An approach for Ewing test selection to support the clinical assessment of cardiac autonomic neuropathy

Andrew Stranieri\(^a\), Jemal Abawajy\(^b\), Andrei Kelarev\(^b\),*, Shamsul Huda\(^a\), Morshed Chowdhury\(^b\), Herbert F. Jelinek\(^c\)

\(^a\) School of Science, Information Technology and Engineering, University of Ballarat, P.O. Box 663, Ballarat, Victoria 3353, Australia
\(^b\) School of Information Technology, Deakin University, 221 Burwood Highway, Victoria 3125, Australia
\(^c\) School of Community Health, Charles Sturt University, P.O. Box 789, Albury, New South Wales 2640, Australia

**Abstract**

Objective: This article addresses the problem of determining optimal sequences of tests for the clinical assessment of cardiac autonomic neuropathy (CAN). We investigate the accuracy of using only one of the recommended Ewing tests to classify CAN and the additional accuracy obtained by adding the remaining tests of the Ewing battery. This is important as not all five Ewing tests can always be applied in each situation in practice.

Methods and material: We used new and unique database of the diabetes screening research initiative project, which is more than ten times larger than the data set used by Ewing in his original investigation of CAN. We utilized decision trees and the optimal decision path finder (ODPF) procedure for identifying optimal sequences of tests.

Results: We present experimental results on the accuracy of using each one of the recommended Ewing tests to classify CAN and the additional accuracy that can be achieved by adding the remaining tests of the Ewing battery. We found the best sequences of tests for cost-function equal to the number of tests. The accuracies achieved by the initial segments of the optimal sequences for 2, 3 and 4 categories of CAN are 80.80, 91.33, 93.97 and 94.14, and respectively, 79.86, 89.29, 91.16 and 91.76, and 78.90, 86.21, 88.15 and 88.93. They show significant improvement compared to the sequence considered previously in the literature and the mathematical expectations of the accuracies of a random sequence of tests. The complete outcomes obtained for all subsets of the Ewing features are required for determining optimal sequences of tests for any cost-function with the use of the ODPF procedure. We have also found two most significant additional features that can increase the accuracy when some of the Ewing attributes cannot be obtained.

Conclusions: The outcomes obtained can be used to determine the optimal sequences of tests for each individual cost-function by following the ODPF procedure. The results show that the best single Ewing test for diagnosing CAN is the deep breathing heart rate variation test. Optimal sequences found for the cost-function equal to the number of tests guarantee that the best accuracy is achieved after any number of tests and provide an improvement in comparison with the previous ordering of tests or a random sequence.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The problem of finding an optimal sequence of tests in diverse knowledge domains has been addressed by various authors including Chi et al. [1] for a classification or a diagnosis, Thompson [2] for determining the sequence of tests that maximizes the predictive accuracy of a disease diagnosis, and Oddi and Cesta [3] for scheduling tasks to manage medical resources. Artificial intelligence (AI) methods were applied to planning and scheduling of tests for a number of diseases in [4–6]. Scheduling tests is a well known topic outside medicine including vehicle fault diagnosis [7] and other domains [8].

This is the first article devoted to a systematic investigation of sequences of the Ewing tests required for the identification of cardiac autonomic neuropathy (CAN). We use data mining methods to find optimal sequences of tests for the clinical risk assessment...
of CAN, to achieve the predictive ability of the tests as high as possible and as quickly as possible. We find the most effective additional tests to supplement the Ewing battery for increasing the predictive accuracy in situations where some Ewing tests cannot be performed. Since not all five tests recommended by Ewing can always be applied in each situation in practice, we determine the accuracy of using each one of the recommended Ewing tests to classify CAN and the additional accuracy obtained by adding the remaining tests recommended by Ewing and even additional supplementary tests. Here complete results are obtained for all subsets of the Ewing features. These results are required for determining the optimal sequences of tests for any individual cost-function. In addition, visual representations are proposed that may simplify the use of experimental results in practice. The outcomes of our experiments can be used, for any given cost-function of performing the tests, to determine the sequences of tests such that the predictive ability of every initial segment of the sequence corresponding to each fixed value of the cost-function is the best possible.

Data mining, as a part of knowledge discovery from databases, can be used to identify novel, valid and useful patterns in data and has been applied extensively to data from the medical domain, for example, in [9–15].

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 [16]. The Ewing battery of tests [17–19] 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 [18] recommended to perform all five tests for the diagnosis of CAN, considered only one sequence of tests and did not provide any data on comparison of this sequence with other possible sequences noting only that in general the question of finding the best sequence is a difficult one. However, the tests take time and not all of them are possible to complete for every patient. 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 [20,21]. Ewing [22] 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 [23–27].

Although the time to perform all five tests, at around 20–30 min, 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. This is why it is important to obtain data on performance of the decision trees for various subsets of the Ewing features. This problem is solved completely in the present article: our experiments assess the performance of decision trees for all subsets of the Ewing battery of tests.

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 is devoted to the methods investigated in the present paper. Section 4 contains the experimental results and discussions. A summary of conclusions is presented in Section 5.

2. Cardiac autonomic neuropathy

We used a new and unique database of the diabetes screening research initiative (DiScri) project [16], which is more than ten times larger than the data set used by Ewing in his original investigation of CAN. DiScri is a diabetes complications screening programme in Australia where members of the general public participate in a comprehensive health review consisting of tests including an electrocardiogram (ECG), Ewing battery, retinal scans, peripheral nerve function and blood supply assessments, for early detection and timely intervention of diabetes and cardio-vascular 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 in [16,23,24,28,29]. 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. Pre-processing 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 [17–19]. In particular, Ewing and Clarke [18] 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 40 mmHg 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.
3. 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 [17–19] from physiological evidence in association with in-field trials. These boundary values are also explained by Ewing et al. [19] in great detail. The categorical variables abnormal, borderline and normal are introduced in the Ewing and Clarke 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 may be identified by decision trees to maximize the predictive accuracy of CAN assessment with missing features.

DiscrI database contains a separate attribute LSNPneg that can take on values FALSE and TRUE. If LSNPneg = TRUE and LSNP ≥ 30, then the result is abnormal. If LSNPneg = TRUE and 29 ≥ LSNP ≥ 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, LSNP, 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. [17,19] defined the five classes for a CAN risk assessment given in Table 2. Ewing et al. [19] 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, 1/2 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. [19] demonstrate that these scoring systems give roughly equivalent categorizations and seem to carry no real advantages.

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. [17,19]. 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 and definite. The fourth classification divides all patients into four classes, allocated each patient to one of the following classes: normal, early, definite and severe.

3. Methods

The optimal decision path finder (ODPF) procedure for determining optimal sequences of tests was proposed and investigated in [1]. The main idea of the procedure is briefly illustrated in Fig. 1 for the purposes of our work. The ODPF uses a pre-specified threshold of confidence 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.

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

<table>
<thead>
<tr>
<th>Test</th>
<th>Value</th>
<th>Normal</th>
<th>Borderline</th>
<th>Abnormal</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAHR (ratio)</td>
<td>≥ 1.21</td>
<td>1.11–1.20</td>
<td>≤ 1.10</td>
<td></td>
</tr>
<tr>
<td>DBHR (beats/min)</td>
<td>≥ 15</td>
<td>11–14</td>
<td>≤ 10</td>
<td></td>
</tr>
<tr>
<td>HGBP (mmHg)</td>
<td>≥ 16</td>
<td>11–15</td>
<td>≤ 10</td>
<td></td>
</tr>
<tr>
<td>LSHR (ratio)</td>
<td>≥ 1.04</td>
<td>1.01–1.03</td>
<td>≤ 1.00</td>
<td></td>
</tr>
<tr>
<td>LSNP (mmHg)</td>
<td>≤ 10</td>
<td>11–29</td>
<td>≥ 30</td>
<td></td>
</tr>
</tbody>
</table>

![Fig. 1. Optimal decision path finder procedure.](image-url)

The cost-functions used in the practical assessment of CAN include (1) the number of tests, (2) time required to perform the tests and (3) the minimization of the individual difficulties faced by each patient in performing the tests and the minimization of the negative impact in those cases where any one or more of the Ewing tests cannot be completed at all.

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 the outcomes of training decision trees for all subsets of tests can be effectively applied to determine the optimal sequence of tests in the ODPF procedure. The decision trees are one of the most important algorithms for clinical applications. Their 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 [14,15], 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 [30], are readily available. We used Waikato Environment for Knowledge Analysis (WEKA) to test the J48 classifier [31]. Decision tree induction algorithms are also available in Rattle [32,33] and many other sources.

Lamb et al. [34] 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. Zupek et al. [35] 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.

4. Experimental results and discussion

In this study, we used the J48 decision tree induction algorithm in WEKA [31] for each single Ewing test, then for each group of 2, 3 and 4 tests. Throughout all experiments, 10-fold cross validation was used to estimate predictive accuracy. The training was repeated for classifications with 2, 3 and 4 classes of Ewing outcomes. Following [14], 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.

Please cite this article in press as: Stranieri A, et al. An approach for Ewing test selection to support the clinical assessment of cardiac autonomic neuropathy. Artif Intell Med (2013), http://dx.doi.org/10.1016/j.artmed.2013.04.007
We carried out a systematic investigation of the Ewing battery of five tests for the diagnosis of CAN. The set of all five Ewing tests has 32 subsets. All of these subsets were included in our experiments. Figs. 2–5 present the outcomes of an exhaustive evaluation of performance of decision trees for the Ewing battery. Fig. 2 shows that the DBHR test is most effective among all Ewing tests. The accuracy achieved by this test alone may be sufficient, for example, for advice on lifestyle based strategies mitigating CAN. The DBHR is the best test that can be used as a single independent attribute for the identification of CAN.

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 and the ODPF procedure to determine the best sequences of Ewing tests for the clinical assessment of CAN and cost-function equal to the number of tests (since the Ewing tests have approximately equal financial costs). These sequences are given in Table 3, which also contains the accuracies achieved after each initial segment of the optimal sequences for cost-function equal to the number of tests.

Furthermore, we calculated the mathematical expectations, or expected values, of the accuracies that can be achieved by a randomly chosen sequence of tests, see Table 3. This is equivalent to running a large number of new tests for many sequences chosen uniformly at random and then averaging the outcomes. The mathematical expectation is equal to a weighted sum of the accuracies for all possible sequences of tests. Assuming that at each step the next test is chosen uniformly at random, the sum coincides with an average of the accuracies presented in Figs. 2–5. We include these mathematical expectations in Table 3. For comparison, it also contains the corresponding accuracies for the diagnosis of CAN with 2, 3 and 4 categories by each initial segment of the sequence recorded previously in [18].

Fig. 6 illustrates the optimal sequence of Ewing tests and predictive accuracies that can be achieved after each step for 2 classes of CAN and the number of tests as cost-function.

Comparing parts (A), (B) and (C) of Table 3 it is easy to see that new sequences provide additional gain. Here we investigate not only the final predictive accuracy reached after performing all tests,
but are interested in the order of tests which provides the best possible accuracy after each number of tests performed during the process (for any value of the cost-function of the performing tests). Optimal sequences considered in our paper always remain optimal after each step: the very first test is chosen so that it provides the best accuracy among all single tests, the second test is then chosen so that the pair of the first two tests provides the best accuracy and so on. Therefore, after any number of steps the clinician can stop if the required level of accuracy has been reached and the preformed sequence of the first tests carried out up to this level will be an optimal one. This is the benefit of the new approach fine tuning previous work on tests for the diagnosis of CAN. Knowledge of such sequences allows the clinician to stop without completing the remaining tests when a satisfactory level of predictive accuracy has been reached.

On the other hand, another important objective of a clinician is to minimize the time required to carry out the tests and the difficulties that individual patients may experience in determining the presence of CAN. Therefore there are alternative cost-functions to be taken into account on individual basis. This may lead to a different subset of tests that has to be used in a particular situation. The clinician has to make an assessment of the time required

---

Fig. 6. Optimal sequence of Ewing tests obtained using ODPF procedure and predictive accuracies after each step for two classes of CAN and the number of tests as cost-function.

---

Fig. 7. Circle diagram.
to undertake tests at the practice and the individual difficulties associated with each case.

During the application of ODPF procedure explained in Fig. 1, a clinician could follow the resulting refinement of the protocol choosing tests so that every next test provides the greatest predictive accuracy for the resulting sequence according to the outcomes obtained in Figs. 2 to 5. These figures could be used to inform clinicians of the best remaining tests to perform for any cost-function and the ensuing predictive accuracies that can be achieved.

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, in [37–39]. It has been applied to the design of plans [40]. Graph-based approaches were used for medical visualization, for instance, in [41]. Hierarchical visualization layouts were considered in [42]. Visualization for inference has been also investigated in [43,44]. For diabetes patients, it was treated in [45].

We include Figs. 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 Fig. 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. The outcomes of our experiments can be used in conjunction with the visual representations or aids to apply the ODPF for the following three categories of cost-functions: (1) cost-function equal to the number of tests, (2) cost function equal to the time required to perform the tests and (3) cost-function expressing individual difficulties in performing the test for a particular patient or in a particular situation. For example, in applying the ODPF procedure with the diagnostic difficulty of tests as a cost-function, the clinician could use the diagrams represented in Fig. 7 and move from the centre of the diagram to the outside circle and choosing the easiest test with appropriate additional gain in predictive accuracy at each step of the process. This is similar to the application of ODPF procedure to determining optimal sequences for cost-function equal to the number of tests.

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 accuracy of classification. Determination of appropriate tests to be used for this has already been considered in the literature, see [46].

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 classifier attribute evaluation. Gain ratio attribute
evaluation assesses the significance of each attribute by calculating its gain ratio using the formula

$$\text{GainR(Class, Attribute)} = \frac{H(\text{Class}) - H(\text{Class|Attribute})}{H(\text{Attribute})},$$  \hspace{1cm} (1)

where \( H(X) \) stands for the entropy of \( X \), see [47]. Information gain attribute evaluation assesses the significance of each attribute by calculating the information gain using the formula

$$\text{InfoGain(Class, Attribute)} = H(\text{Class}) - H(\text{Class|Attribute}).$$  \hspace{1cm} (2)

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 three standard parameters associated to the ECG recordings: ECG interpretation, Grade 10 sec and QRS 10 sec. Let us refer to [48] for more information on ECG. ECG interpretation, Grade 10 sec and QRS 10 sec are standard parameters associated to a 10-ses clinical ECG recording. The ECG interpretation is determined by the cardiologist as a characterization of the ECG recording. Grade 10 sec can take on one of the following values: the categories 1a and 1b are associated with normal recordings of ECG, 2a is associated with ECG that may indicate minor pathology but is not clinically relevant, 2b suggests that it is advisable for the patient to see a doctor and category 3 means that the patient should have an immediate referral to a doctor. The QRS 10 sec refers to the time interval between onset and termination of the QRS complex within the ECG and indicates ventricular depolarisation. Note that the use of ECG data in applications of AI methods has been considered recently, for example, in [49–53].

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. Patients demographic and various clinical data are also contained in DiScrI database for each patients. Feature selection methods were applied to the whole database including the demographic and showed that these data are less significant than Grade 10 sec and QRS 10 sec in their ability to improve classification accuracy for the diagnosis of CAN. This is why we did not include the demographic features in further tests.
Acknowledgements

We carried out a complete evaluation of the predictive accuracy of J48 classifier for the Ewing battery supplemented with the QRS 10sec and Grade 10sec attributes. Figs. 9–12 include experimental results of these tests.

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. Figs. 9–12 can be used to determine the best sequence of Ewing tests, for example, for those patients who already have the values of the Grade 10sec and QRS 10sec attributes determined by the clinicians.

5. Conclusions

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. We have determined the best sequences of Ewing tests for the diagnosis of CAN with 2, 3 and 4 categories and included a table with accuracies after each initial segment of test in these sequences.

A comparison with the sequence considered previously in [18] and the mathematical expectations of the accuracies for a random sequence of tests shows significant improvement that has been achieved and demonstrates that our results can provide additional guidance to the clinicians toward selecting subsets of the whole Ewing battery and determining the order of performing the tests in each individual situation.

Our tables with outcomes contain the predictive accuracies that can be achieved by diagnosing CAN on the basis of each subset of the Ewing features and can be used to determine the optimal sequences of tests for each individual cost-function by following the ODPF procedure. The results show that the best single Ewing test for diagnosing CAN is the deep breathing heart rate variation (DBHR) test, OPTimal sequences found for the cost-function equal to the number of tests guarantee that the best accuracy is achieved after any number of tests, and provide an improvement in comparison with the original ordering of tests. In situations, where some of the Ewing tests cannot be performed, the best additional features that can be recommended to increase the accuracy are the Grade 10sec and QRS 10sec attributes. 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.

Acknowledgements

This research was supported by a Deakin-Ballarat collaboration grant. The authors are grateful to two reviewers for corrections, comments and suggestions that have helped to improve this article.

References