehensive health review consisting of tests including ECG, Ewing battery, retinal scans, peripheral nerve function and blood supply assessments, for early detection and timely intervention of diabetes and cardiovascular disease. Data on over 200 variables from over two thousand attendances have been collected in recent years. The dataset has been used in data mining applications by Cornforth and Jelinek [11], Huda et al. [24] and Jelinek et al. [28], [26], [27]. The presence of CAN from DiScRi data was determined using the Ewing battery of tests.

Several expert editing rules were used to reduce the number of missing values in the database. These rules were determined during discussions with the experts maintaining the database. Preprocessing of data using these rules produced 1177 complete rows with complete values of all Ewing fields, which were used for the experimental evaluation of the performance of data mining algorithms.

The classification of disease progression associated with CAN is important, because it has implications for planning of timely treatment, which can lead to improved wellbeing of the patients and a reduction in morbidity and mortality associated with cardiac arrhythmias in diabetes. As indicated above, the most important tests required for a risk assessment of CAN rely on responses in heart rate and blood pressure to various activities, usually consisting of tests described in [15, 17, 18]. In particular, Ewing and Clarke [17] recommended the tests be performed in a specific sequence as follows:

- (1) Heart rate response to the Valsalva manoeuvre (VAHR); where the patient exhales against 40mmHg pressure while the heart rate is observed.
- (2) Heart rate variation during deep breathing (DBHR); where the patient sits quietly and breathes deeply while an electrocardiogram records the heart rate variation over 6 breathing cycles.
- (2) Blood pressure response to sustained hand-grip (HGBP); where the systolic blood pressure variation is recorded before and after a sustained hand grip.
- (4) Heart rate response to moving from a lying to a standing position (LSHR); where the beat to beat (R-R) interval change in response to standing from a lying position is measured.

- (5) Blood pressure response moving from lying to standing (LSBP); where the blood pressure change in response to standing from a lying position is measured.

Table 1 contains the boundary points for each test derived in [15, 17, 18] from physiological evidence in association with in-field trials. These boundary values are also explained by Ewing et al. [18] in great detail. The categorical variables *Abnormal*, *Borderline* and *Normal* are introduced in the Ewing and Clark formulation for each test. The boundary points illustrated in Table 1 may not necessarily be the optimal boundary points for distinguishing categories of CAN risk when a subset of Ewing tests are used. New boundary points identified by decision trees can be used to maximize the predictive accuracy of CAN assessment.

Test	Value		
	Normal	Borderline	Abnormal
VAHR	$\geq 1.21$	1.11-1.20	$\leq 1.10$
DBHR	$\geq 15$	11-14	$\leq 10$
HGBP	$\geq 16$	11-15	$\leq 10$
LSHR	$\geq 1.04$	1.01-1.03	$\leq 1.00$
LSBP	$\leq 10$	11-29	$\geq 30$

Table 1: Ranges and boundary values determining categorical variables for the Ewing battery

DiScRi database contains a separate attribute LSBPneg that can take on values FALSE and TRUE. If LSBPneg = TRUE and  $LSBP \geq 30$ , then the result is abnormal. If LSBPneg = TRUE and  $29 \geq LSBP \geq 11$ , then the result is borderline. In all other cases the result of this test is normal.

Before applying the cut-offs to DiScRi data for DBHR, LSBP, HGBP we round off fractional parts to the second decimal place. This means that we are a little more conservative. Likewise, for VAHR, LSHR we discard the third and further digits after the decimal point.

Ewing et al. [15, 18] defined the five classes for a CAN risk assessment given in Table 2. Ewing et al. [18] considered alternative approaches to the classification of CAN and compared the categorization given in Table 2 with two scoring systems used by other researchers: (1) giving 0 for a normal test, for a borderline result, and 1 for abnormal result, thus giving a score of 0-5 for each subject; and

(2) the number of tests definitely abnormal, again giving a score of 0-5 for each subject. Ewing et al. [18] demonstrate that these scoring systems give roughly equivalent categorizations and seem to carry no real advantages.

Class	Test values
Normal	All tests normal or one borderline.
Early	One of the three heart rate tests abnormal or two borderline
Definite	Two or more of the heart rate tests abnormal.
Severe	Two or more of the heart rate tests abnormal plus one or both of the blood pressure tests abnormal, or both borderline.
Atypical	Any other combination of abnormal tests.

Table 2: CAN classes defined by Ewing et al. [18]

Since there are very few atypical patients in the DiScRi database, we investigated three original classifications of cardiac autonomic neuropathy progression introduced by Ewing et al. [15, 18]. They have 2, 3 and 4 classes, respectively. The first one divides all patients into two classes allocating each patient either to the *Normal* class, or to *Definite* class. The second one divides all patients into three classes allocating each patient to one of the following classes: *Normal*, *Early*, *Definite*. The fourth classification divides all patients into four classes, allocated each patient to one of the following classes: *Normal*, *Early*, *Definite*, and *Severe*.

Ewing [14] writes that the question of finding an optimal sequence of tests remains a difficult open question that requires further investigation. This is also confirmed by the more recent work of Chen et al. [4], Jelinek et al. [27, 26], Khandoker et al. [30] and Stella et al. [45].

### 3. Measures of Performance for Decision Trees

Following Shouman et al. [43], we used accuracy in assessing the decision trees as the main measure of performance essential for guiding the clinicians in determining the best sequence of tests for a particular patient. It is related to two other measures, sensitivity and specificity, also considered for example



by Shouman et al. [43] and Yearwood et al. [51] and broadly used in clinical practice. Here we include only a brief overview of these measures. The *accuracy* is defined for the whole classifier as the percentage of all patients classified correctly, which means that this definition does not involve weighted averages in the calculation. The accuracy can be expressed as the probability that the prediction of the classifier for an individual patient is correct. *Sensitivity* is the proportion of positives (eg. patients with CAN) that are identified correctly. Sensitivity is also called *True Positive Rate*. *Specificity* is the proportion of negatives (eg. patients without CAN), which are identified correctly. Specificity is equal to  $1 - \text{False Positive Rate}$ . Sensitivity and specificity are measures evaluating binary classifications. For multi-class classifications they can be also used with respect to one class and its complement. Following Shouman et al. [43], these measures can be expressed using formulas:

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Positive} + \text{Negative}} \quad (1)$$

$$\text{Sensitivity} = \text{True Positive/Positive} \quad (2)$$

$$\text{Specificity} = \text{True Negative/Negative} \quad (3)$$

These measures are related to recall and precision. *Precision* of a classifier, for a given class, is the ratio of true positives to combined true and false positives. *Recall* is the ratio of true positives to the number of all positive samples (i.e., to the combined true positives and false negatives). The recall calculated for the class of patients with CAN is equal to sensitivity of the whole classifier.

#### 4. Experimental Results and Discussion

In this paper we used 10 fold cross validation to assess the performance of J48 trained for various subsets of Ewing features. Figures 2 to 5 present the outcomes of an exhaustive evaluation of performance of decision trees for all 32 subsets of the set of features that is recognized as the most important set of features in the diagnosis of CAN and is called the Ewing battery.

These outcomes are required for practical determination of such sequences following the ODPF methodology to minimize individual difficulties faced by each patient in performing the tests. We use these outcomes to determine the best sequence of Ewing tests for the clinical assessment of CAN and cost-function equal to the number of tests, see Table 4. Figure 6 illustrates the optimal sequence of Ewing tests and predictive accuracies that can be achieved after each step for

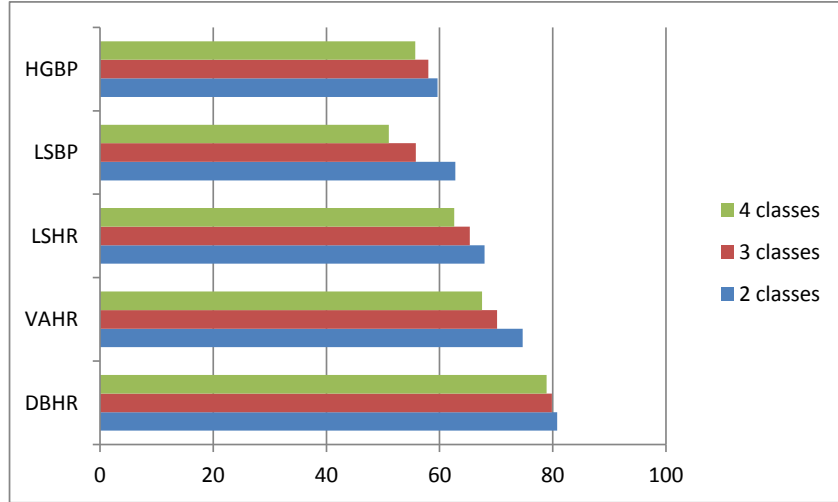


Figure 2: Accuracy of J48 for sets of one Ewing feature

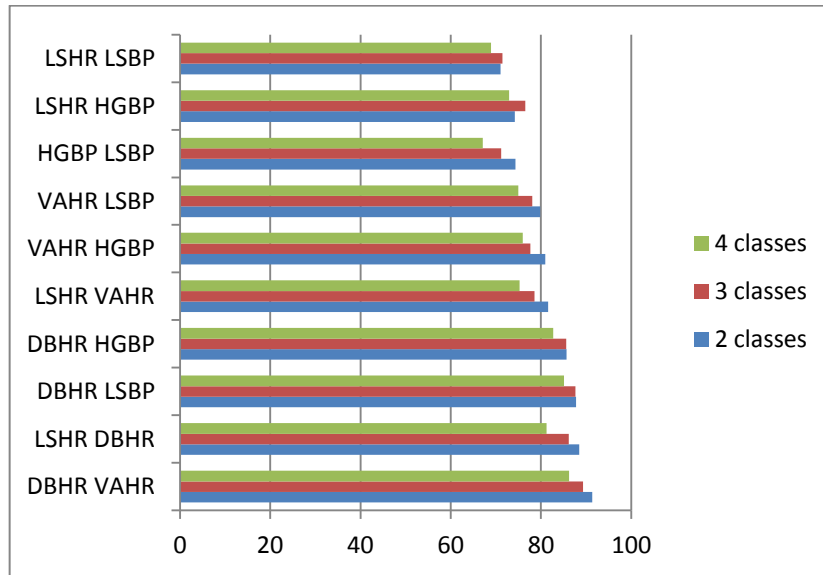


Figure 3: Accuracy of J48 for sets of two Ewing features

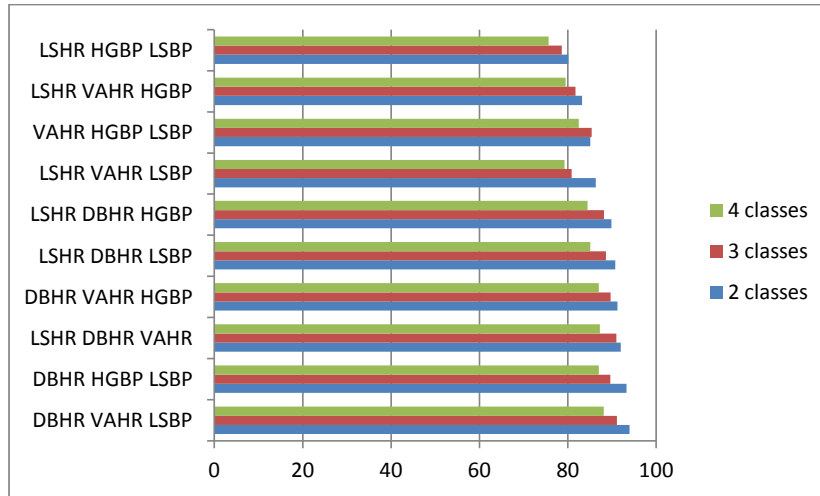


Figure 4: Accuracy of J48 for sets of three Ewing features

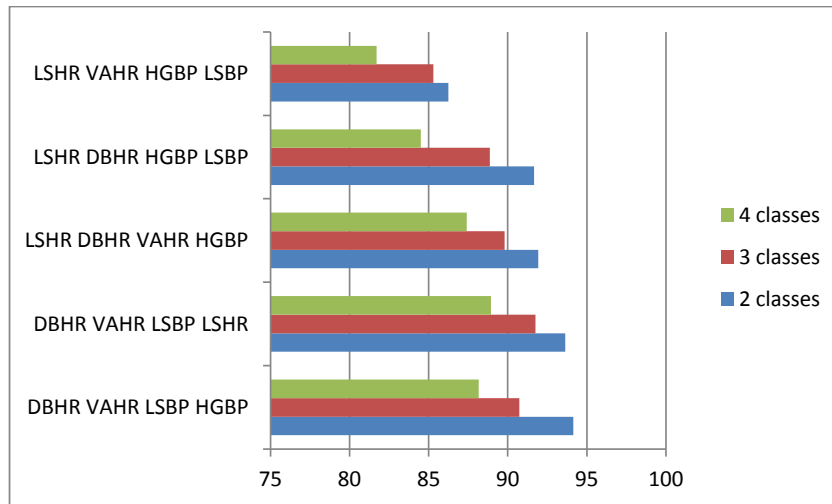


Figure 5: Accuracy of J48 for sets of four Ewing features

2 classes of CAN and the number of tests as cost-function, since the Ewing tests have approximately equal financial costs. On the other hand, another important objective of a clinician is to minimize difficulties that individual patients may experience ... in determining the presence of CAN and so clinicians may have to use other subsets of tests. Therefore there are alternative cost-function to be taken into account on individual basis.

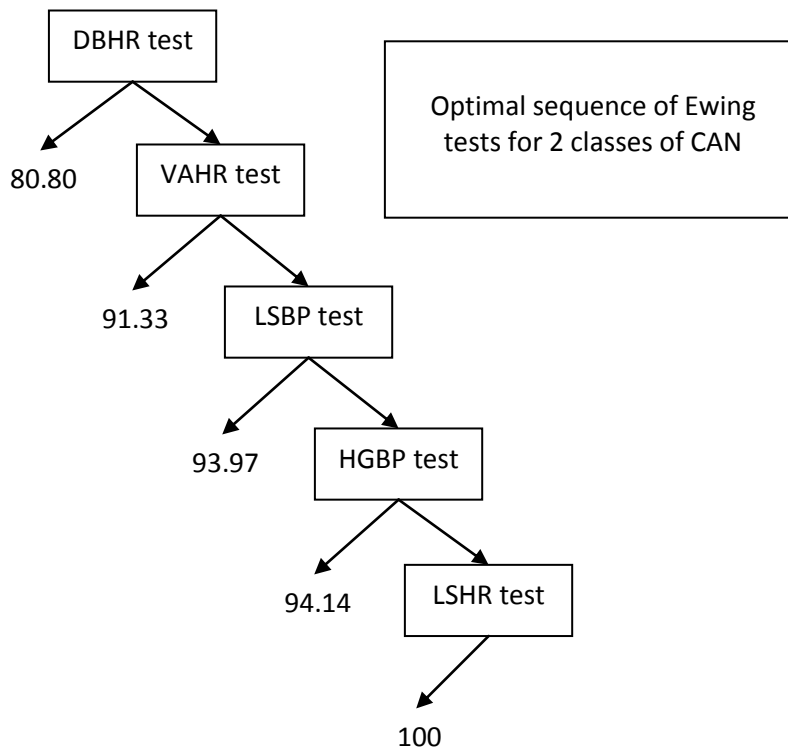


Figure 6: Optimal sequence of Ewing tests and predictive accuracies after each step for 2 classes of CAN and the number of tests as cost-function

Table 3 illustrates the confusion matrix for the J48 classifier using a single Ewing test, the DBHR. The table illustrates that  $159/716 = 22\%$  of patients classified as CAN using the single measure of DBHR are likely to be normal. This error rate may be too high for most diagnostic contexts. However, it may be sufficient for lifestyle based CAN mitigation strategies. The DBHR proved to be the most accurate single test, and so is the test recommended to be performed first.

During the application of ODPF procedure explained in Figure 1, a clinician could follow the resulting refinement of the protocol choosing tests so that every

Classes	Predictions	
	Normal	Definite
Normal	394	67
Definite	159	557

Table 3: Confusion matrix of J48 classifier with 2 CAN classes for the DBHR test

next test provides the greatest predictive accuracy for the resulting sequence according to the outcomes obtained in Figures 2 to 5. These figures can be used to inform clinicians of the best remaining tests to perform and the ensuing predictive accuracies that can be achieved. Therefore clinicians have to make individual assessment of difficulties associated with each case on individual basis.

Rather than embedding an optimal sequence of tests algorithm into a decision support system, we advocate the visualization of all possible Ewing test sequences in a diagram that depicts the accuracy gains in diverse test sequences so that a clinician can easily select the sequence of preference and be informed of the accuracy associated with the chosen sequence. Visualization is one of the most important methods in assisting clinical planning. Information visualization for medical applications has been considered, for example, by Chi, Streen and Katz [8], Combi and Oliboni [10] and Jelinek et al. [25]. It has been applied to the design of plans by Kosara and Miksch [31]. Graph-based approaches were used for medical visualization, for instance, by Plaza et al. [41]. Hierarchical visualization layouts were considered by Tsay et al. [47]. Visualization for inference has been also investigated by Park et al. [39] and Park and Huh [40]. For diabetes patients, it was treated by Cho et al. [9].

We include Figures 7 and 8 representing compressed versions of two diagrams that illustrate visual aids and can be created to include the predictive accuracies of all test sequences facilitating the work of clinicians applying the ODPF procedure in practice. Complete versions for the use of practitioner would include the prediction accuracies achieved at each step as well as average time required to perform the next test. The advantage of the circle diagram in Figure 7 is that at every step of the ODPF process the current cell in the diagram keeps track not only of the final predictive accuracy achieved, but also of the whole sequence of previous tests in the order they were applied.

When it is difficult for a patient to pass one of the standard Ewing battery tests, it may be possible to use the remaining tests to increase the combined predictive

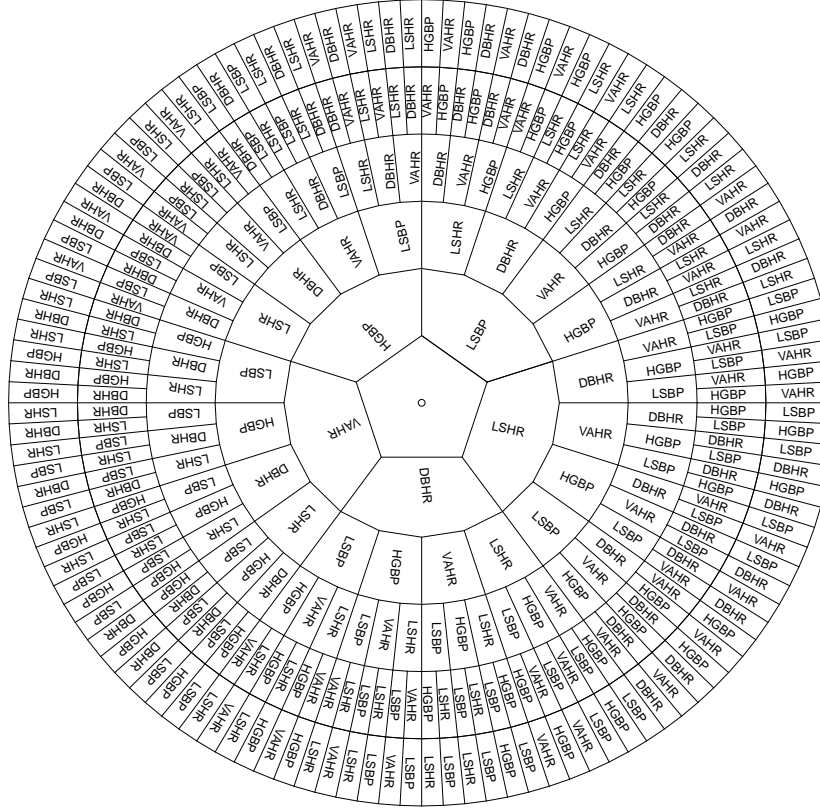


Figure 7: Circle diagram

accuracy of classification. Determination of appropriate tests to be used for this has already been considered in the literature, see [37].

We used several feature selection methods to find a few most effective tests that can be combined with tests in Ewing battery. To rank features in the order of their significance we used three methods: Gain Ratio Attribute Evaluation, Information Gain Attribute Evaluation and OneR Attribute Evaluation. Gain Ratio Attribute Evaluation assesses the significance of each attribute by calculating its gain ratio using the formula

$$\text{GainR}(\text{Class}, \text{Attribute}) = (H(\text{Class}) - H(\text{Class}|\text{Attribute})) / H(\text{Attribute}). \quad (4)$$

Information Gain Attribute Evaluation assesses the significance of each attribute by calculating the information gain using the formula

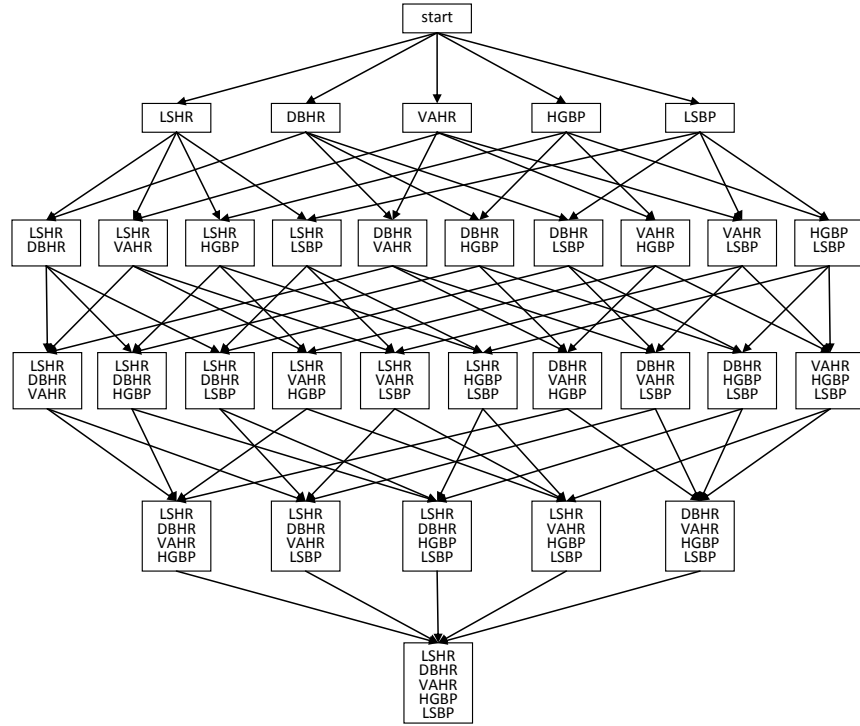


Figure 8: Lattice diagram

$$\text{InfoGain}(\text{Class}, \text{Attribute}) = H(\text{Class}) - H(\text{Class}|\text{Attribute}). \quad (5)$$

Classifier Attribute Evaluation assesses the significance of each attribute by applying it with a user-specified classifier. We used Classifier Attribute Evaluation with J48 classifier. Then we ordered all attributes according to the sum of their ranks in these three assessments. Three most significant features on this list are ECG interpretation, Grade 10sec and QRS 10sec. Note that the use of ECG data in applications of AI methods has been considered recently, for example, by Chao et al. [3], Chen and Yu [5], Chiarugi et al. [7], Gacek and Pedrycz [19] and Jovic and Bogunovic [29]. Further tests have shown that J48 classifier could not use ECG interpretation efficiently, since it is a categorical variable with very large range of values and J48 would have to construct a very large tree to handle it correctly. We carried out a complete evaluation of the predictive accuracy of J48 classifier for all 32 subsets of the Ewing battery supplemented with the QRS 10sec and Grade 10sec attribute added. Bar Charts 9 through to 12 include experimental results of these tests.

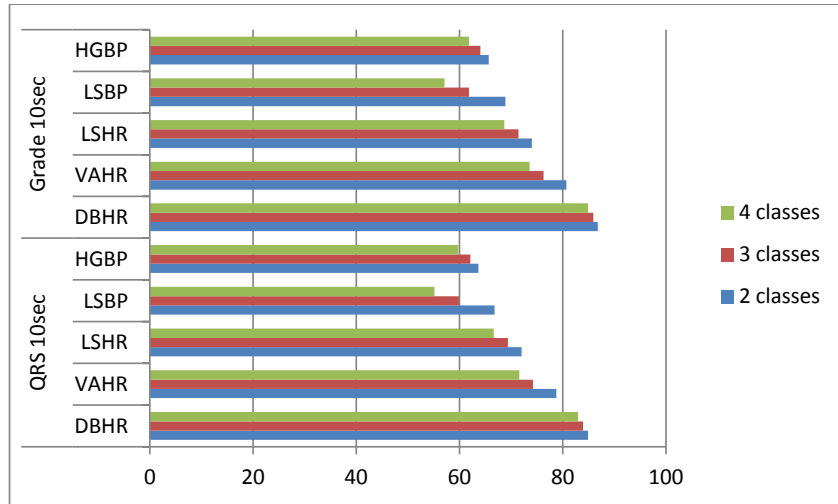


Figure 9: Accuracy of J48 for sets of one Ewing feature with QRS 10sec or Grade 10sec added

	Optimal sequences of Ewing tests
2 classes	DBHR, VAHR, LSBP, HGBP, LSHR
3 classes	DBHR, VAHR, LSBP, LSHR, HGBP
4 classes	DBHR, VAHR, LSBP, LSHR, HGBP

Table 4: Optimal sequences of Ewing tests for the number of tests as cost-function

These outcomes show that Grade 10sec and QRS 10sec produce approximately equivalent improvement in the predictive accuracy of J48, with Grade 10sec slightly better than QRS 10sec. Bar Charts 9 through to 12 can be used to determine the best sequence of Ewing tests, for example, for those patients who already have ECG interpretations and the values of QRS 10sec attribute determined by the clinicians.

## 5. Conclusion

We have applied decision trees to the problem of supporting clinicians in finding optimal sequences of tests for each individual patient for the assessment of cardiac autonomic neuropathy. Our experimental results include the outcomes of a complete evaluation of performance of decision trees for all 32 subsets of



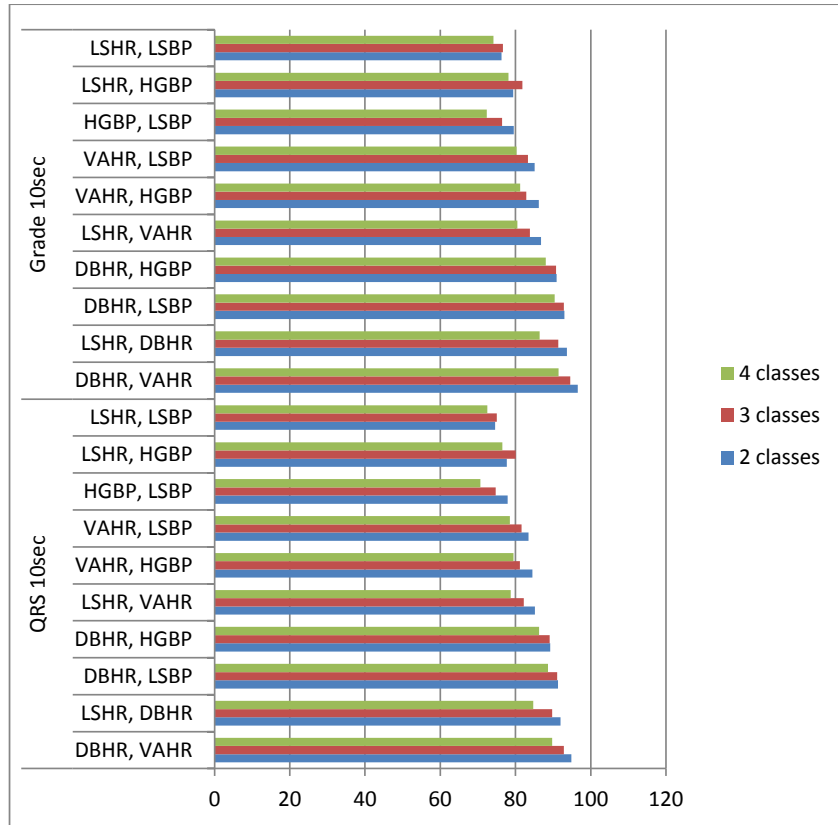


Figure 10: Accuracy of J48 for sets of two Ewing features with QRS 10sec or Grade 10sec added

the Ewing battery as required for practical determination of such sequences using the Optimal Decision Path Finder (ODPF) procedure to minimize individual difficulties faced by each patient in performing the tests. To illustrate this, we have determined the best sequence of Ewing tests for the cost-function equal to the number of tests as one of the appropriate cost-functions used in this setting.

Our experiments show that Grade 10sec and QRS 10sec produce approximately equivalent improvement in the predictive accuracy of J48, with Grade 10sec slightly better than QRS 10sec.

## 6. Acknowledgements

This research was supported by a Deakin-Ballarat collaboration grant.

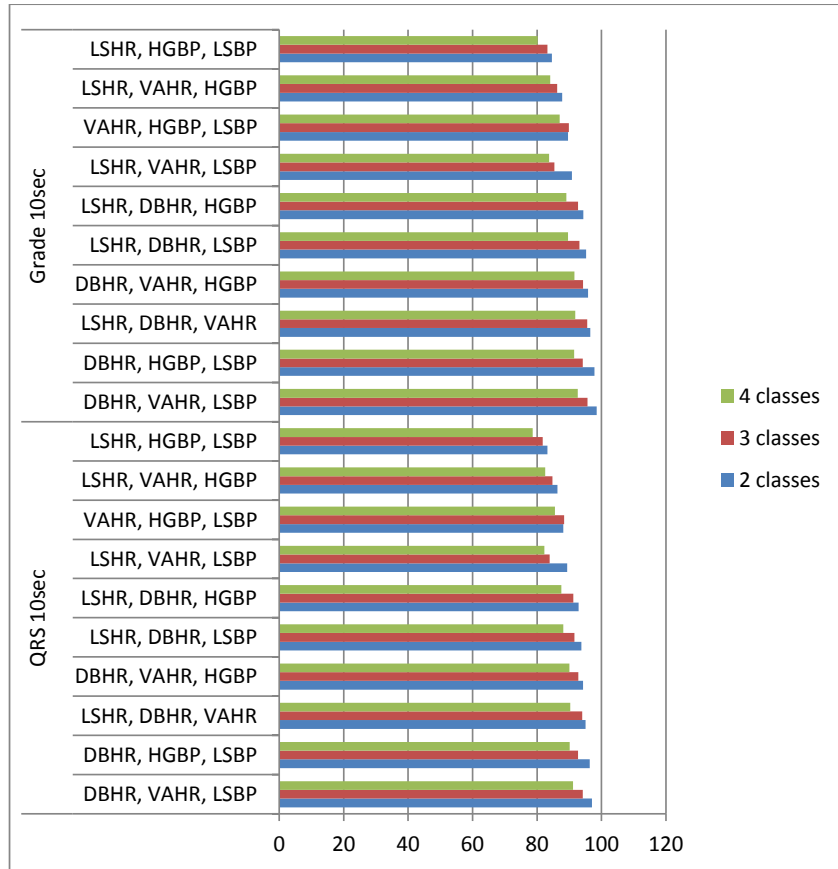


Figure 11: Accuracy of J48 for sets of three Ewing features with QRS 10sec or Grade 10sec added

## References

- [1] Bartels, J.H., Zimmermann, J., 2009. Scheduling tests in automotive R&D projects. *European Journal of Operational Research* 193, 805–819.
- [2] Bellazzi, R., Ferrazzi, F., 2011. Predictive data mining in clinical medicine: a focus on selected methods and applications. *Data Mining and Knowledge Discovery* 1, 416–430.
- [3] Chao, P.K., Wang, C.L., Chan, H.L., 2012. An intelligent classifier for prognosis of cardiac resynchronization therapy based on speckle-tracking echocardiograms. *Artificial Intelligence in Medicine* 54, 181–188.

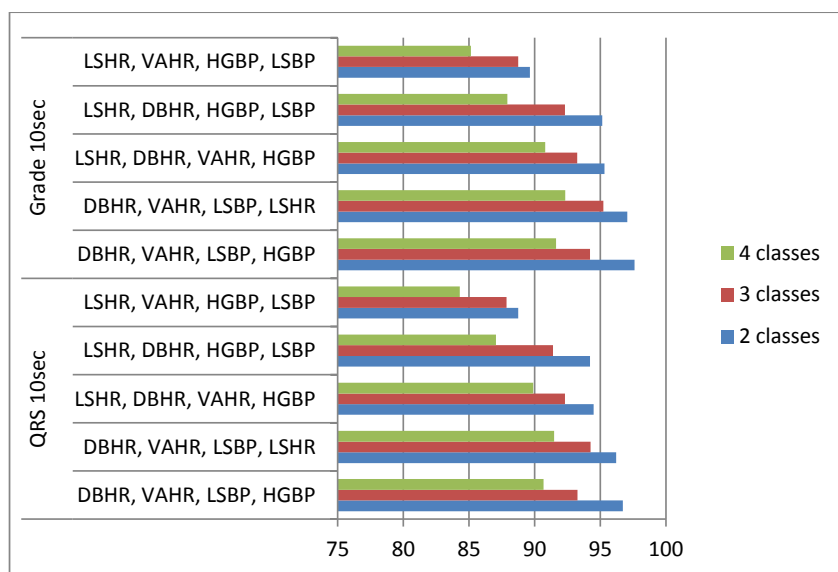


Figure 12: Accuracy of J48 for sets of four Ewing features with QRS 10sec or Grade 10sec added

- [4] Chen, H., Lin, H., Wonb, J., Lee, C., Wu, S., Lin, J., Juan, L., Ho, L., Tang, K., 2008. Cardiovascular autonomic neuropathy, autonomic symptoms and diabetic complications in 674 type 2 diabetes. *Diabetes Research and Clinical Practice* 82, 282–290.
- [5] Chen, Y.H., Yu, S.N., 2012. Selection of effective features for ECG beat recognition based on nonlinear correlations. *Artificial Intelligence in Medicine* 54, 43–52.
- [6] Chi, C., Street, N., Katz, C., 2010. A decision support system for cost-effective diagnosis. *Artificial Intelligence in Medicine* 50, 149–161.
- [7] Chiarugi, F., Colantonio, S., Emmanouilidou, D., Martinelli, M., Moroni, D., Salvetti, O., 2010. Decision support in heart failure through processing of electro- and echocardiograms. *Artificial Intelligence in Medicine* 50, 95–104.
- [8] Chittaro, L., 2001. Information visualization and its application to medicine. *Artificial Intelligence in Medicine* 22, 81–88.
- [9] Cho, B., Yu, H., Kim, K.W., Kim, T., Kim, I., Kim, S., 2008. Application of irregular and unbalanced data to predict diabetic nephropathy using visu-

- alization and feature selection methods. *Artificial Intelligence in Medicine* 42, 37–53.
- [10] Combi, C., Oliboni, B., 2012. Visually defining and querying consistent multi-granular clinical temporal abstractions. *Artificial Intelligence in Medicine* 54, 75–101.
- [11] Cornforth, D., Jelinek, H., 2007. Automated classification reveals morphological factors associated with dementia. *Applied Soft Computing* 8, 182–190.
- [12] Dy, J., Brodley, C., 2000. Feature subset selection and order identification for unsupervised learning, in: *Proc. 17th International Conf. Machine Learning*, pp. 247–254.
- [13] Escalante, H., Montes-y Gomez, M., Gonzalez, J., Gomez-Gil, P., Altamirano, L., Reyes, C., Reta, C., Rosales, A., 2012. Acute leukemia classification by ensemble particle swarm model selection. *Artificial Intelligence in Medicine* 55, 163–175.
- [14] Ewing, D., 1990. Which battery of cardiovascular autonomic function tests? *Diabetologia* 33, 180–181.
- [15] Ewing, D., Campbell, J., Clarke, B., 1980. The natural history of diabetic autonomic neuropathy. *Q. J. Med.* 49, 95–100.
- [16] Ewing, D., Campbell, J., Murray, A., Neilson, J., Clarke, B., 1978. Immediate heart-rate response to standing: simple test for autonomic neuropathy in diabetes. *British Medical Journal* 1, 145–147.
- [17] Ewing, D., Clarke, B., 1982. Diagnosis and management of diabetic autonomic neuropathy. *British Medical Journal* 285, 916–918.
- [18] Ewing, D., Martyn, C., Young, R., Clarke, B., 1985. The value of cardiovascular autonomic function tests: 10 years experience in diabetes. *Diabetes Care* 8, 491–498.
- [19] Gacek, A., Pedrycz, W., 2012. A characterization of electrocardiogram signals through optimal allocation of information granularity. *Artificial Intelligence in Medicine* 54, 125–134.

- [20] Gagliardi, F., 2011. Instance-based classifiers applied to medical databases: Diagnosis and knowledge extraction. *Artificial Intelligence in Medicine* 52, 123–139.
- [21] Gonzalez-Clemente, J.M., Vilardell, C., Broch, M., Megia, A., Caixas, A., Gimenez-Palop, C., Richart, O., Simon, I., Martinez-Riquelme, A., Arroyo, J., Mauricio, D., Vendrell, J., 2007. Lower heart rate variability is associated with higher plasma concentrations of IL-6 in type 1 diabetes. *European Journal of Endocrinology* 157, 31–38.
- [22] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I., 2009. The WEKA data mining software: an update. *SIGKDD Explorations* 11, 10–18.
- [23] Houshyar, A., Khayyal, F., 1990. A mathematical model for scheduling screening tests for progressive diseases. *Socio-Economic Planning Sciences* 24, 187–197.
- [24] Huda, S., Jelinek, H., Ray, B., Stranieri, A., Yearwood, J., 2010. Exploring novel features and decision rules to identify cardiovascular autonomic neuropathy using a hybrid of wrapper-filter based feature selection, in: *Sixth International Conference on Intelligent Sensors, Sensor Networks and Information Processing, ISSNIP 2010*, pp. 297–302.
- [25] Jelinek, H., Cornforth, D., Blackmore, K., 2011a. Visualisation in biomedicine as a means of data evaluation. *Journal of Visualization* 14, 353–359.
- [26] Jelinek, H., Khandoker, A., Palaniswami, M., McDonald, S., 2010. Heart rate variability and QT dispersion in a cohort of diabetes patients. *Computing in Cardiology* 37, 613–616.
- [27] Jelinek, H., Rocha, A., Carvalho, T., Goldenstein, S., Wainer, J., 2011b. Machine learning and pattern classification in identification of indigenous retinal pathology, in: *Proceedings IEEE Conference Eng. Med. Biol. Soc.*, pp. 5951–5954.
- [28] Jelinek, H., Wilding, C., Tinley, P., 2006. An innovative multi-disciplinary diabetes complications screening programme in a rural community: A description and preliminary results of the screening. *Australian Journal of Primary Health* 12, 14–20.

- [29] Jovic, A., Bogunovic, N., 2011. Electrocardiogram analysis using a combination of statistical, geometric, and nonlinear heart rate variability features. *Artificial Intelligence in Medicine* 51, 175–186.
- [30] Khandoker, A., Jelinek, H., Palaniswami, M., 2009. Identifying diabetic patients with cardiac autonomic neuropathy by heart rate complexity analysis. *BioMedical Engineering OnLine* 8, <http://www.biomedical-engineering-online.com/content/8/1/3>.
- [31] Kosara, R., Miksch, S., 2001. Metaphors of movement: a visualization and user interface for time-oriented, skeletal plans. *Artificial Intelligence in Medicine* 22, 111–131.
- [32] Kukar, M., Kononenko, I., Groselj, C., 2011. Modern parameterization and explanation techniques in diagnostic decision support system: A case study in diagnostics of coronary artery disease. *Artificial Intelligence in Medicine* 52, 77–90.
- [33] Lamb, S., McCabe, C., Becker, C., Fried, L., Guralnik, J., 2008. The optimal sequence and selection of screening test items to predict fall risk in older disabled women: The women’s health and aging study. *The Journal of Gerontology, Series A* 63, 1082–1088.
- [34] Liang, G., Zhang, C., 2011. Empirical study of bagging predictors on medical data, in: Vamplew, P., Stranieri, A., Ong, K.L., Christen, P., Kennedy, P.J. (Eds.), *Australasian Data Mining Conference, AusDM 2011*, ACS, Ballarat, Australia. pp. 31–40.
- [35] Liberatore, M., Nydick, R., 2008. The analytic hierarchy process in medical and health care decision making: A literature review. *European Journal of Operational Research* 189, 194–207.
- [36] Marinagi, C., Spyropoulos, C., Papatheodorou, C., Kokkotos, S., 2000. Continual planning and scheduling for managing patient tests in hospital laboratories. *Artificial Intelligence in Medicine* 20, 139–154.
- [37] May, O., Arildsen, H., 2000. Assessing cardiovascular autonomic neuropathy in diabetes mellitus: How many tests to use? *Journal of Diabetes and Its Complications* 14, 7–12.

- [38] Oddi, A., Cesta, A., 2000. Toward interactive scheduling systems for managing medical resources. *Artificial Intelligence in Medicine* 20, 113–138.
- [39] Park, C., Godtliebsen, F., Taqqu, M., Stoev, S., Marron, J., 2007. Visualization and inference based on wavelet coefficients, sizer and sinos. *Computational Statistics & Data Analysis* 51, 5994–6012.
- [40] Park, C., Huh, J., 2012. Statistical inference and visualization in scale-space using local likelihood. *Computational Statistics & Data Analysis* , in print.
- [41] Plaza, L., Diaz, A., Gervas, P., 2011. A semantic graph-based approach to biomedical summarisation. *Artificial Intelligence in Medicine* 53, 1–14.
- [42] Quinlan, R., 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA.
- [43] Shouman, M., Turner, T., Stocker, R., 2011. Using decision tree for diagnosing heart disease patients, in: Vamplew, P., Stranieri, A., Ong, K.L., Christen, P., Kennedy, P. (Eds.), *Australasian Data Mining Conference, AusDM 11*, ACS, Ballarat, Australia. pp. 23–30.
- [44] Spyropoulos, C., 2000. AI planning and scheduling in the medical hospital environment. *Artificial Intelligence in Medicine* 20, 101–111.
- [45] Stella, P., Ellis, D., Maser, R., Orcharda, T., 2000. Cardiac autonomic neuropathy (expiration and inspiration ratio) in type 1 diabetes incidence and predictors. *Journal of Diabetes and Its Complications* 14, 1–6.
- [46] Thompson, M., 2003. Assessing the diagnostic accuracy of a sequence of tests. *Biostatistics* 4, 341–351.
- [47] Tsay, J.J., Wu, B.L., Jeng, Y.S., 2010. Hierarchically organized layout for visualization of biochemical pathways. *Artificial Intelligence in Medicine* 48, 107–117.
- [48] Van, A., Gay, V., Kennedy, P., Barin, E., Leijdekkers, P., 2011. Understanding risk factors in cardiac rehabilitation patients with random forests and decision trees, in: Vamplew, P., Stranieri, A., Ong, K.L., Christen, P., Kennedy, P. (Eds.), *Australasian Data Mining Conference, AusDM 2011*, ACS, Ballarat, Australia. pp. 11–22.

- [49] Williams, G., 2009. Rattle: a data mining GUI for R. *The R Journal* 1, 45–55.
- [50] Williams, G., 2011. *Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery (Use R!)*. Springer, New York, Dordrecht, Heidelberg, London.
- [51] Yearwood, J., Kang, B., Kelarev, A., 2008. Experimental investigation of classification algorithms for ITS dataset, in: *Pacific Rim Knowledge Acquisition Workshop, PKAW 2008, Hanoi, Vietnam, 15–16 December 2008*. pp. 262–272.
- [52] Zuzek, A., Biasizzo, A., Novak, F., 2000. Sequential diagnosis tool. *Microprocessors and Microsystems* 24, 191–197.