

# International Journal of Artificial Intelligence and Expert Systems (IJAE)

ISSN : 2180-124X

Volume 1, Issue 4

Number of issues per year: 6

# **International Journal of Artificial Intelligent and Expert Systems (IJAE)**

**Volume 1, Issue 4, 2010**

**Edited By**  
**Computer Science Journals**  
[www.cscjournals.org](http://www.cscjournals.org)

**Editor in Chief Dr. Bekir Karlik**

## **International Journal of Artificial Intelligent and Expert Systems (IJAE)**

Book: 2010 Volume 1, Issue 4

Publishing Date: 20-12-2010

Proceedings

ISSN (Online): 2180-124X

This work is subjected to copyright. All rights are reserved whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illusions, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication of parts thereof is permitted only under the provision of the copyright law 1965, in its current version, and permission of use must always be obtained from CSC Publishers. Violations are liable to prosecution under the copyright law.

IJAE Journal is a part of CSC Publishers

<http://www.cscjournals.org>

© IJAE Journal

Published in Malaysia

Typesetting: Camera-ready by author, data conversation by CSC Publishing Services – CSC Journals, Malaysia

**CSC Publishers**

## **Editorial Preface**

The International Journal of Artificial Intelligence and Expert Systems (IJAE) is an effective medium for interchange of high quality theoretical and applied research in Artificial Intelligence and Expert Systems domain from theoretical research to application development. This is the fourth issue of volume first of IJAE. The Journal is published bi-monthly, with papers being peer reviewed to high international standards. IJAE emphasizes on efficient and effective Artificial Intelligence, and provides a central for a deeper understanding in the discipline by encouraging the quantitative comparison and performance evaluation of the emerging components of Expert Systems. IJAE comprehensively cover the system, processing and application aspects of Artificial Intelligence. Some of the important topics are AI for Service Engineering and Automated Reasoning, Evolutionary and Swarm Algorithms and Expert System Development Stages, Fuzzy Sets and logic and Knowledge-Based Systems, Problem solving Methods Self-Healing and Autonomous Systems etc.

IJAE give an opportunity to scientists, researchers, and vendors from different disciplines of Artificial Intelligence to share the ideas, identify problems, investigate relevant issues, share common interests, explore new approaches, and initiate possible collaborative research and system development. This journal is helpful for the researchers and R&D engineers, scientists all those persons who are involve in Artificial Intelligence and Expert Systems in any shape.

Highly professional scholars give their efforts, valuable time, expertise and motivation to IJAE as Editorial board members. All submissions are evaluated by the International Editorial Board. The International Editorial Board ensures that significant developments in image processing from around the world are reflected in the IJAE publications.

IJAE editors understand that how much it is important for authors and researchers to have their work published with a minimum delay after submission of their papers. They also strongly believe that the direct communication between the editors and authors are important for the welfare, quality and wellbeing of the Journal and its readers. Therefore, all activities from paper submission to paper publication are controlled through electronic systems that include electronic submission, editorial panel and review system that ensures rapid decision with least delays in the publication processes.

To build its international reputation, we are disseminating the publication information through Google Books, Google Scholar, Directory of Open Access Journals (DOAJ), Open J Gate, ScientificCommons, Docstoc and many more. Our International Editors are working on establishing ISI listing and a good

impact factor for IJAE. We would like to remind you that the success of our journal depends directly on the number of quality articles submitted for review. Accordingly, we would like to request your participation by submitting quality manuscripts for review and encouraging your colleagues to submit quality manuscripts for review. One of the great benefits we can provide to our prospective authors is the mentoring nature of our review process. IJAE provides authors with high quality, helpful reviews that are shaped to assist authors in improving their manuscripts.

**Editorial Board Members**

International Journal of Artificial Intelligence and Expert Systems (IJAE)

# **Editorial Board**

## **Editor-in-Chief (EiC)**

**Dr. Bekir Karlik**  
*Mevlana University (Turkey)*

## **Associate Editors (AEiCs)**

**Assistant Professor. Tossapon Boongoen**  
*Royal Thai Air Force Academy (Thailand)*

**Assistant Professor. Ihsan Omur Bucak**  
*Mevlana University (Turkey)*

## **Editorial Board Members (EBMs)**

**Professor Yevgeniy Bodyanskiy**  
*Kharkiv National University of Radio Electronics (Ukraine)*

**Assistant Professor. Bilal Alatas**  
*Firat University (Turkey)*

**Associate Professor Abdullah Hamed Al-Badi**  
*Sultan Qaboos University (Oman)*

# Table of Content

Volume 1, Issue 4, December 2010

## Pages

75 - 87      Decision Tree Classifiers to determine the patient's Post-operative Recovery Decision

**Shanthi Dhanushkodi, G.Sahoo, Saravanan Nallaperumal**

## Decision Tree Classifiers to Determine the Patient's Post-Operative Recovery Decision

**D.Shanthi**

*Department of Computer Science and IT  
Mazoon University College  
Seeb, P.O.Box 101, Muscat, Sultanate of Oman*

dshan71@gmail.com

**Dr.G.Sahoo**

*Department of Information Technology  
Birla Institute of Technology  
Mesra, Ranchi, India*

gsahoo@bitmesra.ac.in

**Dr.N.Saravanan**

*Department of Computer Science and IT  
Mazoon University College  
Seeb, P.O.Box 101, Muscat, Sultanate of Oman*

saranshanu@gmail.com

---

### Abstract

Machine Learning aims to generate classifying expressions simple enough to be understood easily by the human. There are many machine learning approaches available for classification. Among which decision tree learning is one of the most popular classification algorithms. In this paper we propose a systematic approach based on decision tree which is used to automatically determine the patient's post-operative recovery status. Decision Tree structures are constructed, using data mining methods and then are used to classify discharge decisions.

**Keywords:** Data Mining, Decision Tree, Machine Learning, Post-operative Recovery.

---

### 1. INTRODUCTION

Decision support systems help physicians and also play an important role in medical decision making. They are based on different models and the best of them are providing an explanation together with an accurate, reliable and quick response. Clinical decision-making is a unique process that involves the interplay between knowledge of pre-existing pathological conditions, explicit patient information, nursing care and experiential learning [1]. One of the most popular among machine learning approaches is decision trees. A data object, referred to as an example, is described by a set of attributes or variables and one of the attributes describes the class that an example belongs to and is thus called the class attribute or class variable. Other attributes are often called independent or predictor attributes (or variables). The set of examples used to learn the classification model is called the training data set. Tasks related to classification include regression, which builds a model from training data to predict numerical values, and clustering, which groups examples to form categories. Classification belongs to the category of supervised learning, distinguished from unsupervised learning. In supervised learning, the training data consists of pairs of input data (typically vectors), and desired outputs, while in unsupervised learning there is no a priori output. For years they have been successfully used in many medical decision making applications. Transparent representation of acquired knowledge and fast algorithms made decision trees what they are today: one of the most often used symbolic machine learning approaches [2]. Decision trees have been already successfully used in



medicine, but as in traditional statistics, some hard real world problems cannot be solved successfully using the traditional way of induction [3]. A hybrid neuro-genetic approach has been used for the selection of input features for the neural network and the experimental results proved that the performance of ANN can be improved by selecting good combination of input variables and this hybrid approach gives better average prediction accuracy than the traditional ANN [4].

Classification has various applications, such as learning from a patient database to diagnose a disease based on the symptoms of a patient, analyzing credit card transactions to identify fraudulent transactions, automatic recognition of letters or digits based on handwriting samples, and distinguishing highly active compounds from inactive ones based on the structures of compounds for drug discovery [5]. Seven algorithms are compared in the training of the multi-layered Neural Network Architecture for the prediction of patient's post-operative recovery area and the best classification rates are compared [6].

In this study, the Gini Splitting algorithm is used with promising results in a crucial way and at the same time complicated classification problem concerning the prediction of post-operative recovery area.

## 2. MOTIVATIONS AND RELATED WORK

In data mining, there are three primary components: model representation, model evaluation and search. The two basic types of search methods used in data mining consist of two components: Parameter Search and Model Search [15]. In parameter search, the algorithm searches for the set of parameters for a fixed model representation, which optimizes the model evaluation criteria given the observed data. For relatively simple problems, the search is simple and the optimal parameter estimates can be obtained in a closed form. Typically, for more general models, a closed form solution is not available. In such cases, iterative methods, such as the gradient descent method of back-propagation for neural networks, are commonly used. The gradient descent method is one of the popular search techniques in conventional optimization [16]. Decision trees are considered to be one of the most popular approaches for representing classifiers. It induces a decision tree from data. A decision tree is a tree structured prediction model where each internal node denotes a test on an attribute, each outgoing branch represents an outcome of the test, and each leaf node is labeled with a class or class distribution. Decision trees are often used in classification and prediction. It is simple yet a powerful way of knowledge representation. The models produced by decision trees are represented in the form of tree structure. Learning a decision tree involves deciding which split to make at each node, and how deep the tree should be. A leaf node indicates the class of the examples. The instances are classified by sorting them down the tree from the root node to some leaf node [2].

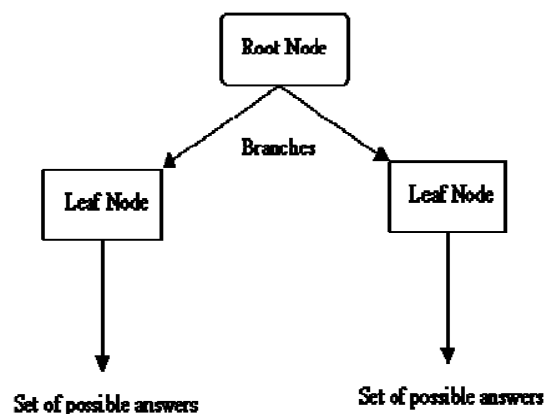


FIGURE 1: Decision Tree Structure

Decision trees (DTs) are either univariate or multivariate [17]. Univariate decision trees (UDTs) approximate the underlying distribution by partitioning the feature space recursively with axis-parallel hyperplanes. The underlying function, or relationship between inputs and outputs, is approximated by a synthesis of the hyper-rectangles generated from the partitions. Multivariate decision trees (MDTs) have more complicated partitioning methodologies and are computationally more expensive than UDTs. A typical decision tree learning algorithm adopts a top-down recursive divide-and-conquer strategy to construct a decision tree. Starting from a root node representing the whole training data, the data is split into two or more subsets based on the values of an attribute chosen according to a splitting criterion. For each subset a child node is created and the subset is associated with the child. The process is then separately repeated on the data in each of the child nodes, and so on, until a termination criterion is satisfied. Many decision tree learning algorithms exist. They differ mainly in attribute-selection criteria, such as information gain, gain ratio [7], gini index [8], termination criteria and post-pruning strategies. Post-pruning is a technique that removes some branches of the tree after the tree is constructed to prevent the tree from over-fitting the training data. Representative decision tree algorithms include CART [8] and C4.5 [7]. There are also studies on fast and scalable construction of decision trees. Representative algorithms of such kind include RainForest [9] and SPRINT [10].

### 3. METHODS AND METHODOLOGY

The tool used in this study is known as DTREG [14]. It basically follows the same principles as many other decision tree building tools, but it also implements different extensions. One of those extensions is called dynamic discretization of continuous attributes, which was used in our experiments with success. For this study, we have taken classification problem. The problem consists of determining the class (General Hospital, Go-home, Intensive care) for a certain input vector. For this study, the data was taken from UCI Machine Learning Repository. (<http://www.ics.uci.edu/~mlearn/MLSummary.html>) [11]. The data were originally created by Sharon Summers (School of Nursing, University of Kansas) and Linda Woolery (School of Nursing, University of Missouri) and donated by Jerzy W. Grzymala-Busse. The goal is to determine where patients in a postoperative recovery area should be sent to next. Because hypothermia is a significant concern after surgery. The attributes correspond roughly to body temperature measurements. The dataset contains 90 records with 8 characteristics of a patient's state in a postoperative period.

- 1) InternalTemp - patient's internal temperature in C: high ( $> 37$ ), mid ( $\geq 36$  and  $\leq 37$ ), low ( $< 36$ )
  - 2) SurfaceTemp - patient's surface temperature in C: high ( $> 36.5$ ), mid ( $\geq 36.5$  and  $\leq 35$ ), low ( $< 35$ )
  - 3) OxygenSat - oxygen saturation in %: excellent ( $\geq 98$ ), good ( $\geq 90$  and  $< 98$ ), fair ( $\geq 80$  and  $< 90$ ), poor ( $< 80$ )
  - 4) BloodPress - last measurement of blood pressure: high ( $> 130/90$ ), mid ( $\leq 130/90$  and  $\geq 90/70$ ), low ( $< 90/70$ )
  - 5) SurfTempStab - stability of patient's surface temperature: stable, mod-stable, unstable
  - 6) IntTempStab - stability of patient's internal temperature: stable, mod-stable, unstable
  - 7) BloodPressStab - stability of patient's blood pressure: stable, mod-stable, unstable
  - 8) Comfort - patient's perceived comfort at discharge, measured as an integer between 0 and 20
- A. Target Column*
- 9) Discharge Decision - discharge decision (Intensive-Care, Go-Home, General-Hospital):  
Intensive-Care (patient sent to Intensive Care Unit),  
Go-Home (patient prepared to go home),  
General-Hospital (patient sent to general hospital floor).

### 4. EXPERIMENTAL RESULTS AND EVALUATION

The input to a classification problem is a dataset of training records and each record has several attributes. Attributes whose domains are numerical are called numerical attributes (continuous attributes), whereas attributes whose domains are not are called categorical attributes (discrete

attributes). There is one distinguished attribute, called *class label*, which is a categorical attribute with a very small domain. The remaining attributes are called *predictor attributes*; they are either continuous or discrete in nature. The goal of classification is to build a concise model of the distribution of the class label in terms of the predictor attributes [12].

| SI.No | Attributes     |
|-------|----------------|
| 1.    | InternalTemp   |
| 2.    | SurfaceTemp    |
| 3.    | OxygenSat      |
| 4.    | BloodPress     |
| 5.    | SurfTempStab   |
| 6.    | IntTempStab    |
| 7.    | BloodPressStab |
| 8.    | Comfort        |

**TABLE 1:** Input Attributes

In this data analysis the last column will be considered as the target one and other columns will be considered as input columns. The target classes are depicted below in Table 2.

| SI.No | Output           |
|-------|------------------|
| 1.    | General-Hospital |
| 2.    | Go Home          |
| 3.    | Intensive Care   |

**TABLE 2:** Target Classes

In this data set total numbers of variables are 9 and for data sub setting has been done using all data rows. The total numbers of data rows are 88 and total weights for all rows are 88. From analysis it has been noticed that number of rows with missing target or weight values are 0 and rows with missing predictor values are 3 and were discarded because these variables had missing values. Decision tree inducers are algorithms that automatically construct a decision tree from a given dataset. Typically the goal is to find the optimal decision tree by minimizing the generalization error. However, other target functions can be also defined, for instance, minimizing the number of nodes or minimizing the average depth.

| Target variable                     | Discharge Decision                  |
|-------------------------------------|-------------------------------------|
| Number of predictor variables       | 8                                   |
| Type of model                       | Single tree                         |
| Maximum splitting levels            | 10                                  |
| Type of analysis                    | Classification                      |
| Splitting algorithm                 | Gini                                |
| Category weights (priors)           | Data file distribution              |
| Misclassification costs             | Equal (unitary)                     |
| Variable weights                    | Equal                               |
| Minimum size node to split          | 10                                  |
| Max. categories for continuous      | 200                                 |
| Use surrogate splitters for missing | Yes                                 |
| Always compute surrogate splitters  | No                                  |
| Tree pruning and validation method  | Cross validation                    |
| Tree pruning criterion              | Minimum cost complexity (0.00 S.E.) |
| Number of cross-validation folds    | 10                                  |

**TABLE 3:** Decision Tree Parameters

The Gini splitting algorithm is used for classification trees. Each split is chosen to maximize the heterogeneity of the categories of the target variable in the child nodes. Cross Validation method is used to evaluate the quality of the model. In this study, single decision tree model was built with maximum splitting level is 10 and misclassification cost (Equal) are same for all categories. Surrogate splitters are used for the classification of rows with missing values.

There are various top-down decision trees inducers such as ID3 [13], C4.5 [7], CART [8]. Some consist of two conceptual phases: growing and pruning (C4.5 and CART). Other inducers perform only the growing phase. In most of the cases, the discrete splitting functions are univariate. Univariate means that an internal node is split according to the value of a single attribute. Consequently, the inducer searches for the best attribute upon which to split. There are various univariate criteria. These criteria can be characterized in different ways:

- According to the origin of the measure: information theory, dependence, and distance.
- According to the measure structure: impurity based criteria, normalized impurity based criteria and Binary criteria.

Gini index is an impurity-based criterion that measures the divergences between the probability distributions of the target attribute's values. In this work Gini index has been used and it is defined as

$$Gini(y, S) = 1 - \sum_{c_j \in dom(y)} \left( \frac{|\sigma_{y=c_j} S|}{|S|} \right)^2 \tag{1}$$

Consequently the evaluation criterion for selecting the attribute  $a_i$  is defined as:

$$GiniGain(a_i, S) = Gini(y, S) - \sum_{v_{i,j} \in dom(a_i)} \frac{|\sigma_{a_i=v_{i,j}} S|}{|S|} \cdot Gini(y, \sigma_{a_i=v_{i,j}} S) \tag{2}$$

The summary of variables section displays information about each variable that was present in the input dataset. The first column shows the name of the variable, the second column shows how the variable was used; the possibilities are Target, Predictor, Weight and Unused. The third column shows whether the variable is categorical or continuous, the fourth column shows how

many data rows had missing values on the variable, and the fifth column shows how many categories (discrete values) the variable has, as shown in table 4.

| No | Variable           | Class     | Type        | Missing rows | Categories |
|----|--------------------|-----------|-------------|--------------|------------|
| 1  | InternalTemp       | Predictor | Categorical | 0            | 3          |
| 2  | SurfaceTemp        | Predictor | Categorical | 0            | 3          |
| 3  | OxygenSat          | Predictor | Categorical | 0            | 2          |
| 4  | BloodPress         | Predictor | Categorical | 0            | 3          |
| 5  | SurfaceTempStab    | Predictor | Categorical | 0            | 2          |
| 6  | IntTempStab        | Predictor | Categorical | 0            | 3          |
| 7  | BloodPressStab     | Predictor | Categorical | 0            | 3          |
| 8  | Comfort            | Predictor | Continuous  | 3            | 4          |
| 9  | Discharge Decision | Target    | Categorical | 0            | 3          |

**TABLE 4:** Summary of variables

In Table 4, Discharge Decision is a target variable and the values are to modeled and predicted by other variables. There must be one and only one target variable in a model. The variables such as InternalTemp, SurfaceTemp, OxygenSat, etc are predictor variables and the values of this variable are used to predict the value of the target variable. In this study there are 8 predictor variables and 1 target variable are used to predict the model. All the variables except comfort are categorical variables.

|  |                     |
|--|---------------------|
| Maximum depth of the tree                    | 9                   |
| Total number of group splits                 | 16                  |
| No of terminal (leaf) nodes of a tree        | 9                   |
| The minimum validation relative error occurs | with 9 nodes.       |
| The relative error value                     | 1.1125              |
| Standard error                               | 0.1649              |
| The tree will be pruned                      | from 28 to 7 nodes. |

**TABLE 5:** Model Size - Summary Report

Table 5 displays information about the maximum size tree that was built, and it shows summary information about the parameters that were used to prune the tree.

| Nodes | Val Cost | Val std err | RS cost | Complexity |
|-------|----------|-------------|---------|------------|
| 9     | 1.1125   | 0.1649      | 0.7200  | 0.000000   |
| 1     | 1.0000   | 0.0000      | 1.0000  | 0.009943   |

**TABLE 6:** Validation Statistics

Table 6 displays information about the size of the generated tree and statistics used to prune the tree. There are five columns in the table:

**Nodes** – This is the number of terminal nodes in a particular pruned version of the tree. It will range from 1 up to the maximum nodes in the largest tree that was generated. The maximum number of nodes will be limited by the maximum depth of the tree and the minimum node size allowed to be split on the Design property page for the model.

**Val cost** – This is the validation cost of the tree pruned to the reported number of nodes. It is the error cost computed using either cross-validation or the random-row-holdback data. The displayed cost value is the cost relative to the cost for a tree with one node.

The validation cost is the best measure of how well the tree will fit an independent dataset different from the learning dataset.

**Val std. err.** – This is the standard error of the validation cost value.

**RS cost** – This is the re-substitution cost computed by running the learning dataset through the tree. The displayed re-substitution cost is scaled relative to the cost for a tree with one node. Since the tree is being evaluated by the same data that was used to build it, the re-substitution cost does not give an honest estimate of the predictive accuracy of the tree for other data.

**Complexity** – This is a *Cost Complexity* measure that shows the relative tradeoff between the number of terminal nodes and the misclassification cost.

| Category         | Actual    |           | Misclassified |           | Percent       | Cost         |
|------------------|-----------|-----------|---------------|-----------|---------------|--------------|
|                  | Count     | Weight    | Count         | Weight    |               |              |
| General-Hospital | 63        | 63        | 1             | 1         | 1.587         | 0.0616       |
| Go-Home          | 23        | 23        | 15            | 15        | 65.217        | 0.652        |
| Intensive-Care   | 2         | 2         | 2             | 2         | 100.00        | 1.000        |
| <b>Total</b>     | <b>88</b> | <b>88</b> | <b>18</b>     | <b>18</b> | <b>20.455</b> | <b>0.205</b> |

TABLE 7: Misclassification Table: Training Data

Table 7 shows the misclassifications for the training dataset and it describes the number of rows for “General – Hospital” misclassified by the tree 1, for Go-Home is 15 and for Intensive-Care is 2. This misclassification cost for Intensive-Care is high.

| Category         | Actual    |           | Misclassified |           | Percent       | Cost         |
|------------------|-----------|-----------|---------------|-----------|---------------|--------------|
|                  | Count     | Weight    | Count         | Weight    |               |              |
| General-Hospital | 63        | 63        | 8             | 8         | 12.698        | 0.127        |
| Go-Home          | 23        | 23        | 18            | 18        | 78.261        | 0.783        |
| Intensive-Care   | 2         | 2         | 2             | 2         | 100.00        | 1.000        |
| <b>Total</b>     | <b>88</b> | <b>88</b> | <b>28</b>     | <b>28</b> | <b>31.818</b> | <b>0.318</b> |

TABLE 8: Misclassification Table: Validation Data

Table 8 shows the misclassifications for the validation dataset and total misclassification percent is 31.818%. A Confusion Matrix provides detailed information about how data rows are classified by the model. The matrix has a row and column for each category of the target variable. The categories shown in the first column are the actual categories of the target variable. The categories shown across the top of the table are the predicted categories. The numbers in the cells are the weights of the data rows with the actual category of the row and the predicted category of the column. Table 9 & 10 shows confusion matrix for training data and for validation data.

| Actual Category  | Predicated Category |         |                |
|------------------|---------------------|---------|----------------|
|                  | General-Hospital    | Go-Home | Intensive-care |
| General-Hospital | 62                  | 1       | 0              |
| Go-Home          | 15                  | 8       | 0              |
| Intensive-Care   | 2                   | 0       | 0              |

**TABLE 9:** Confusion Matrix- Training Data

The numbers in the diagonal cells are the weights for the correctly classified cases where the actual category matches the predicted category. The off-diagonal cells have misclassified row weights.

| Actual Category  | Predicated Category |         |                |
|------------------|---------------------|---------|----------------|
|                  | General-Hospital    | Go-Home | Intensive-care |
| General-Hospital | 58                  | 8       | 0              |
| Go-Home          | 18                  | 5       | 0              |
| Intensive-Care   | 1                   | 1       | 0              |

**TABLE 10:** Confusion Matrix - Validation Data

**A. Lift/Gain for Training Data**

The lift and gain table is a useful tool for measuring the value of a predictive model. Lift and gain values are especially useful when a model is being used to target (prioritize) marketing efforts. The following 3 table's shows Lift/Gain for the three different outputs such as General-Hospital, Go-Home, and Intensive Care.

Lift/Gain for Discharge Decision = General-Hospital

| Bin Index | Class % of Bin | Cum % Population | Cum % of Class | Gain | % of Population | % of class | Lift |
|-----------|----------------|------------------|----------------|------|-----------------|------------|------|
| 1         | 88.89          | 10.23            | 12.70          | 1.24 | 10.23           | 12.70      | 1.24 |
| 2         | 66.67          | 20.45            | 22.22          | 1.09 | 10.23           | 9.52       | 0.93 |
| 3         | 77.78          | 30.68            | 33.33          | 1.09 | 10.23           | 11.11      | 1.09 |
| 4         | 88.89          | 40.91            | 46.03          | 1.13 | 10.23           | 12.70      | 1.24 |
| 5         | 77.78          | 51.14            | 57.14          | 1.12 | 10.23           | 11.11      | 1.09 |
| 6         | 77.78          | 61.36            | 68.25          | 1.11 | 10.23           | 11.11      | 1.09 |
| 7         | 100.00         | 71.59            | 82.54          | 1.15 | 10.23           | 14.29      | 1.40 |
| 8         | 66.67          | 81.82            | 92.06          | 1.13 | 10.23           | 9.52       | 0.93 |
| 9         | 44.44          | 92.05            | 98.41          | 1.07 | 10.23           | 6.35       | 0.62 |
| 10        | 14.29          | 100.00           | 100.00         | 1.00 | 7.95            | 1.59       | 0.20 |

**TABLE 11:** Lift and Gain for training data (General-Hospital)

Average gain = 1.112

Percent of cases with Discharge Decision = General-Hospital: 71.59%

Lift/Gain for Discharge Decision = Go-Home

| Bin Index | Class % of Bin | Cum % Population | Cum % of Class | Gain | % of Population | % of class | Lift |
|-----------|----------------|------------------|----------------|------|-----------------|------------|------|
| 1         | 88.89          | 10.23            | 34.78          | 3.40 | 10.23           | 34.78      | 3.40 |
| 2         | 0.00           | 20.45            | 34.78          | 1.70 | 10.23           | 0.00       | 0.00 |
| 3         | 33.33          | 30.68            | 47.83          | 1.56 | 10.23           | 12.04      | 1.28 |
| 4         | 11.11          | 40.91            | 52.17          | 1.28 | 10.23           | 4.35       | 0.43 |
| 5         | 22.22          | 51.14            | 60.87          | 1.19 | 10.23           | 8.70       | 0.85 |
| 6         | 22.22          | 61.36            | 69.57          | 1.13 | 10.23           | 8.70       | 0.85 |
| 7         | 22.22          | 71.59            | 78.26          | 1.09 | 10.23           | 8.70       | 0.85 |
| 8         | 22.22          | 81.82            | 86.96          | 1.06 | 10.23           | 8.70       | 0.85 |
| 9         | 11.11          | 92.05            | 91.30          | 0.99 | 10.23           | 4.35       | 0.43 |
| 10        | 28.57          | 100.00           | 100.00         | 1.00 | 7.95            | 8.70       | 1.09 |

**TABLE 12:** Lift and Gain for training data (Go-Home)

Average gain = 1.441

Percent of cases with Discharge Decision = Go-Home: 26.14%

Lift/Gain for Discharge Decision = Intensive-Care

| Bin Index | Class % of Bin | Cum % Population | Cum % of Class | Gain | % of Population | % of class | Lift |
|-----------|----------------|------------------|----------------|------|-----------------|------------|------|
| 1         | 2.27           | 10.00            | 10.00          | 1.00 | 10.00           | 10.00      | 1.00 |
| 2         | 2.27           | 20.00            | 20.00          | 1.00 | 10.00           | 10.00      | 1.00 |
| 3         | 2.27           | 30.00            | 30.00          | 1.00 | 10.00           | 10.00      | 1.00 |
| 4         | 2.27           | 40.00            | 40.00          | 1.00 | 10.00           | 10.00      | 1.00 |
| 5         | 2.27           | 50.00            | 50.00          | 1.00 | 10.00           | 10.00      | 1.00 |
| 6         | 2.27           | 60.00            | 60.00          | 1.00 | 10.00           | 10.00      | 1.00 |
| 7         | 2.27           | 70.00            | 70.00          | 1.00 | 10.00           | 10.00      | 1.00 |
| 8         | 2.27           | 80.00            | 80.00          | 1.00 | 10.00           | 10.00      | 1.00 |
| 9         | 2.27           | 90.00            | 90.00          | 1.00 | 10.00           | 10.00      | 1.00 |
| 10        | 2.27           | 100.00           | 100.00         | 1.00 | 10.00           | 10.00      | 1.00 |



**TABLE 13:** Lift and Gain for training data (Intensive -Care)

Average gain = 1.000

Percent of cases with Discharge Decision = Intensive-Care: 2.27%

Lift and Gain for validation data

Lift/Gain for Discharge Decision = General-Hospital

| Bin Index | Class % of Bin | Cum % Population | Cum % of Class | Gain | % of Population | % of class | Lift |
|-----------|----------------|------------------|----------------|------|-----------------|------------|------|
| 1         | 77.78          | 10.23            | 11.11          | 1.09 | 10.23           | 11.11      | 1.09 |
| 2         | 55.56          | 20.45            | 19.05          | 0.93 | 10.23           | 7.94       | 0.78 |
| 3         | 77.78          | 30.68            | 30.16          | 0.98 | 10.23           | 11.11      | 1.09 |
| 4         | 77.78          | 40.91            | 41.27          | 1.01 | 10.23           | 11.11      | 1.09 |
| 5         | 88.89          | 51.14            | 53.97          | 1.06 | 10.23           | 12.70      | 1.24 |
| 6         | 88.89          | 61.36            | 66.67          | 1.09 | 10.23           | 12.70      | 1.24 |
| 7         | 100.00         | 71.59            | 80.95          | 1.13 | 10.23           | 14.29      | 1.40 |
| 8         | 44.44          | 81.82            | 87.30          | 1.07 | 10.23           | 6.35       | 0.62 |
| 9         | 44.44          | 92.05            | 93.65          | 1.02 | 10.23           | 6.35       | 0.62 |
| 10        | 57.14          | 100.00           | 100.00         | 1.00 | 7.95            | 6.35       | 0.80 |

**TABLE 14:** Lift and Gain for Validation data (General Hospital)

Average gain = 1.037

Percent of cases with Discharge Decision = General-Hospital: 71.59%

Lift/Gain for Discharge Decision = Go-Home

| Bin Index | Class % of Bin | Cum % Population | Cum % of Class | Gain | % of Population | % of class | Lift |
|-----------|----------------|------------------|----------------|------|-----------------|------------|------|
| 1         | 22.22          | 10.23            | 8.70           | 0.85 | 10.23           | 8.70       | 0.85 |
| 2         | 44.44          | 20.45            | 26.09          | 1.28 | 10.23           | 17.39      | 1.70 |
| 3         | 44.44          | 30.68            | 43.48          | 1.42 | 10.23           | 17.39      | 1.70 |
| 4         | 33.33          | 40.91            | 56.52          | 1.38 | 10.23           | 13.04      | 1.28 |
| 5         | 22.22          | 51.14            | 65.22          | 1.28 | 10.23           | 8.70       | 0.85 |
| 6         | 22.22          | 61.36            | 73.91          | 1.20 | 10.23           | 8.70       | 0.85 |
| 7         | 11.11          | 71.59            | 78.26          | 1.09 | 10.23           | 4.35       | 0.43 |
| 8         | 22.22          | 81.82            | 86.96          | 1.06 | 10.23           | 8.70       | 0.85 |
| 9         | 22.22          | 92.05            | 95.65          | 1.04 | 10.23           | 8.70       | 0.85 |
| 10        | 14.29          | 100.00           | 100.00         | 1.00 | 7.95            | 4.35       | 0.55 |

**TABLE 15:** Lift and Gain for Validation data (Go-Home)

Average gain = 1.160

Percent of cases with Discharge Decision = Go-Home: 26.14%

Lift/Gain for Discharge Decision = Intensive-Care

| Bin Index | Class % of Bin | Cum % Population | Cum % of Class | Gain | % of Population | % of class | Lift |
|-----------|----------------|------------------|----------------|------|-----------------|------------|------|
| 1         | 2.27           | 10.00            | 10.00          | 1.00 | 10.00           | 10.00      | 1.00 |
| 2         | 2.27           | 20.00            | 20.00          | 1.00 | 10.00           | 10.00      | 1.00 |
| 3         | 2.27           | 30.00            | 30.00          | 1.00 | 10.00           | 10.00      | 1.00 |
| 4         | 2.27           | 40.00            | 40.00          | 1.00 | 10.00           | 10.00      | 1.00 |
| 5         | 2.27           | 50.00            | 50.00          | 1.00 | 10.00           | 10.00      | 1.00 |
| 6         | 2.27           | 60.00            | 60.00          | 1.00 | 10.00           | 10.00      | 1.00 |
| 7         | 2.27           | 70.00            | 70.00          | 1.00 | 10.00           | 10.00      | 1.00 |
| 8         | 2.27           | 80.00            | 80.00          | 1.00 | 10.00           | 10.00      | 1.00 |
| 9         | 2.27           | 90.00            | 90.00          | 1.00 | 10.00           | 10.00      | 1.00 |
| 10        | 2.27           | 100.00           | 100.00         | 1.00 | 10.00           | 10.00      | 1.00 |

**TABLE 16:** Lift and Gain for Validation data (Intensive Care)

Average gain = 1.000

Percent of cases with Discharge Decision = Intensive-Care: 2.27%

Table 17 shows the relative importance of each predictor variables. The calculation is performed using sensitive analysis where the values of each variables are randomized and the effect on the quality of the model is measured.

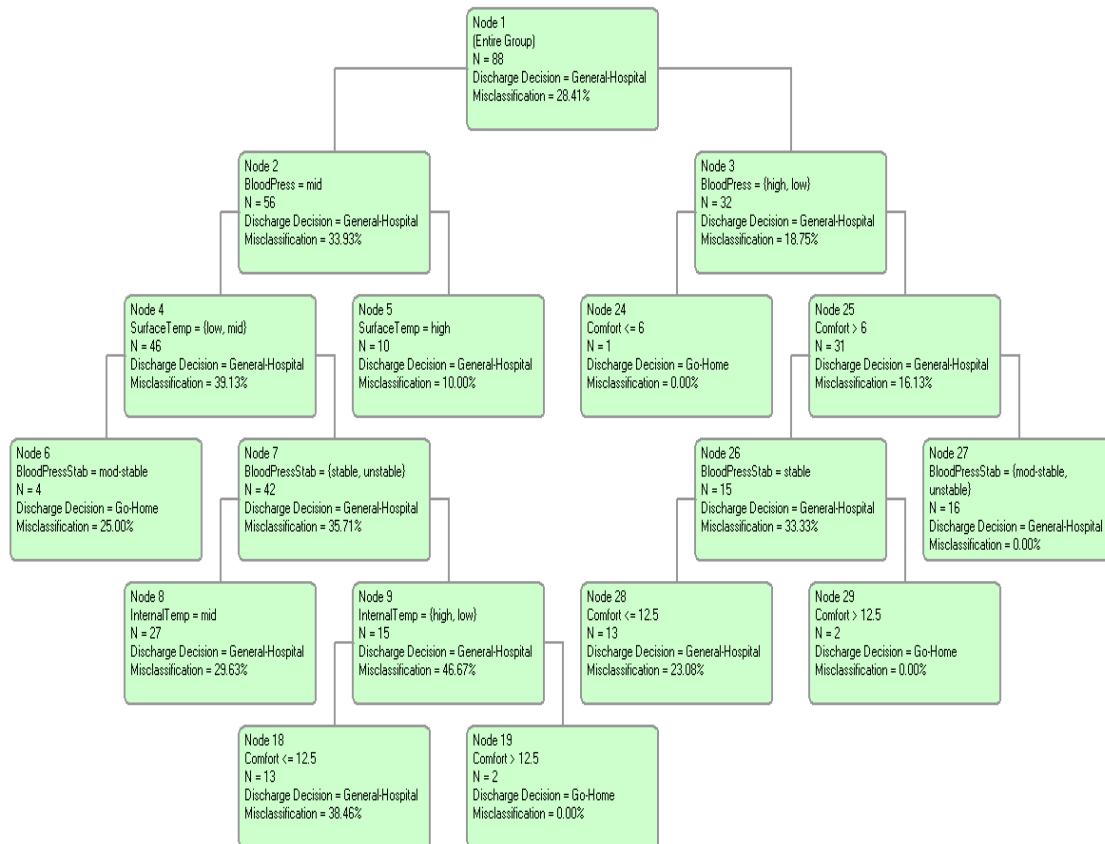
| SI.No | Variable       | Importance |
|-------|----------------|------------|
| 1     | Comfort        | 100.00     |
| 2     | BloodPressStab | 51.553     |
| 3     | InternalTemp   | 34.403     |
| 4     | SurfaceTemp    | 25.296     |
| 5     | IntTempStab    | 20.605     |
| 6     | BloodPress     | 20.045     |

**TABLE 17:** Over all Importance of Variables

The results of this experimental research through the above critical analysis using this decision tree based method shows more accurate data prediction and helping the clinical experts in taking decisions such as General Hospital, Go-Home and Intensive-care. More accuracy will be guaranteed for large data sets.

## 5. CONCLUSION

This paper has presented a decision tree classifier to determine the patient’s post-operative recovery decisions such as General Hospital, Go-Home and Intensive-care. Figure 2 depicts the decision tree model for the prediction of patient’s post-operative recovery area.



**FIGURE 2:** Decision tree for the prediction of post-operative recovery area

In figure 2, each node represents set of records from the original data set. Nodes that have child nodes such as 1, 2, 3, 4, 25, 7, 26, 9 are called interior nodes. Nodes 5, 24, 6, 27, 8, 28, 29, 18, 19 are called leaf nodes. Node 1 is root node. The value N in each node shows number of rows that fall on that category. The target variable is Discharge Decision. Misclassification percentage in each node shows the percentage of the rows in this node that had target variable categories different from the category that was assigned to the node. In other words, it is the percentage of rows that were misclassified. For instance, from the tree it is clearly understood the value of predictor variable comfort is  $\leq 6$  the decision is Go-Home. If Comfort is  $>6$  additional split is required to classify. The total time consumed for analysis is 33ms.

The Decision tree analysis discussed in this paper highlights the patient sub-groups and critical values in variables assessed. Importantly the results are visually informative and often have clear clinical interpretation about risk factors faced by these subgroups. The results shows 71.59% of cases with General-Hospital 26.14% of cases with Go-Home and 2.27% cases with Intensive-Care.

## 6. REFERENCES

1. Banning, M. (2007). "A review of clinical decision making models and current research," Journal of clinical nursing, [online] available at: <http://www.blackwellsynergy.com/doi/pdf/10.1111/j.1365-2702.2006.01791.x>
2. T. af Klercker (1996): "Effect of Pruning of a Decision-Tree for the Ear, Nose and Throat Realm in Primary Health Care Based on Case-Notes". Journal of Medical Systems, 20(4): 215-226.

3. M. Zorman, V.Podgorelec, P.Kokol, M.Peterson, J. Lane (2000). "Decision tree's induction strategies evaluated on a hard real world problem." In: 13th IEEE symposium on computer-based medical systems 22-24 June 2000, Houston, Texas, USA: proceedings, Los Alamitos, IEEE Computer Society 19-24.
4. Shanthi D, Sahoo G, Saravanan N. (2008). "Input Feature Selection using Hybrid Neuro-Genetic Approach in the diagnosis of Stroke." International Journal of Computer Science and Network Security, ISSN: 1738-7906, 8(12):99-107.
5. F.S,Khan, R,M,Anwer, of Torgersson, and G. Falkman (2009), "Data Mining in Oral Medicine Using Decision Trees", International Journal of Biological and Medical Sciences 4:3.
6. Shanthi D, Sahoo G, Saravanan N. (2009), "Comparison of Neural Network Training Algorithms for the prediction of the patient's post-operative recovery area", Journal of Convergence Information Technology, ISSN: 1975-9320,4(1):24-32.
7. Quinlan JR (1993) *C4.5: "programs for machine learning"*. California: Morgan Kaufmann Publishers.
8. Breiman L, Friedman JH, Olshen RA, et al., (1984). "Classification and regression trees." Belmont Calif: Wadsworth International Group.
9. Gehrke, J., Ramakrishnan, R., & Ganti, V. (1998). "Rain-Forest A framework for fast decision tree construction of large datasets". Proceedings of the 24<sup>th</sup> International Conference on Very Large Data Bases, pp. 416–427.
10. Shafer, J., Agrawal, R., & Mehta, M. (1996). "SPRINT: A scalable parallel classifier for data mining." Proceedings of the 22<sup>th</sup> International Conference on Very Large Data Bases, pp. 544–555.
11. "UCI Machine Learning Repository" [online] available at: (<http://www.ics.uci.edu/~mllearn/MLSummary.html>). The data was originally created by Sharon Summers (School of Nursing, University of Kansas) and Linda Woolery (School of Nursing, University of Missouri) and donated by Jerzy W. Grzymala-Busse.
12. Abdel-badeeh M.Salem, Abeer M.Mahmoud (2002). "A hybrid genetic algorithm –Decision tree classifier," IJICIS, 2(2)
13. Quinlan, J. R. (1986). "Induction of decision trees". Machine Learning, 1, 81–106.
14. DTREG version 9.1 "Decision Tree Software for Predictive Modeling and Forecasting".
15. Fayyad, U., Piatetsky-Shaprio, G., Smyth, P. & Uthurusamy, R. (1996). (eds.), "Advances in Knowledge Discovery and Data Mining", MIT Press, Cambridge, MA
16. Hillier, F. & Lieberman, G. (2001). "Introduction to Operations Research," McGrawHill, Boston.
17. Abbass, H.A., Towsey M., &Finn G. (2001). C-Net: "A Method for Generating Non-deterministic and Dynamic Multivariate Decision Trees." Knowledge and Information Systems: An International Journal, Springer-Verlag, 5(2).

# CALL FOR PAPERS

**Journal:** International Journal of Artificial Intelligence and Expert Systems (IJAE)

**Volume:** 2 **Issue:** 1

**ISSN:** 2180-124x

**URL:** <http://www.cscjournals.org/csc/description.php?JCode=IJAE>

The main aim of International Journal of Artificial Intelligence and Expert Systems (IJAE) is to provide a platform to AI & Expert Systems (ES) scientists and professionals to share their research and report new advances in the field of AI and ES. IJAE is a refereed journal producing well-written original research articles and studies, high quality papers as well as state-of-the-art surveys related to AI and ES. By establishing an effective channel of communication between theoretical researchers and practitioners, IJAE provides necessary support to practitioners in the design and development of intelligent and expert systems, and the difficulties faced by the practitioners in using the theoretical results provide feedback to the theoreticians to revalidate their models. IJAE thus meets the demand of both theoretical and applied researchers in artificial intelligence, soft computing and expert systems.

IJAE is a broad journal covering all branches of Artificial Intelligence and Expert Systems and its application in the topics including but not limited to technology & computing, fuzzy logic, expert systems, neural networks, reasoning and evolution, automatic control, mechatronics, robotics, web intelligence applications, heuristic and AI planning strategies and tools, computational theories of learning, intelligent system architectures.

To build its International reputation, we are disseminating the publication information through Google Books, Google Scholar, Directory of Open Access Journals (DOAJ), Open J Gate, ScientificCommons, Docstoc and many more. Our International Editors are working on establishing ISI listing and a good impact factor for IJAE.

## **IJAE List of Topics:**

The realm of International Journal of Artificial Intelligence and Expert Systems(IJAE) extends, but not limited, to the following:

- AI for Web Intelligence Applications
- AI Parallel Processing Tools
- AI Tools for Computer Vision and Speech Understand
- Application in VLSI Algorithms and Mobile Communic
- AI in Bioinformatics
- AI Tools for CAD and VLSI Analysis/Design/Testing
- AI Tools for Multimedia
- Automated Reasoning

- Case-based reasoning
- Derivative-free Optimisation Algorithms
- Evolutionary and Swarm Algorithms
- Expert Systems Components
- Fuzzy Sets and logic
- Hybridisation of Intelligent Models/algorithms
- Inference
- Intelligent Planning
- Intelligent System Architectures
- Knowledge-Based Systems
- Logic Programming
- Multi-agent Systems
- Neural Networks for AI
- Parallel and Distributed Realisation of Intelligen
- Reasoning and Evolution of Knowledge Bases
- Rule-Based Systems
- Uncertainty
- Data and Web Mining
- Emotional Intelligence
- Expert System Development Stages
- Expert-System Development Lifecycle
- Heuristic and AI Planning Strategies and Tools
- Image Understanding
- Integrated/Hybrid AI Approaches
- Intelligent Search
- Knowledge Acquisition
- Knowledge-Based/Expert Systems
- Machine learning
- Neural Computing
- Object-Oriented Programming for AI
- Problem solving Methods
- Rough Sets
- Self-Healing and Autonomous Systems
- Visual/linguistic Perception

## **IMPORTANT DATES**

**Volume:** 2

**Issue:** 1

**Paper Submission:** January 31, 2011

**Author Notification:** March 01, 2011

**Issue Publication:** March / April 2011

## CALL FOR EDITORS/REVIEWERS

CSC Journals is in process of appointing Editorial Board Members for ***International Journal of Artificial Intelligent and Expert Systems (IJAE)***. CSC Journals would like to invite interested candidates to join **IJAE** network of professionals/researchers for the positions of Editor-in-Chief, Associate Editor-in-Chief, Editorial Board Members and Reviewers.

The invitation encourages interested professionals to contribute into CSC research network by joining as a part of editorial board members and reviewers for scientific peer-reviewed journals. All journals use an online, electronic submission process. The Editor is responsible for the timely and substantive output of the journal, including the solicitation of manuscripts, supervision of the peer review process and the final selection of articles for publication. Responsibilities also include implementing the journal's editorial policies, maintaining high professional standards for published content, ensuring the integrity of the journal, guiding manuscripts through the review process, overseeing revisions, and planning special issues along with the editorial team.

A complete list of journals can be found at <http://www.cscjournals.org/csc/byjournal.php>. Interested candidates may apply for the following positions through <http://www.cscjournals.org/csc/login.php>.

*Please remember that it is through the effort of volunteers such as yourself that CSC Journals continues to grow and flourish. Your help with reviewing the issues written by prospective authors would be very much appreciated.*

Feel free to contact us at [coordinator@cscjournals.org](mailto:coordinator@cscjournals.org) if you have any queries.

## **Contact Information**

### **Computer Science Journals Sdn Bhd**

M-3-19, Plaza Damas Sri Hartamas  
50480, Kuala Lumpur MALAYSIA

Phone: +603 6207 1607  
          +603 2782 6991  
Fax:     +603 6207 1697

### **BRANCH OFFICE 1**

Suite 5.04 Level 5, 365 Little Collins Street,  
MELBOURNE 3000, Victoria, AUSTRALIA

Fax: +613 8677 1132

### **BRANCH OFFICE 2**

Office no. 8, Saad Arcad, DHA Main Bulevard  
Lahore, PAKISTAN

### **EMAIL SUPPORT**

Head CSC Press: [coordinator@cscjournals.org](mailto:coordinator@cscjournals.org)  
CSC Press: [cscpress@cscjournals.org](mailto:cscpress@cscjournals.org)  
Info: [info@cscjournals.org](mailto:info@cscjournals.org)



COMPUTER SCIENCE JOURNALS SDN BHD  
M-3-19, PLAZA DAMAS  
SRI HARTAMAS  
50480, KUALA LUMPUR  
MALAYSIA