

Seismic vulnerability assessment using Support Vector Machine classification for remote sensing and in-situ Data

Matsuka Panagiota (1,2), Pathier Erwan (1) , Gueguen Philippe (1) & Chanussot Jocelyn (2)

*ISTerre, University Joseph Fourier Grenoble, CNRS/IFSTTAR, France (1)
Grenoble Images Parole Signal Automatique (GIPSA-lab),France (2)*



SUMMARY:

A new framework for evaluating seismic vulnerability classes at urban scale on a building basis is proposed in this paper. The proposed methodology utilizes the capabilities of remote sensing techniques combined with information from in-situ evaluation by expert of a small percentage of buildings. The end scope is an estimated vulnerability map classifying all the buildings according to the EMS98 vulnerability classes, providing initial information to decision support system for seismic risk in the context of vulnerability assessment. We tested the methodology in the city of Grenoble (France), using only 2 building attributes from remote sensing: roof type and building height (extracted from a very high resolution image, and an 1-m-resolution digital elevation model). Results show that with only these 2 attributes it is difficult to classify the building in the 4 existing classes. However using two classes (merging EMS98 classes A and B and classes C and D) give satisfactory results.

Keywords: support vector classification, remote sensing, seismic vulnerability classes, in-situ data

1. INTRODUCTION

Assessment of building seismic vulnerability, such as an urban spatial information basis, can provide an important guide for decision-makers to develop mitigation strategies. In addition it can help to raise public awareness of risks in the forefront of an expected disaster. Vulnerability can be defined by a mixture of different components including physical, demographic, social, economic, ecological and political aspects. For seismic vulnerability assessment, several methodologies have been developed depending on the resources and the scale of assessment. Most of methods developed for assessment at the urban scale (e.g., HAZUS, GNDT or RiskUE methods) were calibrated on post-seismic observations. Building vulnerability matrix were defined, by crossing some building attributes (e.g., material type, building location and foundation, age of construction, type of roof, plan regularity, elevation regularity, position of the structure in the block) and the damage level observed for a given ground motion intensity (e.g., macroseismic intensity). In order to assess the seismic vulnerability at urban scale is due to the variability of the building characteristics, and the number of buildings concerned that may require important resources.

The development of fast and automatic methods for collecting building attributes involved in seismic vulnerability is of major importance. For that, remote sensing can be an effective tool for providing automatically building attributes that can be used for seismic vulnerability studies in a large space within urban areas. For instance, Mueller et al. (2008) gave example of potentials of remote sensing contribution to the identification of physical parameters of the buildings used for vulnerability. Valero et al. (2008) focused on the estimate of the nature of the roof for each building by using remote sensing data. Polli et al. (2009) in a recent work attempt extraction of required parameters through analysis of very high resolution remote sensing images.

In situ visual analysis of buildings can be tedious work in large scale areas, while remote sensing can provide faster and cheaper solutions. . A lot of different parameters are taken into account in order to assess seismic vulnerability. That is why there is not any potential for RS to substitute existing,

available, vulnerability assessment methodologies, but there is strong potential for supporting and providing faster as well as smarter solutions for specific building attributes. Combining In-situ data with remote sensing data can be an efficient solution for improving the seismic vulnerability study. The motivation of this paper arises from the combination of in-situ data and information from remote sensing. A non linear support vector classification, is applied using in-situ and remote sensed data. Classification of seismic vulnerability classes according to the the standard EMS98 is applied in buildings with in-situ data calculated by method VULNERALP (Gueguen et al. 2007) tested in Grenoble (France). Based on remote sensed data and corresponding in-situ measurements the use of support vector machines generates a relation between them, which can subsequently be used to classify unknown building vulnerability classes from additional remote sensing data.

The general framework we propose is presenting in Figure1. Section 2 presents SVM classification. The data sets and the feature extraction are presented in Section 3 The experimental results are given in Section 4. We conclude in Section 5.

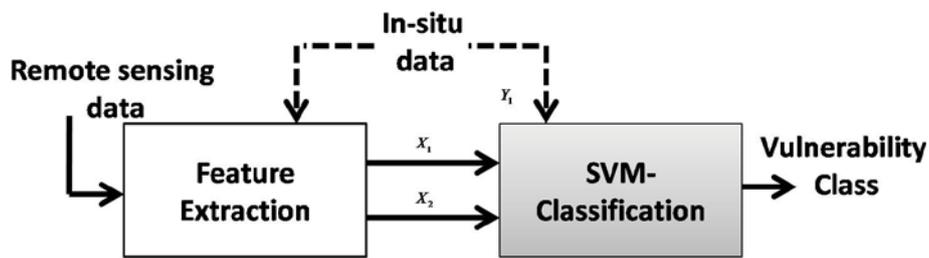


Figure 1.The general flowchart of the proposed methodology is described in the above figure. The first step includes the extraction of relevant attributes from remote sensing data and the second step applies support vector classification.

2.METHODOLOGY

Within a supervised classification framework, we use a support vector machines (SVM) statistical learning algorithm to label the buildings according to the desired EMS98 standard for seismic vulnerability classes. We assume the existence of a relation between remote sensed data and in-situ seismic vulnerability data. Based on this relation we adapt a supervised classification scheme based on SVM to classify the seismic vulnerability over the buildings. The availability of a VHR imagery allows us a fine resolution on a building basis analysis on an urban area. We assume features related with seismic vulnerability extracted from remote sensed data (roof type and building height) described in Section 3.2. A pixel is related by a feature/vector $x_i^d \in \mathfrak{R}$ where each component corresponds to a particular extracted feature, d represents the number of different features extracted from remote sensing. Seismic vulnerability classes correspond to values y_i . Given a set of training samples $(x_i^d, y_i)_{i=1}^n \in \mathfrak{R} \times \mathfrak{R}$ with known corresponding seismic vulnerability classes, SVM generates an approximation of the relation between the remote sensing data (x_i^d) with the corresponding seismic vulnerability class y_i . We build a supervised SVM classification model (Burges, 1998) such that it accurately classify the outputs y corresponding to a new set of input examples $x_j^d, j \neq d$.

Since SVMs are covered in a large number of recent papers, we give a brief introduction to SVMs, for a more systematic description interested readers may refer to Christianini et al.(2000), Burges (1998) and Vapnik (1995).

Linear SVMs

For simplicity, we consider a supervised binary classification problem. Let us assume that the training set consists $x_i^d \in \mathfrak{R}(i=1,2,\dots,n)$. Each vector x_i^d is associated with a target $y_i \in \mathfrak{R}\{-1,+1\}$. The linear SVM classification aims of finding the hyperplane that maximizes the margin (i.e the distance to the closest training data points in both classes). The hyperplane H_p is defined as:

$$w \cdot x + b = 0, \forall x \in H_p \quad (2.1)$$

where $w \in \mathfrak{R}^d$ defined as is a vector normal to the hyperplane, $b \in \mathfrak{R}$ is a bias and $w \cdot x$ is the dot product between w and x . If $x \notin H_p$ then $f(x) = w \cdot x + b = 0$ is the distance from x to the hyperplane. Therefore such as hyperplane has to be define according to:

$$y_i(w \cdot x_i + b) > 1, \forall i \in [1, N] \quad (2.2)$$

Also the optimal hyperplane has to maximize the margin: $2/\|w\|$ which leads to the criterion of error minimization:

$$\min \left[\frac{\|w\|^2}{2} \right], \text{ with respect to (2.2)} \quad (2.3)$$

For non-linearly separable data, slack variables ξ are introduced and Eq. (2.2) becomes:

$$y_i(w \cdot x_i + b) > 1 - \xi_i, \xi_i > 0, \forall i \in [1, N] \quad (2.4)$$

which leads to the final optimization problem:

$$\min \left[\frac{\|w\|^2}{2} + C \sum_{i=1}^N \xi_i \right], \text{ with respect to (2.4)} \quad (2.5)$$

where C constant controls penalty errors. The final equation leads to a quadratic programming solution (Vapnik, 1995).

The membership decision rule is based on the sign function and the classification is done by

$y_{new} = \text{sgn}(w \cdot x_{new} + b)$ where (w, b) are the hyperplane parameters found during the training process and x_{new} is an unseen sample.

Multiclass SVMs

SVM are intrinsically binary classifiers but for remote sensing several classes are usually of interest. Different multiclass classification strategies can be adopted C.W.Hsu et al. (2002). In our experiments we adapt the pairwise classification to solve an m -class problem.

The pairwise classification is the case where $\frac{m(m-1)}{2}$ binary classifiers are applied on each pair of classes. Each sample is assigned to the class getting the highest number of votes. A vote is defined for each class as a classifier assigning the pattern to that class.

Nonlinear SVMs

Kernel methods are a generalization of SVMs providing nonlinear hyperplanes and thus improving classification abilities.

Input data are mapped into a higher dimensional space H by using a nonlinear function Φ such as:

$$\begin{aligned}
\mathfrak{R}^d &\rightarrow H \\
x &\rightarrow \Phi(x) \\
x_i \cdot x_j &\rightarrow \Phi(x_i) \cdot \Phi(x_j).
\end{aligned}
\tag{2.6}$$

The expensive computation of $\Phi(x_i) \cdot \Phi(x_j)$ in H is reduced using the kernel trick Scholkopf et al. (2005)

$$\Phi(x_i) \cdot \Phi(x_j) = K(x_i, x_j) \tag{2.7}$$

The selection of kernel K has to satisfy the Mercer's condition (Burges, 1998). There exist several choices of kernel function K . In this work we test and present the SVM classification using Gaussian kernels.

$$GaussK(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \tag{2.8}$$

The tuning of the SVM requires some parameters (kernel parameter γ , constant C that controls penalty errors) to be well fitted before we run the algorithm.

3. DATA SETS AND FEATURES FROM REMOTE SENSING

3.1.Data set

For several years, Grenoble has been as a test site for estimating regional and local risk, since is one of the most high risk cities in French Alps region because of the number of inhabitants, the economic activity, the presence of high-tech companies and industrial facilities (chemical and nuclear). Those reasons and the availability of in-situ and remote sensing data motivated the development and validation of this methodology over Grenoble.

The available remote sensing data sets include a very high resolution (VHR) orthorectified panchromatic image (airborne data, 25 cm resolution) and a digital elevation model (DEM) (airborne acquisition, 1m resolution in three dimensions). These two modalities are used to provide related vulnerability attributes. The available in-situ data includes building footprints and seismic vulnerability class according to the vulnerability EMS-98 standard. The EMS98 classes in Grenoble were defined using visual inspection of existing buildings and following the Grenoble Building Typology defined for the Vulneralp project (Gueguen et al., 2007; Michel et al., 2012). Fig3.1 shows the test area of Grenoble, where in Fig3.1 (a) we present the location of the test area in the city of Grenoble and in Fig3.1 (b) we show the small area we focused on (test area) with the available seismic classes per building and in the background the VHR image. The test area chosen because it shows a mix of buildings typologies representative of the Grenoble metropolitan area. Table3.1 shows the number of buildings in each seismic vulnerability class in the test area presented in Fig3.1(b).

Table 3.1. Number of buildings for the four seismic vulnerability classes (A,B,C,D) found in the in situ data in the test area of Grenoble (see Fig.3.1)

Seismic Vulnerability class(EMS 98)	ClassA	ClassB	ClassC	ClassD	SUM
Buildings Number	125	233	86	40	484

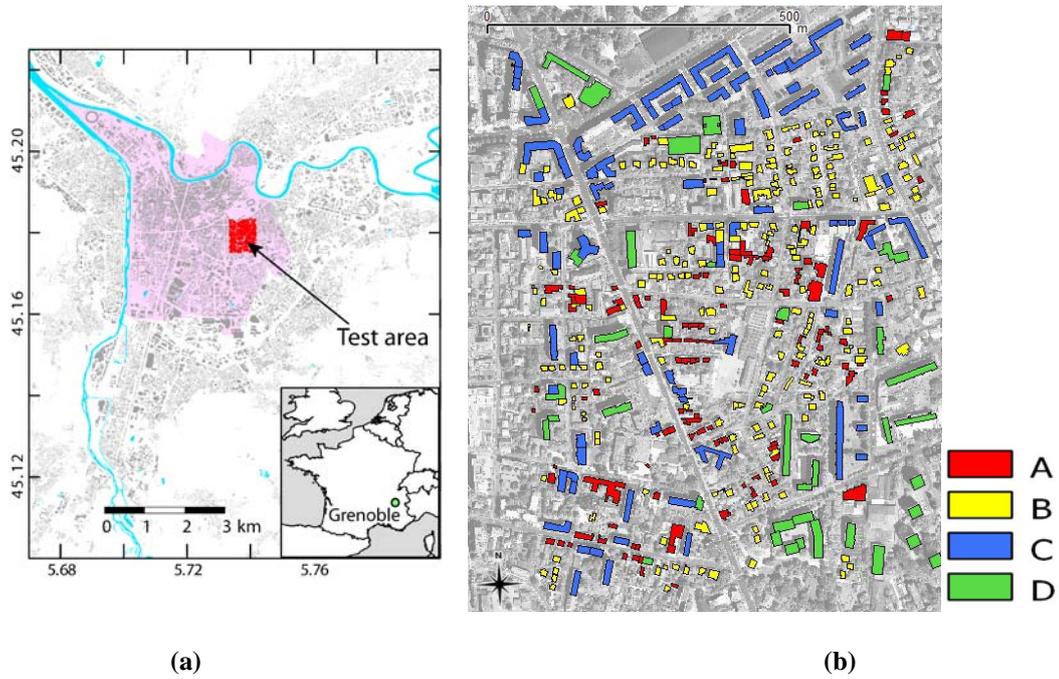


Figure 3.1. (a) The figure shows the location of the test area, in red, located within the city of Grenoble. The pink area correspond to the administrative limits of the city of Grenoble (156. 000 inhabitants) and the grey polygons are building footprints of the Grenoble metropolitan area (500.000 inhabitants). Rivers or small lakes are in blue. (b) VHR orthoimage with available footprints and corresponding in situ data. The colorbar represents the different seismic vulnerability classes according to the EMS98 standard (class A, class B, class C, class D) and they represent the in situ information provided by the experts. The area represents a small area of the Grenoble city.

3.2. Feature extraction from remote sensing

In this section we introduce the attributes extracted from remote sensing. For this study two different features related with seismic vulnerability were selected:

1) Roof type: To identify the roof of the building we implement the methodology described in Valero et al. (2008). The aim here is to discriminate between flat roofs and gable by fusing VHR panchromatic image and DEM and by using available building footprints.

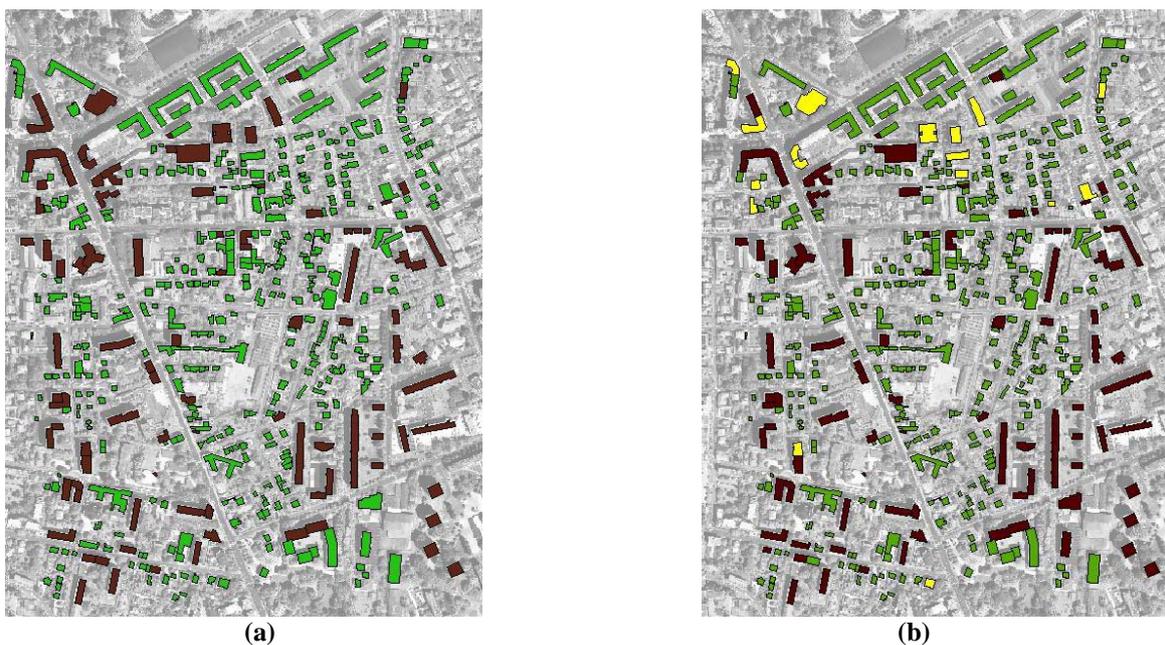


Figure 3.2. (a) In-situ data available for the roof. Brown colour correspond to gable roofs and green to flat. (b)

Classification results after implementing the methodology described in Valero et al. (2008). Yellow colour corresponds to miss classification cases.

2)Building Height: The height of the building is calculated by using the DEM and the building footprints. We estimate the ground level from the pixels of the DEM that are outside building footprints and each building height correspond to the difference between the ground level and the average median of the elevation of the pixels inside of each building footprint. All the features are calculated on a building base.

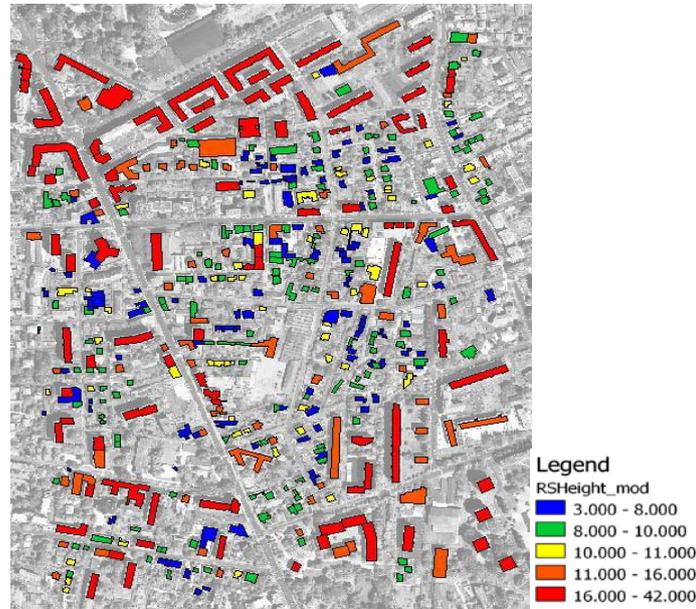


Figure 3.3. Height calculation by using the digital elevation model as described above.

Table 3.2. Probability table between remote sensing features (height and roof) extracted as described in Section 3.2 and in-situ data (seismic vulnerability classes) all over the test area of Grenoble.

Elevation	Roof	P(A)	P(B)	P(C)	P(D)
Between 3-8 m	Flat	0.2500	0.2500	0.1667	0.3333
	Gable	0.3067	0.3067	0	0.1250
Between 8-11 m	Flat	0	0	0.0202	0.3733
	Gable	0.2586	0.3485	0.6267	0.0603
Between 11-42	Flat	0	0	0.6267	0.3733
	Gable	0.2586	0.4224	0.2586	0.0603

Correction on the building footprints has been applied by using erosion to the individual building footprints. Roof identification and building height selected to be extracted from remote sensing.

In order to implement the above methodology (roof type and building height extraction) the building footprints are needed. Techniques for extracting building foot-prints using remote sensing are mainly based on edge and line primitives detection from optical image LIU et al. (2005) is not a trivial work and it is beyond the scope of this paper. In our case we used building footprints extracted from a GIS vector data base IGN, BDTPOPO. (2009).

We evaluate the performance of the roof classification using available in-situ data. Those two features can be calculated fast and with a satisfactory accuracy by implementing several techniques according with the availability of the remote sensing data. Finally in Table3.2 we present the probability table between remote sensing features extracted as described above and seismic vulnerability classes (A,B,C,D) available from insitu data by the experts. Elevation and Roof values correspond to all of the buildings in the test area shown in Figure3.2(b). Probabilities $P(A)$, $P(B)$, $P(C)$, $P(D)$ are

calculating as $P(i) = \frac{N_i}{N_T}$, $i = A, B, C, D$, where N_i is the number of buildings belong in class i and

N_T is the total number of buildings. Using the information extracted from remote sensing we continue to the classification scheme and our experiments.

4.CLASSIFICATION SCHEME AND EXPERIMENTS

After feature extraction from remote sensing we execute SVM. We generated separate training and testing data (training data are not included in testing data). We use the following notation: n_{tr} is the training number of buildings, and n_{te} is the number of testing buildings. $x_i^d \in \mathfrak{R}$ denotes the feature vector extracted from remote sensing, $i \in 1, \dots, n_{tr}$ and $d \in 1, 2$, where d represents the number of the extracted features (roof, height) and $y_i \in \mathfrak{R}$, $i \in 1, \dots, n_{tr}$ correspond to the in-situ seismic vulnerability class. The SVMs were computed using SVM and kernel methods Matlab toolbox (Canu et al, 2005), and the program was modified to include cross validation. During the training process the kernel parameter adjusted to maximize the overall accuracy, which was computed using 5-fold Cross validation. The experiments were repeated ten times (with ten independent training subsets) and the mean accuracy values were reported.

For the classification experiments we used $d = 2$, and tested using different training buildings: $n_{tr} = 22, 43, 87, 130, 174, 260, 347$. (example: for 5% as training samples we have $n_{tr} = 22$ and $n_{te} = 412$, see Table3.1 total size of available building is 434).

Table 4.1. Classification accuracies in percentage for several training set size

Accuracy Vs training buildings	5%	10%	20%	30%	40%	60%	80%
OA	55.97	58.53	60.47	60.27	61.17	61.66	62.50
Kappa	29.14	32.43	34.81	34.44	35.59	35.96	36.86

Table 4.1 summarizes the results obtained using the Gaussian kernel. These values extracted from the confusion matrix. The overall accuracy (OA) is the percentage of correctly classified pixels. Kappa coefficient is another criterion classically used in remote sensing classification to measure the degree of agreement. It is the percentage agreement corrected by the level of agreement that could be expected due to chance (Fauvel, 2007). The classification accuracies are between 55.97 for 5% to 62.50 for 80% of training data set.

Table 4.2. Confusion matrix for classification of the seismic vulnerability classes A,B,C,D (EMS98) over theGrenoble test area using remote sensing and 10% of in-situ data

Insitu data	SVM classification using remote sensing data						Commission error %
		Class D	Class C	Class B	Class A	Sum	
Class D	9	21	5	1	36	75.00%	
Class C	11	53	11	2	77	31.17%	
Class B	2	6	181	20	209	13.40%	
Class A	1	1	99	11	112	90.18%	
Sum	23	81	296	34	434		
Omission error %	60.87%	34.57%	38.85%	67.65%		Overall Accuracy 58.53%	

Table 4.2 presents the confusion matrix by 10% of training data set where the OA is 58.53%. The differences between the data provided by experts (in-situ) and the SVM classification using remote sensing data and 10% of training data are quantified using commission error, defined as the percentage of buildings incorrectly assigned to each damage grade, and, omission error, defined as the percentage of buildings incorrectly omitted from each vulnerability class. Considering the errors for each vulnerability class, class B and class C shows the smallest errors with commission errors 13.40%, 31.17% respectively and omission errors 38.85%, 34.57%. However the commission and omission errors for the other classes are significantly increased, especially for Class A with commission error of 90.18% (i.e. 90.18% of the buildings identified in Class A were not in Class A) and an omission error of 67.65% (i.e. 67.65% of the actual buildings in class A were classified in other classes). Those results indicate the difficulty to distinguish buildings with high vulnerability class by using only roof and height from remote sensing. Additionally comparing the number of buildings in class A (11 from remote sensing vs. 113 from in-situ data) it shows that remote sensing observations underestimate the numbers of buildings belong in Class A. This result is due to the fact that very vulnerable buildings belong in class A may not be visible from the aerial view of the remote sensing data by using only the roof and the height information.

Table 4.3. Confusion matrix for classification by merging seismic vulnerability (EMS98) class A with B and class C with D over the area of Grenoble using remote sensing and 10% of in-situ data

Insitu data	SVM classification using remote sensing data		SUM	Commission error %
	Class D&C	Class B&A		
Class D&C	94	19	113	16.81%
Class B&A	10	311	321	3.11%
SUM	104	330	434	
Omission Error %	9.61%	5.75%		Overall Accuracy 93.32%

If the data are combined such that class A and Class B are combined into a single category and class C and Class D in another category, the overall accuracy increase to 93.32% and the commission and omission errors are with less value at 3.11% and maximum at 16.8% as shown in Table 4.3. The Table 3.1 shows the classification accuracies after merging the data into two classes using different training buildings.

Table 3.1. Classification accuracies in percentage for several training set size after merging class A with B and class C with D

Accuracy Vs training buil.	5%	10%	20%	30%	40%	60%	80%
OA	91.97	93.32	93.28	93.22	93.47	94.30	94.79
Kappa	78.64%	82.19%	82.24%	82.02%	82.86%	84.86%	86.30%

The area of Grenoble has been tested using SVM classification and extracting roof and height from remote sensing data. The availability of the ground truth by the experts offered a great way to evaluate this methodology quantitative (section 3) and qualitative (section 4).

5. CONCLUSIONS

Primary results using only two attributes from remote sensing (roof type and building height) are encouraging and support vector machine classification appears to be promising. The high correlation between remote sensing features and in-situ data has a strong impact on support vector classification performance. An investigation to further improve our results includes the use of other possible features

extracted by remote sensing (such as the volume of the building, area, age, material, building irregularity) and analyze their contribution. Future work will include also the extension of our methodology in the vulnerability assessment in homogeneous areas based on the existing results in a building basis.

ACKNOWLEDGEMENT

This work has been supported by the ANR national research agency as part of its RiskNat program (URBASIS project, no. ANR-09-RISK-009).

REFERENCES

- Burges C.J.C., (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*. **Vol.2**: pp. 121-167
- Canu S., Grandvalet Y., Guigue V, Rakotomanonjy A.. (2005). SVM and Kernel methods Matlab Toolbox. Perception systemes et Information, INSA de Rouen, Rouen, France
- Christianini, N. and Shawe-Taylor, J. (2000). An introduction to Support Vector Machines and other Kernel-Based Learning Methods. Cambridge University Press . <http://www.support-vector.net>
- Fauvel M. (2007). Spectral and Spatial Methods for the classification of urban remote sensing data. Doctoral thesis.
- Guéguen P, Michel C, Le Corre L.(2007) A simplified approach for vulnerability assessment in moderate-to-low seismic hazard regions: application to Grenoble (France). *Bulletin Earthquake Engineering*5(3):467-490
- Hsu C.W. and Lin. C.J. (2002), A comparison of methods for multiclass support vector machines. *IEEE Trans. Neural Netw.*, **Vol 13**, no.2, pp.415-425
- IGN-(French National Geographic institute) (December 2009), BD TOPO VERSION2 www.professionnels.ign.fr
- LIU.W.L.Z and WANG.J, (2005) Building extraction from high resolution imagery based on multi-scale object oriented classification and probabilistic hough transform. *Geoscience and Remote Sensing Symposium. IGARSS*, **Vol.4**, pp. 2250-2253
- Michel, C., Guéguen, P., and Causse, M. (2012). Seismic vulnerability assessment to slight damage based on experimental modal parameters. *Earthquake Engineering & Structural Dynamics* **41(1)**, 81-98.
- Mueller M., Segl K., Heiden U. and Kaufmann H.. (2008). Potential of high-resolution satellite data in the context of vulnerability of buildings. *Natural hazards* **Vol.38**:pp247-258.
- Polli D., F.Dell'Acqua and P.Gamba. (2009). First steps towards a framework for earth observation (EO)- Based seismic vulnerability evaluation. *Environmental Seismiotics*, **Vol.2**, no.1, pp16-30
- Scholkopf B. and Smola A.J. , (2002) Learning with kernels, MIT Press.
- Valero S., Chanussot J. and Gueguen P.. (2008). Classification of basic roof types based on vhr optical data and digital elevation model. *Geoscience and Remote Sensing Symposium. IGARSS. IEEE International*, **Vol.4**: pp. IV-149
- Vapnik, V.N. (1995). The nature of statistical learning theory. Springer