# SEISMIC RISK MITIGATION IN URBAN AREAS BASED ON ARTIFICIAL INTELLIGENCE METHODS

**G.M. Atanasiu[1], F. Leon[2], B.F. Popa[3]**

[1] *Professor, Dept. of Civil Engineering,"Gh. Asachi" Technical University, Iasi, Romania*
[2] *Lecturer, Dept. of Computer Science and Engineering,"Gh. Asachi" Technical University, Iasi, Romania*
[3] *Researcher, Dept. of Civil Engineering,"Gh. Asachi" Technical University, Iasi, Romania*
*Email: atanasiu@ce.tuiasi.ro*

**ABSTRACT :**

This paper presents a GIS-based methodology for monitoring the seismic performance, while taking into account the deteriorations revealed during scenarios, aiming at the identification of the seismic serviceability of the structure. By using geospatial data, one can develop useful scenarios to improve the knowledge on structural vulnerability of the urban built infrastructure. Scenarios of modeling, simulation and nonlinear seismic analysis are described and applied to a class of damaged models for some of the structures typical of the existing urban infrastructure of Iaşi, Romania. Some data mining techniques, especially decision trees are presented as a tool for awareness and mitigation of seismic effects of possible future events in the urban area.

**KEYWORDS:** data mining, seismic risk mitigation, urban areas management, geographic information systems

## 1. INTRODUCTION

The Vrancea region in Romania, located where the Carpathians Mountains Arch bends, is a source of subcrustal seismic activity, being the contact zone of four tectonic plates and subplates: Intra-Alpine, East-European, Moesian and Black Sea.

The earthquake loss estimation methodology is intended to provide local, state and regional officials with the tools necessary to assess the risks from earthquakes. Traditional loss estimation methodologies can be characterized as static: inventory data and geologic attributes were collected, one or more scenarios were evaluated and a report was written. Emphasis was given to one parameter over another based on what the author(s) considered the "controlling factor" and there was no mechanism to carry out what-if-analysis to account for the inventory variability, the geo-hazard data accuracy, and the uncertainty in the overall approach (Bouhafs, Si, Lawson & Bouabid, 1997).

Geographical Information Systems (GIS) try to change this approach. Faced with limited resources and competing priorities, the decision-maker requires accurate and accessible information. One of the greatest challenges in developing adequate information resources is interoperability, or the need to accommodate multiple users, data providers, hazard stages, scenario simulations, and mitigation goals (Wood & Stein, 2001). A fundamental principle of risk assessment is that risk due to natural hazards such as earthquakes, hurricanes and floods is location dependent. The process of risk assessment involves hazard assessment and vulnerability analysis. The probability of earthquake occurrence varies depending on location, and local site conditions also play a vital role in determining the intensity of the earthquake. Zoning of hazard prone regions is a common practice. The vulnerability of buildings, other critical structures and population is dependent on their exposure to the hazard, which varies from location to location. The spatial characteristics of hazard and vulnerability justify the application of mapping and spatial technologies such as GIS in the risk assessment process.

The whole process of digital management of the vulnerability of constructions in built urban environment is an integrated activity with multidisciplinary features, involving civil engineers as well as architects, IT administrators, and the public administration sector. The strategic objective of this process addresses the following purposes (Atanasiu, Leon & Gâlea, 2006; Atanasiu & Gâlea, 2008):

- vulnerability assessment of existing infrastructure for planning the preventive measures of human safety against earthquake;
- creating instruments for the emergency management of situations based on a possible seismic scenario;
- education goals for enhancing the social culture in crises management during and post catastrophic events;
- building of safety patterns to seismic hazard in various urban samples, which will lead to a digital city map for evaluation of seismic vulnerability.

Through the analysis of urban seismic risk and with the help of seismic scenarios, the city stakeholders can create a risk management plan in order to define the actions that should take place in order to improve the quality of constructions, lifelines and the response of the institutions and people to a possible earthquake. This can be a key step in risk mitigation, with the consequence of reducing damages and human loses in case of a seismic event.

We organize the paper as follows: in section 2 we outline the methodology for the creation of seismic risk scenarios, in section 3 we present the three data mining classification algorithms, in section 4 we present and discuss the results of the algorithms, and then we show how a Geographical Information System (GIS) can be used for the visualization of vulnerability of buildings in a densly populated urban area of Iasi city, Romania.

## 2. CREATING SEISMIC RISK SCENARIOS

A fundamental principle of risk assessment is that risk due to natural hazards such as earthquakes, hurricanes and floods is location dependent. The process of risk assessment involves hazard assessment and vulnerability analysis. The probability of earthquake occurrence varies depending on location, and local site conditions also play a vital role in determining the intensity of the earthquake. Seismic risk scenarios are intended to provide local, state and regional officials with the tools necessary to assess the risks from earthquakes. Geographical Information Systems can be used for all the disaster prevention phases, and especially for physical planning in the mitigation phase, by taking the disaster risk into consideration.

In order to evaluate the damage state of a civil structure, a R.C. diaphragm structure with 5 floors located in the existing urban infrastructure, we developed a deterministic methodology based on a scenario of structural degradation. The scenario is built on *in situ* information collected through visual screening by technical experts.

Our research has been focused on one class of structure, typical for Iasi infrastructure, for which models are presented in figures 1, 2, and 3.

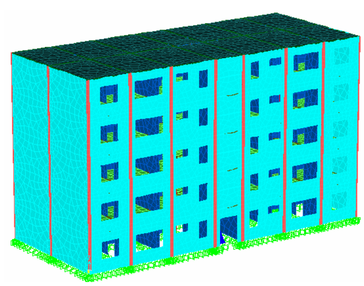

Figure 1 Structural model with no degradation

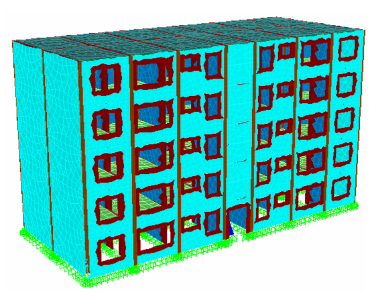

Figure 2 Structural model with small surface degradations

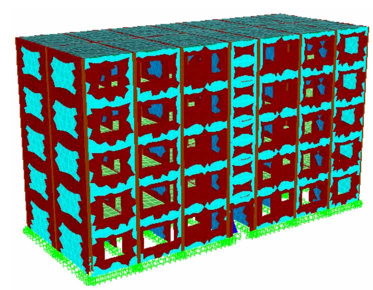

Figure 3 Structural model with largest surface degradations

The scenario consists in numerical experiments for the following cases and has been based on observations of technical experts:

- the degradation of the reinforced concrete diaphragm structure is modeled by decreasing the stiffness of the connection between walls and of the openings (doors, windows);
- the degradation of the reinforced concrete diaphragm structure is modeled by decreasing the stiffness of the connection between walls, of the walls and of the openings (doors, windows).

For each case a damage index (DI), called Maximum Softening by DiPasquale and Cakmak (1990), measures the overall damage of the structure and is given by:

$$DI = 1 - \frac{T_0}{T_{equivalent}} \tag{1}$$

where: $T_0$ is the initial natural period and $T_{equivalent}$ is the equivalent period of the softer structural model having less stiffness (elasticity modulus degradation).

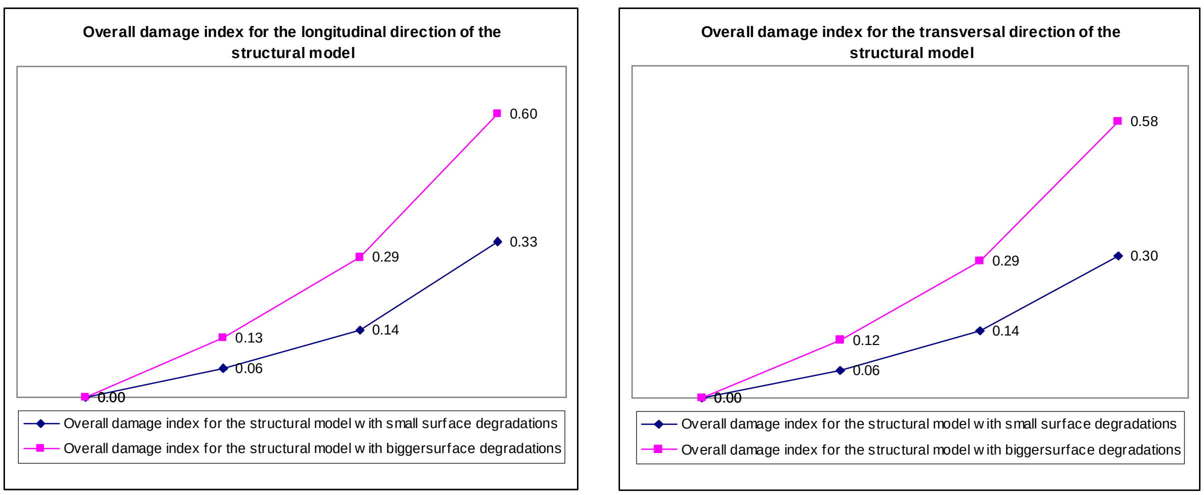


Figure 4 Overall damage index representation

## 3. DATA MINING TECHNIQUES FOR SEISMIC ANALYSIS

From the computer analysis, several data were exported on which data mining techniques were applied. These techniques were used to build a mapping model between the focal distance of an earthquake, its magnitude, direction of propagation, peak ground acceleration (PGA), and building degradation, as inputs, and the maximum displacement of the building, as output. The analysis was performed on each of the three cases described above: a building with no degradation, a building with small degradation, and a building with severe degradations, respectively.

This situation can be viewed as a typical classification problem. Classification is a procedure in which individual items (or objects, often referred to as instances) are placed into groups (or classes) based on quantitative information on one or more of their characteristics (referred to as attributes). Classification is a supervised technique, i.e. the model is built based on a training set of instances whose classes are known. Formally, given the training data $\{(\vec{x}_1, y_1), \ldots, (\vec{x}_n, y_n)\}$, $\vec{x}_i \in X$ and $y_i \in Y$, the goal is to determine a classifier $h: X \rightarrow Y$ which is a good approximation of the mapping of any object $\vec{x}_i$ to its classification label (or "class") $y_i$. More simply, the purpose of a classification algorithm is to find a hypothesis *h* from examples, i.e. pairs $(\vec{x}, f(\vec{x}))$, where $f$ is a function with a discrete codomain, such that $h \approx f$ holds not only for the given training pairs, but also for other arguments $\vec{x}_i$ from their domain. Thus, the information contained in the training set (with instances whose corresponding class labels are known) is used to classify new, previously unseen instances.

In order to test our results, we used three classification algorithms: the *C4.5* and *Random Tree* algorithms, which are decision tree inducers, and *Non-Nested Generalized Exemplars* (NNGE), which is a generalized instance-based approach, that can provide classification rules. We chose these classes of classification algorithms since their results are explicit and can be helpful for a civil engineer not only to classify new instances of verify existing situation, but also to understand the model and see in which way the variation of inputs affect the output.

C4.5 is an algorithm designed by Quinlan (1993), which generates a decision tree for the given dataset by recursive partitioning of data. The decision regarding the partition of the data on a node is made by considering an information theoretic measure of impurity of the resulting split, such as entropy. The goal is to create subsets as pure as possible, i.e. containing instances that belong to a majority class. In this way, the resulting tree has less test nodes, and therefore it is quite small. A further pruning operation can be performed after the initial classification model has been built, in order to increase the generalization capability of the decision tree, although often at the expense of a greater error rate.

The random tree classifier was proposed by Frank and Kirkby. It builds a tree that considers $k$ random features at each node and performs no pruning, therefore its error rate on the training set alone is rather small.

The simple Nearest Neighbor approach classifies a new instance based on the class of the closest existing instance in the training set. The k-Nearest Neighbor algorithm extends this approach from the closest instance to the closest k instances, which can increase the robustness of the model and can prove useful especially if the data is affected by noise. The Generalized Exemplars theory is an extension of the nearest neighbor classification methods, leading to a learning paradigm based on class exemplars, where an induced hypothesis has the graphical shape of a set of hyper-rectangles in an n-dimensional Euclidean space. Exemplars of classes are either hyper-rectangles or single training instances, i.e. points, known as “trivial” hyper-rectangles. A representative of this class of algorithms is NNGE (Martin, 1995). It tries to generalize new examples to their nearest neighbor of the same class, but if this is impossible due to intervening negative examples, no generalization is performed. If a generalization later conflicts with a negative example, it is modified, i.e. the existing hyper-rectangle is split to maintain the consistency of the model.

More information about these algorithms and their implementations can be found in the book by Witten and Frank (2000), and in an article by Leon, Zaharia and Gâlea (2004) which contains a detailed benchmarking analysis regarding the performance of several classification algorithms on different classification problems.

## 4. CLASSIFICATIONS RESULTS

Since the output must be discrete for all three classification algorithms, we discretized the maximum displacement into six classes: *Very Low* ($d_{max} < 0.08$), *Low* ($0.08 \leq d_{max} < 0.17$), *Medium Low* ($0.17 \leq d_{max} < 0.5$), *Medium High* ($0.5 \leq d_{max} < 0.8$), *High* ($0.8 \leq d_{max} < 1.1$), and *Very High* ($1.1 \leq d_{max}$).

The results provided by the C4.5 algorithm are displayed in figures 5, 6, and 7.

```
PGA <= 0.669: Low
PGA > 0.669
|   Focal-distance <= 100: MediumLow
|   Focal-distance > 100: Low
```

Figure 5 Decision tree for the "No degradation" case

```
PGA <= 0.528
|   PGA <= 0.368
|   |   Degradation <= 60: VeryLow
|   |   Degradation > 60: Low
|   PGA > 0.368: Low
PGA > 0.528
|   Degradation <= 60
|   |   Focal-distance <= 91: MediumLow
|   |   Focal-distance > 91
|   |   |   PGA <= 0.582: MediumLow
|   |   |   PGA > 0.582
|   |   |   |   PGA <= 1.095: Low
|   |   |   |   PGA > 1.095: MediumLow
|   Degradation > 60
|   |   Focal-distance <= 91
|   |   |   PGA <= 1.064: MediumLow
|   |   |   PGA > 1.064: MediumHigh
|   |   Focal-distance > 91: MediumLow
```

Figure 6 Decision tree for the "Small degradation" case

```
Degradation <= 60
|   Focal-distance <= 79
|   |   PGA <= 0.368: VeryLow
|   |   PGA > 0.368: Low
|   Focal-distance > 79
|   |   Focal-distance <= 91: MediumLow
|   |   Focal-distance > 91
|   |   |   Focal-distance <= 100: MediumLow
|   |   |   Focal-distance > 100
|   |   |   |   Direction = Longitudinal: Low
|   |   |   |   Direction = Transversal: MediumLow
Degradation > 60
|   Mw <= 6.4: MediumLow
|   Mw > 6.4
|   |   Direction = Longitudinal: High
|   |   Direction = Transversal
|   |   |   PGA <= 1.064: High
|   |   |   PGA > 1.064: MediumHigh
```

Figure 7 Decision tree for the "Severe degradation" case

The main advantage of using C4.5 algorithm is that the size of the trees is quite small and therefore the model is quite compact and easy to interpret. The main drawback is that the error rate on the training set can be rather high. In our case, the error rate are those presented in figure 8.

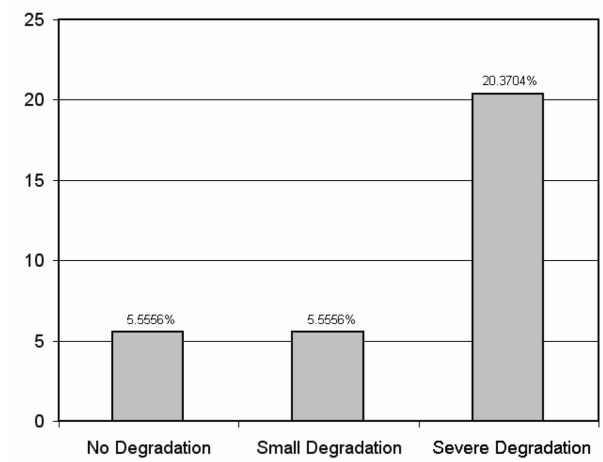


Figure 8 Error rates of C4.5 algorithm

Conversely, the error rate of Random Tree algorithm is 0% in all three cases, however the trees produced by random partitioning can be rather large. Also, because of the randomness involved in choosing the attribute for splitting, the algorithm can produce different trees on different runs. We can make an evaluation regarding the size of the tree by comparing the number of leaves in the C4.5 tree with the number of leaves of a corresponding random tree. The results are displayed in figure 9.

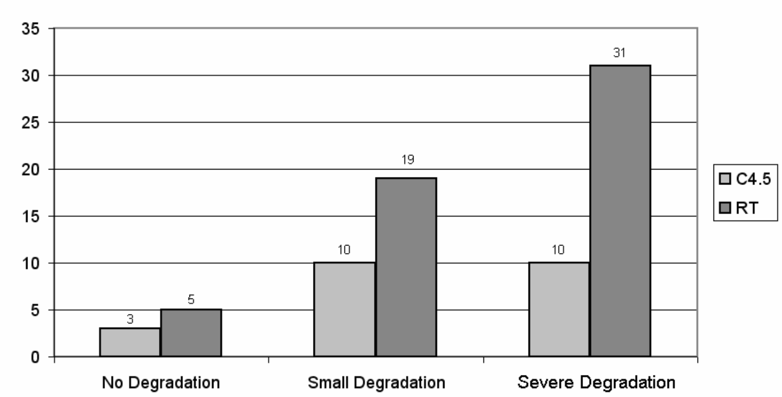


Figure 9 Comparison between the sizes of C4.5 trees and random trees

One can notice that the C4.5 produces an equal sized model for the "Severe degradation" case as for the "Small degradation" case. Since the former is more complex, the error rate of C4.5 is much higher. The Random Tree algorithm preserves the 0% error by increasing the size of its induced model.

The NNGE algorithm creates hyper-rectangles instead of decision trees. In urn, these hyper-rectangles can be interpreted as classification rules. For example, for the "No degradation" problem, the induced rules are as follows:

**CLASS VeryLow** : Focal-distance=79.0 AND Mw=6.4 AND Direction IN {Transversal} AND PGA=0.312 AND Degradation=0.0 (1)
**CLASS MediumLow** : 79.0<=Focal-distance<=91.0 AND 6.4<=Mw<=7.0 AND Direction IN {Longitudinal,Transversal} AND 0.76<=PGA<=1.356 AND Degradation=0.0 (7)
**CLASS Low** : 79.0<=Focal-distance<=100.0 AND 6.0<=Mw<=6.4 AND Direction IN {Longitudinal,Transversal} AND 0.368<=PGA<=0.658 AND Degradation=0.0 (6)
**CLASS Low** : Focal-distance=133.0 AND Mw=7.3 AND Direction IN {Longitudinal,Transversal} AND 0.641<=PGA<=1.464 AND Degradation=0.0 (4)

The number of instances that belong and create the corresponding rule are presented in brackets. For the "Small degradation" case, because of space, we only listed the rules with the greatest number of instances for each class, which also correspond to the most important decision regions:

**CLASS VeryLow** : Focal-distance=79.0 AND Mw=6.4 AND Direction IN {Transversal} AND PGA=0.312 AND 30.0<=Degradation<=60.0 (2)
**CLASS Low** : 79.0<=Focal-distance<=133.0 AND 6.4<=Mw<=7.3 AND Direction IN {Longitudinal,Transversal} AND 0.458<=PGA<=0.669 AND 30.0<=Degradation<=60.0 (10)
**CLASS MediumLow** : Focal-distance=91.0 AND Mw=7.0 AND Direction IN {Longitudinal,Transversal} AND 0.9<=PGA<=1.356 AND 30.0<=Degradation<=60.0 (12)
**CLASS MediumHigh** : Focal-distance=91.0 AND Mw=7.0 AND Direction IN {Transversal} AND PGA=1.095 AND Degradation=90.0 (1)

The same approach was followed for the "Severe degradation" case. The selected results are given below:

**CLASS VeryLow** : Focal-distance=79.0 AND Mw=6.4 AND Direction IN {Longitudinal,Transversal} AND 0.312<=PGA<=0.368 AND Degradation=30.0 (2)
**CLASS Low** : 79.0<=Focal-distance<=133.0 AND 6.4<=Mw<=7.3 AND Direction IN {Longitudinal,Transversal} AND 0.458<=PGA<=0.76 AND Degradation=30.0 (6)
**CLASS MediumLow** : Focal-distance=91.0 AND Mw=7.0 AND Direction IN {Longitudinal,Transversal} AND 0.9<=PGA<=1.356 AND Degradation=30.0 (6)
**CLASS MediumHigh** : Focal-distance=79.0 AND Mw=6.4 AND Direction IN {Longitudinal} AND 0.496<=PGA<=0.76 AND Degradation=90.0 (2)
**CLASS High** : 91.0<=Focal-distance<=133.0 AND 7.0<=Mw<=7.3 AND Direction IN {Longitudinal} AND 0.669<=PGA<=1.262 AND Degradation=90.0 (3)
**CLASS VeryHigh** : Focal-distance=133.0 AND Mw=7.3 AND Direction IN {Transversal} AND PGA=1.464 AND Degradation=90.0 (1)

In all three case, the error rate is 0%.

## 5. GEOGRAPHICAL INFORMATION SYSTEM

A key concept for the evaluation of vulnerability, developed primarily for seismic events, is the structural *fragility curve*. Fragility curves can be used for modelling the effects of a possible natural hazard event on structures , as a method of analyzing the behaviour of built existing infrastructure in urban areas under different hazard scenarios. The fragility curve is defined as the mathematical expression that relates the conditional probability of reaching or exceeding a particular structural safety level, given a particular level of the hazard (Leon & Atanasiu, 2006). HAZUS (FEMA, 1999) specifies the safety levels in terms of four damage states: slight, moderate, severe, and complete damage state. The GIS map of the vulnerability of buildings in a urban sample of Iaşi city, Romania, is displayed in figure 10. Given a certain seismic event, green stands for minor damage, blue means moderate damage, yellow represents major damage, and red stands for near-collapse.

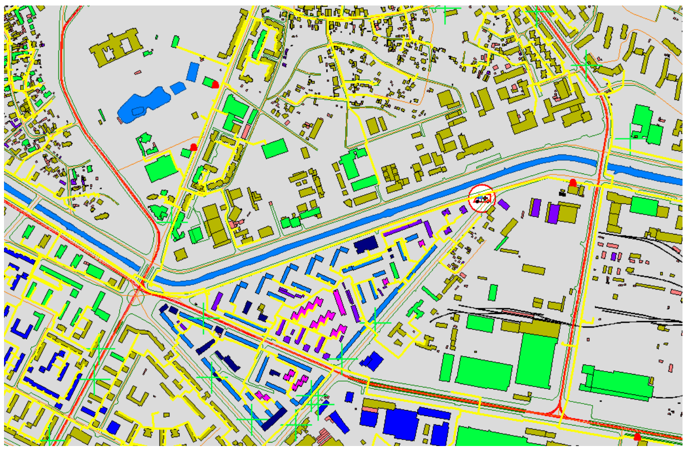

Figure 10 GIS map of the vulnerability of buildings

## 6. CONCLUSIONS

This paper describes a methodology of creating seismic risk scenarios. The results are analysed using different classification algorithms. Each algorithms has its own advantages and disadvantages, but their combination can help a stakeholder to make informed decisions about the vulnerability of a structure during a seismic event. The visualization of the building risk states can be made with the help of a Geographical Information System, GIS, which can give the decision makers a more intuitive inspection of the site under analysis, a useful characteristic especially for densely populated urban areas.

## REFERENCES

Atanasiu, G. M., Gâlea, D., eds. (2008). GIS Monitoring of the Urban Seismic Risk, Politehnium, Iaşi, Romania.

Atanasiu, G. M., Leon, F., Gâlea, D. (2006). GIS-Based Monitoring of Seismic Vulnerability for a Class of Concrete Structures during Lifetime Serviceability, *Proceedings of the European Symposium on Service Life and Serviceability of Concrete Structures, ESCS-2006*, 183-188, Concrete Association of Finland.

Bellifemine, F. L., Caire, G., Greenwood, D. (2007). Developing Multi-Agent Systems with JADE, Wiley Series in Agent Technology.

Bouhafs M., Si C., Lawson R. S., Bouabid J. (1997). GIS implementation of a nationwide seismic risk assessment methodology, ESRI International Users Conference, Paper 375; http://www.esri.com/library/userconf/proc97/proc97/to400/pap375/p375.htm.

DiPasquale E., Cakmak A. S. (1989). On the relation between local and global damage indices, Technical Report NCEER-89-0034, Princeton University.
FEMA (1999). HAZUS, Earthquake Loss Estimation Methodology, Technical Manual, National Institute of Building Sciences for the Federal Emergency Management Agency.
Forgy, C. (1982). Rete: A Fast Algorithm for the Many Pattern/Many Object Pattern Match Problem, *Artificial Intelligence*, **19**, 17-37.
Giarratano, J. C. (2002). CLIPS User's Guide, www.ghg.net/clips/download/documentation/usrguide.pdf.
Leon F., Zaharia, M. H., Gâlea, D. (2004). Performance Analysis of Categorization Algorithms, *Proceedings of the 8th International Symposium on Automatic Control and Computer Science*, Iasi, Romania.
Quinlan, R. (1993). C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, CA.
Witten, I. H., Frank, E. (2000). Data Mining: Practical machine learning tools with Java implementations, Morgan Kaufmann, San Francisco.
Wood, N., Stein D. (2001). A GIS-based vulnerability assessment of Pacific Northwest ports and harbors to tsunami hazards, ITS 2001 Proceedings, Sessions 1, Number 1-13, 367-374.