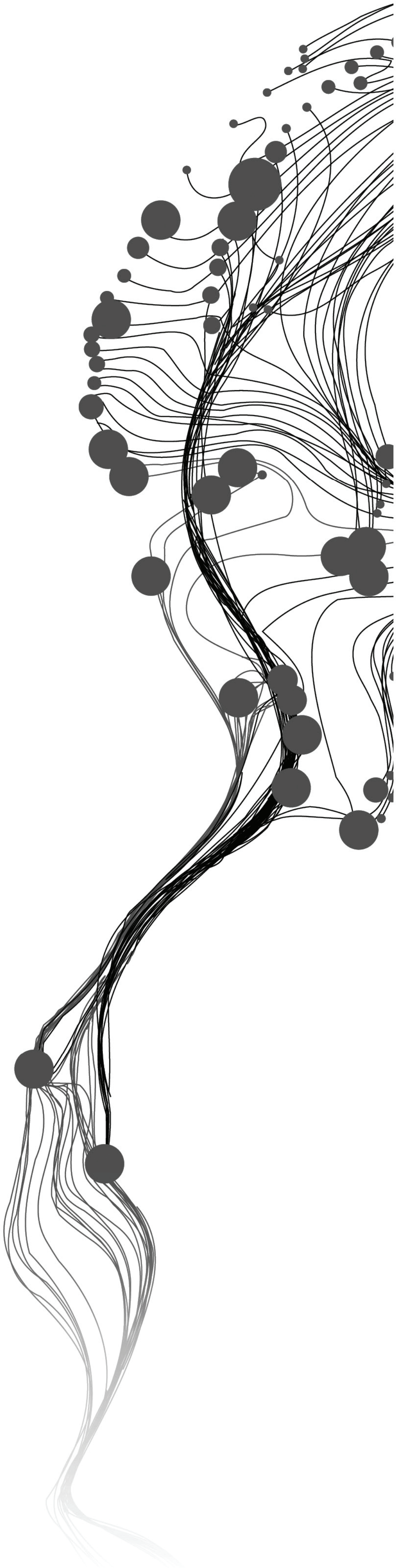
Using statistics to optimise the detection of collapsed building from laser scanner data

S.A.MAJID FATHI
February 2011

SUPERVISORS:
Dr. Ir. S.J. Sander Oude Elberink
Prof. Dr. Ir. M.G. George Vosselman



Using statistics to optimise the detection of collapsed building from laser scanner data

S.A.MAJID FATHI
Enschede, The Netherlands, [February, 2011]

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Geoinformatics

SUPERVISORS:
Dr. Ir. S.J. Sander Oude Elberink
Prof. Dr. Ir. M.G. George Vosselman

THESIS ASSESSMENT BOARD:
Chair: Prof.Dr.Ir. M.G. Vosselman
External examiner: Dr. K. Khoshelham
Supervisor: Dr.Ir. S.J. Oude Elberink
Second supervisor: Prof.Dr.Ir. M.G. Vosselman

ABSTRACT

In the present study, a method for the detection of collapsed buildings from post event Lidar data is presented.
Strong earthquakes require extensive and immediate field investigation to record damage patterns. The investigations of the collapsed building and its spatial distribution after an earthquake are of primary importance for planning the rescue activities and for evaluating the level of damages in affected area. Effective disaster management requires real-time data to various decision makers.

Airborne Lidar scanner (ALS) as a state- of -the art technique is capable of delivering large amounts and very accurate point clouds of our interested area in a relatively short time. ALS data is a suitable technique as a basis for damage analysis because it can be acquired directly after a disaster, independent of weather conditions and during days and nights. However, considering the amount of captured data, the automatic detection and interpretation of ALS data remains a challenge to several scientists in the field. Up to the present, a wide variety of algorithms for processing of ALS data has been already introduced and developed and nowadays, the extraction of classes such as buildings, vegetation, etc. is interesting for many applications in Geomatics.

One approach for classifying ALS data is to employ machine learning techniques, for which many statistical methods and tools are applicable. This research has been conducted to assess the capability of maximum entropy (Maxent) approach to automate the detection of collapsed buildings from ALS point clouds after an earthquake. Maxent can be considered as a new method for one-class classification. The output of Maxent is the probability distributions of the introduced class. Post event Lidar data of Haiti with a density of 2 points per square meter was used after segmentation of the planar surface in a region growing algorithm, to calculate some features as input predictors for our classification. In this study, 281 collapsed building records have been used to train and evaluate the classifier. The classification for 8516 points was done using Maxent. The importance and contribution of each and every variable was calculated by Jacknife test. The model was evaluated using one threshold independent technique, the area under receiver operating characteristic curve (AUC), and two threshold-dependent techniques, Kappa and true scale statistics (TSS).

According to the results, the most important variable is the number of points per segment, which suggests the size of a segment contains useful information. The results also showed that the ratio of unsegmented points to the segmented points, and the distance to DTM were the second and third important variables, respectively. The average behaviour of Maxent in 30 bootstrap simulations using all features revealed that some features (e.g. Density of points in 2D, density of points in 3D and residuals to planarity) had the least predictive power. The evaluation of Maxent suggests that this technique can be considered as a fairly accurate model to detect the collapsed building in a one-class classification problem.

Keywords: Maximum entropy, Maxent, classification, Lidar, collapsed building

ACKNOWLEDGEMENTS

This research study would not be successfully completed without the help and encouragement of people, who deserve my acknowledgement. I would like to take this opportunity to appreciate the people who helped me during this thesis research.

First, I would like to express my special thanks to my first supervisor **Dr. Ir. S.J. Sander Oude Elberink**. Thanks for your guidance during my work; also I appreciate your encouragement and constructive comments.

And also, I am thankful to my second supervisor **Prof. Dr. Ir. M.G. George Vosselman**. It was my pleasure to have you in my supervision team. I do appreciate your supports and also your encouragement. Thanks for your invaluable comments and guidance.

Also I would like to appreciate all staff and faculty members of GFM department especially Mr. Huurneman for their invaluable help during my study in ITC.

My Special thanks go to Babak Naimi (PhD candidate) for all his help and supports. I really appreciate your invaluable guidance during the process of my thesis.

I am so thankful to my international colleagues in (GFM 2010-2011) for their kindly friendship during the 18 months of study in ITC. It was great experience to study with nice people from many countries all around the world.

And finally, I would like to express my very special thanks to my lovely Wife, Nasrin, for all her nice and kindly supports during this 18 month study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Due to the possibility of acquiring precise data of large areas rapidly, airborne laser scanning systems seem to be suitable for obtaining information about the damaged area immediately after a natural disaster like earthquake in large scale. This research aims to present and optimize a technique for detection and classification of damaged (collapsed) buildings in affected areas.

1.1. Motivation and problem statement

Disaster Management issues, have become important worldwide issues. For time-critical situations, fast decision making could save many lives. In order to support search activities and achieve an efficient use of the available resources for rescue and saving as many trapped persons as possible, a fast and extensive damage analysis is required. Generally, remote sensing data analysis procedures prove to be an excellent method in terms of benefits to emergency responders. Airborne Laser Scanning (ALS) shows an excellent performance for such environments like earthquakes due to fast and geometrically precise 3D data. While 3D data are valuable but there are some difficulties in analyzing these data to identify and detect destroyed or damaged buildings (for example misclassification due to similar characteristics of vegetation and debris (surface roughness)). Although ALS datasets are very precise and accurate, processing manually of these datasets is not easy, because the size of ALS datasets is very huge. Therefore an automatic method for detection of collapsed buildings in ALS datasets is necessary. There are some methods for analyzing pre-event and post-event data together in remote sensing but sometimes pre-event are not available at that time or has not been acquired before. Therefore a method that can analyze the amount of damage just by post-event data is necessary.

On the one hand, problem of damaged building detection can be considered as a one-class classification problem. On the other hand, although there are many techniques to classify a multiple class problem, the only method that has been used for one-class classification was supported vector machine (SVM). Therefore, investigation to explore and develop of new methods for the one-class classification seems necessary.

1.2. Research objectives

The main objective of the proposed research is to develop and optimize an automatic damaged building detection and classification method by analyzing the post event captured ALS data. This research comprises the following specific objectives:

- To detect the damaged buildings using an appropriate classification method.

- To optimize the model (classifier) by selecting the more relevant features as inputs for classification.
- To evaluate the capability of maximum entropy approach for one-class classification of damaged buildings using ALS data.
- To evaluate the potential of post-event ALS data to detect collapsed buildings.

1.3. Research questions

To achieve the specific objectives, following research questions need to be answered:

- What are the most important suitable features for distinction of damaged buildings?
- Whether maximum entropy approach can be used as a one-class classifier to detect the damaged buildings based on ALS data?
- In which level of accuracy the damaged buildings can be detected by using the Maxent approach?

1.4. Structure of thesis

To fulfil the objective and question mentioned above, this thesis is divided in to six chapters

Chapter 1-introducion

Introduction to disaster management and application of Lidar for rescue activities also objectives and research questions are described.

Chapter 2- Research Background

Review of related work in the literature regarding Methods of Classification

Chapter 3-Methods and Materials

Description and Methodology of classification by Maxent (Maximum entropy approach) and its implementation are described.

Chapter 4- Results

This chapter dedicated to analyze the achieved results of Maxent implementation.

Chapter 5- Discussion

In this chapter the achieved result are discussed and will be compared with other existing results.

Chapter 6- Conclusion and recommendation

This chapter provides the conclusion of this research and gives some recommendation for improvement and future works

2. RESEARCH BACKGROUND

A Lidar sensor delivers 3D point clouds with some attributes; Extensive post-processing is required to extract accurate terrain or semantic information from the Lidar point clouds. Lidar provides data in the form of 3D point clouds while the users often require specific information on land cover. This information can be obtained by segmentation which is following by a classification. In fact, classification assigns label of a land cover class to segments.

2.1. Classification

A wide variety of algorithms for interpretation, modelling and classification of the point cloud are employed to create additional information of acquired data. Some of these methods are using the point cloud, while others perform classification based on segments, some are using only geometry and some are using additional information. Filin (2004) presented a feature space classification by using surface texture and variation in surface trend. The position of point, the parameters of the tangent plane to the point and the relative height difference between the point and its neighbours can be used as features for classification. This method uses the point cloud directly. Rutzinger et al. (2008) used full-waveform capabilities for classification of point cloud for detecting of trees. In this method, the echo strength and the width of the pulse are used for classification. Darmawati (2008) proposed a method for classification of point cloud by using multiple echo information. They used the number of points in the segment, percentage of echo and topological relations as feature for classification. Vosselman (2009) used three segment attributes including the number of points in a segment, the average height of a segment above a local minimum height and the percentage of last echo points in a segment to classify the segments.

2.1.1. Segmentation

2.1.1.1. Scan line segmentation

Scan line segmentation method was proposed for segmentation of point cloud (Sithole and Vosselman 2005). In this method point cloud is partitioned into three directions to have their profiles and in order to find segments of line, points are connected in the profiles. Finally, at the last stage, these three partitions are superposed (See figure 2-1) to generate the segments.

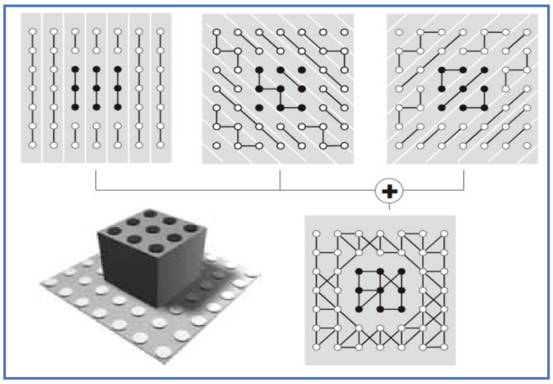


Figure 2-1 Segmenting a point cloud, Two surfaces are segmented using three different profile directions. (Sithole and Vosselman 2005)

Similar segmentation method have been presented by Han et al. (2007), in this method every point is compared with last point based on scan line segmentation. If the height of new point is similar to the last one, then the assigned segment will be same as the last and if not, it is segmented as a new one.

2.1.1.2. Surface growing

In the mostly used algorithms, first a small set of nearby points with a good fit to a plane are found. This set of points then constitutes seed for a surface growing procedure in which adjacent points are added to the segment if they meet some criteria like distance to the plane or locally defined smooth surface. Once there are no points left that meet required criteria, further seed segments are selected and expanded until all points have been assigned to a segment (Vosselman, 2009). In these algorithms planar are generated by detected seed points in the point cloud and then they start to grow. The first step in this method is performed by 3D Hough transform, RANSAC search or robust plane fitting and then in the next step based on the criteria such as difference in curvature, difference in normal vector or deference in slop, point will be connected. Tóvári & Pfeifer (2005) has suggested importance of factors like height, slope or curvature difference in the context of geometrical relation of neighborhood. Vosselman (2004) presented a method in which the base criteria for surface growing are proximity of points, locally planar and smooth normal vector. while Pu & Vosselman (2009) have used surface growing method for segmentation of terrestrial laser point .

2.2. Machine learning techniques

A new approach for classification of ALS data is to employ machine learning techniques. There are several techniques of machine learning approach including Support Vector Machine (SVM), decision tree, Random forests (RF), Maxent and neural networks. These methods are varied in how they present the distribution of the interested classes, select predictor features, analyze feature contributions and capable to interaction. Generally, considering their output, classification approaches can be divided into two main categories: one-class classification and multiple-class classification. One class classifiers can be used to classify given data in order to detect the interested class in the field. The multiple-class classifiers can be used to classify given reference data into multiple predefined classes rather than only two, at one simultaneous running process or by a hierarchical multi-segmentation.

2.2.1. Multi-class classification

Several machine learning techniques have been applied as multiple-class classification technique. Performance of classifier can be evaluated and compared by using statistical methods. The Random forest is recently emerged as a state-of-the-art multi-class machine learning technique which considers ensembles of decision trees, rather than single tree generated by a base classifier. Random Forests (RF) are a variant of bagging created by Breiman (Breiman, 2001). As Breiman indicated, it is a decision-tree-based ensemble classifier that can achieve classification accuracy comparable to other classifier like boosting; Random Forests can be explained as combination of trees which have been resulted of independently sampling. Each tree presents a vector as unit vote for the most popular class. The final label is determined by a majority vote of all trees. The Random Forest classifier has two parameters: the number of trees (T) and the number of variables (M) which are randomly chosen at each split. Error in Random Forest depends on two parameters: first the correlation between any pair of trees and second the strength of each individual tree in the created forest. It is clear that like any classifier, Increasing the correlation between features, increases the forest error rate while increasing the strength of the individual trees decreases this misclassification rate (Breiman, 2001), for application with large datasets such as Lidar data, it is efficient and runs fast, It does not over fit, It does not require assumptions on the distribution of the data, which is interesting when different types or scales of input features are used. These outstanding advantages make it suitable for remote sensing multi-class classification. We should note that, this is crucial to analyse the relevance of each feature for classification. The big advantage of Random Forests is its measure of feature importance. Out of bag (OOB) processed is used to create such output (Breiman, 2001). For calculating feature importance, a high prediction accuracy decrease by omitting a feature shows the importance of that feature.

Generally, in the statistical modelling, different strategies for data partitioning are used, the simplest method is splitting the available reference data in train and test data, for example you can randomly take 70% of your data as training dataset, and use the rest 30% as test data to validate your model's output. In a method called sub-sampling, data are drawn without replacement, but in bootstrapping the data are drawn with replacement. It means that the same segments (points) could be included in the test more than once. The third partitioning method is cross-validation in which at the beginning of the procedure the reference dataset will be divided into N-fold. The model runs N times, each time one fold is used as test data and the other (N-1) folds are used as training data. Therefore, at the end, all N folds are used as test data.

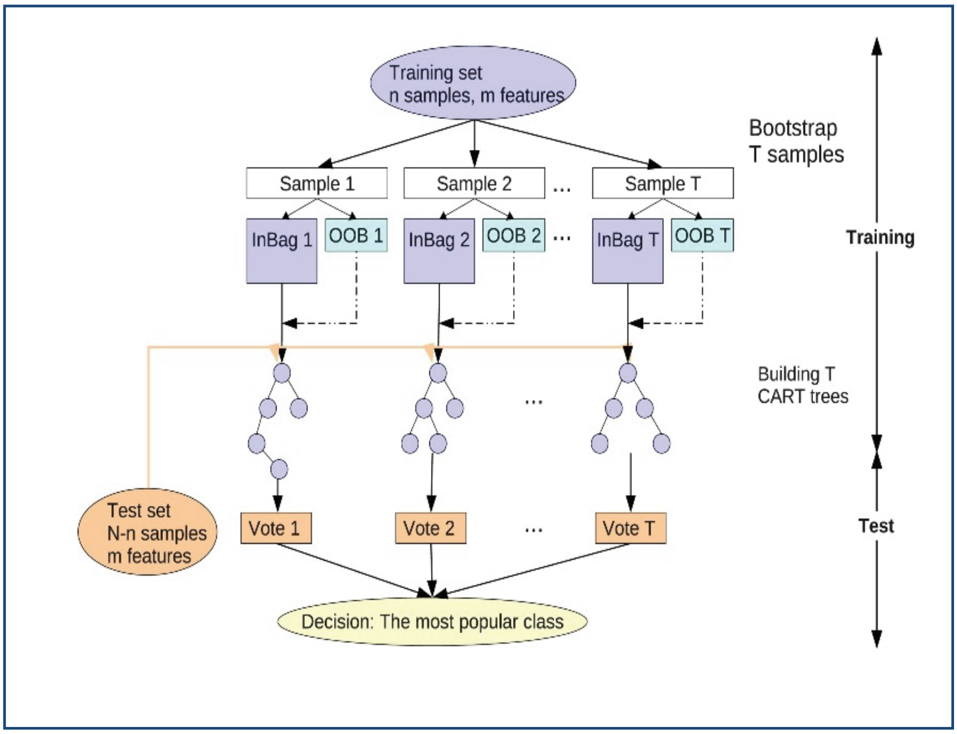


Figure 2-2 Main steps of Random Forests (Chehata et al., 2009)

Recently Kim et al. (2010) presented a 3D classification by employing the Random Forest method to classify power-line scene where a few structures including trees, transmission lines and pylons would be vertically overlapped. The research has extracted 21 features to characterize each class from different segment scale. A sensitivity analysis has been conducted in terms of feature extraction scale, feature importance and class distribution over test datasets. Their Experiments showed that an optimized classification performance of 96% success rate by Random Forest can be achieved with point-based feature extraction. Chehata et al. (2009) presented a classification of Full-wave form Lidar and airborne image data. Provided measures of feature importance for each class, accuracy and also efficient classification were their motivation of employing Random Forest as classifier.

Another machine learning algorithm is AdaBoost; Lodha et al. (2007) have used this method for classification of Lidar data in four classes including road, grass, trees and buildings. Five features have been employed including height, height variation Lidar return intensity, image intensity and normal vector. Also in their experiment in order to create three classes in which the road and grass classes are merged, they used only those features which have been derived from Lidar. They have implemented with several variations within the AdaBoost family of algorithms and suggested that their results are stable over all the various tests and algorithmic variations. The result of the different researches which have been employing multi-class machine learning techniques show promise to good result but, it is considerable that, all these classifiers require the availability of exhaustively labeled training sets of all classes of interest for adequate training of the classification methods.

2.2.2. One-class classification

Second approach for classification is binary classifiers which they seek to classify given data into two features which is often used for solving the fore-to-background problem. One of the popular binary classifier is Support Vector Machine (SVM), Ability and good performance of SVM in variety of research domains make it attractive. SVM gives improved results with respect to traditional classifiers like maximum likelihood. Support vector machine was introduced in 1992 to generate maximal margin for non-separable training data in feature space by hyperplanes (Vapnik & Kotz, 2006). SVM was a binary classifier primitively which labelled classes as + 1 and − 1. The idea for this classification is separating these two classes with maximum margin. SVM constructs an optimal hyperplane for getting to maximized margin (Tso & Mather, 2009). In fact hyperplanes are decision boundaries for separating classes in feature space which use training data that lie on the edge of class distribution. Figure 2-3 shows how optimal hyperplane divide two classes based on maximum margin (Vapnik & Kotz, 2006).

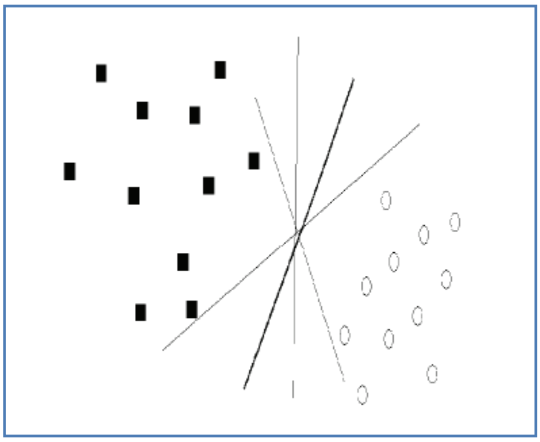


Figure 2-3 Optimal hyperplane of SVM

Although, the SVM classifier described above, distinguishes between two classes only; however, it is possible to be modified to a multi-class classifier. Numerous methods are available for modifying SVMs into a multiclass setting, including:

A. Training a classifier to distinguish each class from all other classes, this method commonly is called "One vs. all",

B. Training a classifier to distinguish between each pair of classes

C. Using some extension of the concept of the margin to include more than two classes, and performing optimization directly on this quantity.

Lodh et al. (2006) have classified 3D aerial Lidar scattered height data into buildings, trees, roads, and grass using the SVM algorithm. To do so they have used five features: height, height variation, and Lidar returns intensity, image intensity and normal variation. In order to have three classes in which the road and grass classes are merged they used only features which have been created from Lidar data. They have implemented with several variations of the SVM algorithm and observed that the results are robust by comparing them against the ground truth.

In the applications of remote sensing, users are sometimes interested in classifying only one specific land cover type, without considering the other land covered classes (Foody et al., 2006). The classifiers seek to extract one specific land type of interest, given only the training sample of the class of interest which is referred to as positive and other land types are referred as negative data. One of the used modified SVM for such condition is One-class support vector machines (OCSVM) which has been proposed by Scholkopf et al. (2001). Recently, this method has been employed to classify remote sensing imagery and has shown good results in some research (Foody et al., 2006). However, its output is sensitive to parameters that are difficult to tune (Manevitz and Yousef 2001). Foody et al. (2006) has indicated that manually collecting training data of classes of interest is time-consuming. Since, traditional supervised classifiers are not efficient in one-class classification, and we need to develop new method to discriminate the single class of interest from the other classes with only positive training data (Li and Guo 2010). There are a few methods that can be used to model the distribution using only true-positive data. Recently, Maximum Entropy approach (Maxent) for modelling distributions of data with a precise and simple mathematical formulation, has been introduced by Phillips et al., (2006). Li and Guo (2010) investigated a Maxent approach to one-class classification of remote sensing imagery, i.e. classifying a single land class. They evaluated the classification accuracy and effectiveness of the maximum-entropy approach and compared it to OCSVM. They have indicated that the differences in the classification accuracies are statistically significant at the 95% level of confidence, which indicates that Maxent produced higher accuracy than OCSVM in their study.

Since in our case of remote sensing classification we are interested in detecting only collapsed building as one specific land class without considering other classes, which is referred to as one-class classification. Given our airborne Lidar data and our limits to extraction of validated information from airborne images and also in the other hand in order to simulate the real time-critical situations after a natural disaster like earthquake Management issues in which more than everything a fast and extensive damage analysis is required, we consider our situation subjects to one-class classification. Recently, the capability of maximum entropy approach was discussed for one-class classification problem of remote sensing data (Li & Guo, 2010), but its application for Lidar data has not been investigated yet. Therefore, in this research, we have investigated the capability of this approaches a one-class classification for Lidar data, and classifying a single class of collapsed building.

2.2.3. Advantage of Maxent

Maxent is the current state-of-the-art method of modeling from only true-positive data (Phillips et al., 2004). In addition to the benefit of needing only true-positive locations without requiring negative data, Maxent is less sensitive than other classifiers to the number of true-positive locations which are required for developing an accurate prediction model (Phillips et al., 2006). The main reason for such characteristic of Maxent is a procedure which is called regularization and by such function it prevents over fitting when there are only a few location for training sample which resulting in relative insensitivity to sample size, to do so, Maxent constrains the estimated distribution of a feature in such a manner that its average value would be close to the empirical average of that feature rather than exactly equal to it. This is called regularization procedure and in the application of Maxent, users can set the optimization parameters for this procedure in a way that potentially compensate the classifier for small sample sizes (Baldwin, 2009) . The relatively small sample size required for accurate model building is a very beneficial characteristic of the Maxent classification due to the lack of reliable data in the spatial distribution modeling. Providing such data are often time and labor consuming, thereby, the effort of manually collecting training data for classification by Maxent can be significantly reduced. Finally the best advantage of Maxent especially for our research is its capability to generate feature contribution measure (feature importance) which as it was mentioned earlier, is important to know how each feature influences the occurrence of collapsed building and, subsequently, which features have the least and the most influence on the model. In line with this output there is response curve which in addition to feature importance, it shows the manner of influence to collapsed occurrences for all included features in model.

2.2.4. Could Maxent be considered as a new approach for Lidar classification?

The recommendations of the different papers are generally suggesting that performance of Maxent for classification of one-class of interest is good. Considering the total advantages of Maxent, especially being binary classifier and needing presence data only which other land classes are not required for training purpose as in all multi-class classifiers and also the small size of training data which, Thereby, the effort of manually collecting training data for classification can be significantly reduced, are really valuable beneficial aspect of the Maxent approach given critical situation of post-disaster that finding reliable data is difficult, so minimizing required input for classifier and gaining acceptable output could be a good indicator to consider Maxent as a new suitable approach for Lidar classification .

3. METHODS AND MATERIALS

3.1. Conceptual Model

This chapter describes materials and methods of this study. The Conceptual modeling of collapsed building is shown in Figure (3-1) that was drafted and used as a guideline during implementation of the processes. The conceptual model shows a flow chart including three main steps of input, modeling and output that will be implemented. As it is shown, the bootstrapping method is used for data collection which has been replicated 30 times (70% training and 30% test) for all modeling.

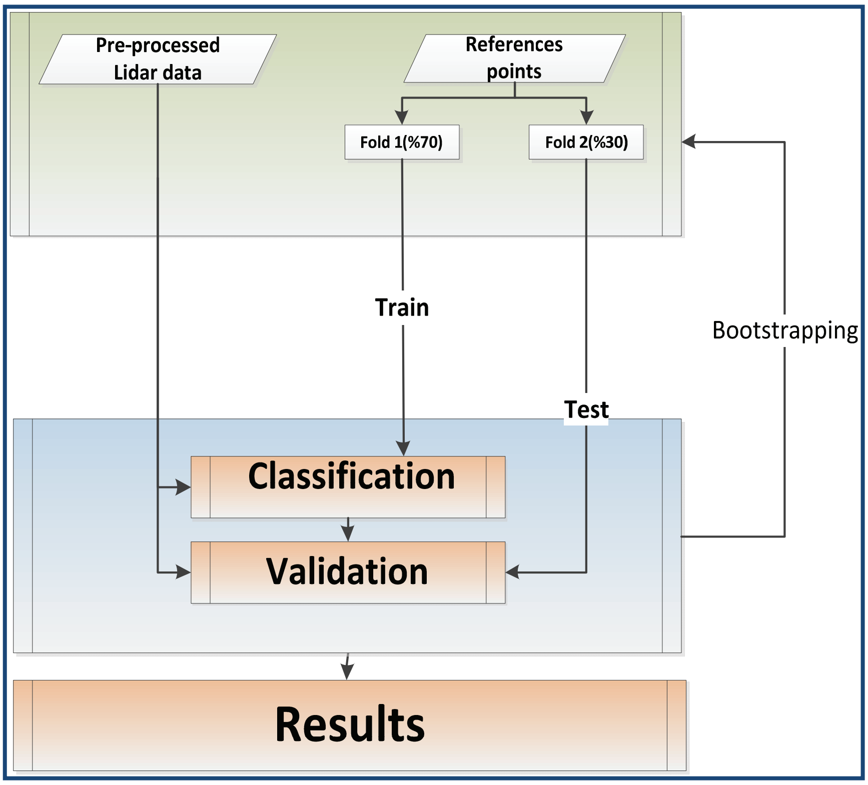


Figure 3-1 Conceptual Model showing input, process and output

3.2. Research workflow

Also the general methodology for spatial modeling of collapsed building is shown in Figure (3-2). These steps will be described in more detail.

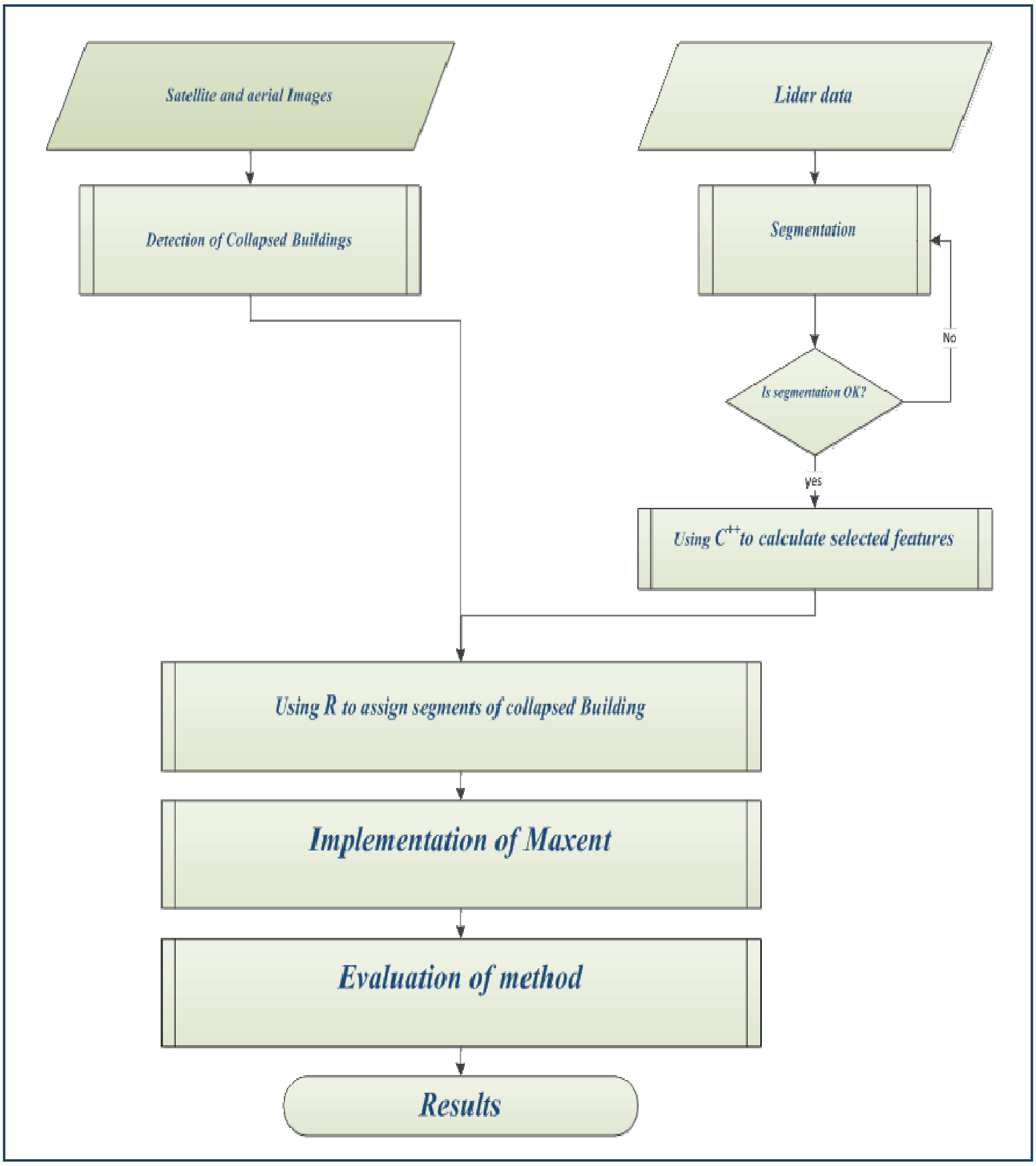


Figure 3-2 General Methodology for spatial Modelling of collapsed building

Understanding the relationship between a signature and the value of features belong to the area in which it occurs is fundamental when developing a Predictive model. In our application, Maxent as Predictive modelling is based on these relationships through calculated features of any single segments by using C++. This study tests the application of Maxent as a method for generating information of collapsed buildings from the airborne laser data. Maxent generate a target probability distribution in a given study area. The principal of method in based on the estimation of probability that is the closest to uniform distribution subject to some defined constraints which are our incomplete information of the target distribution. This information is a set of real-valued features and the constraints are that the expected values of features should match their empirical average. We used corresponding segments of the location of collapsed building created by UN from airborne images and also their features calculated by C++ from Lidar data as set of value for each segments. We followed statistical rule for random selection of sampling data and in order to prevent repeated data we used R software to assign only one which is the closest segment to each collapsed building for training purpose. These segments make up the space on which the probability distribution is defined, and the calculated features of segments are predictor variables. The inputs of Maxent are two sets, one reference segments (calculated features for segments of collapsed building) and second background segments (calculated feature for all segments). 70% (selected randomly) of Reference segments are used for training and 30% are used for test purpose. This process is replicated 30 times to consider their variances and use the averages of outputs. Subsequently, the outputs are evaluated by well recognised standard statistic methods to achieve reliable results. Finally in addition to create probability distribution map, response curve for each feature which shows the manner of their variation (upward or downward) and also feature importance by using Jacknife test are created. This information helps us to discover the influence of features in predicting the likely occurrence of collapsed building.

3.3. Study area

Our study area was Haiti which is located in the west of Hispaniola and Dominica, between approximately 19 N and 73 W in the middle of the Caribbean Sea and the North Atlantic Ocean (Millar, 2010). The population of Haiti is about 7 million and it covers a total area of approximately 27,750 km2 (United Nations, 2010). Figures (3-3) below shows the location of Haiti, an image of Port-au-Prince, the area of interest and also projected map of our Lidar data is available in the next figure (3-4).

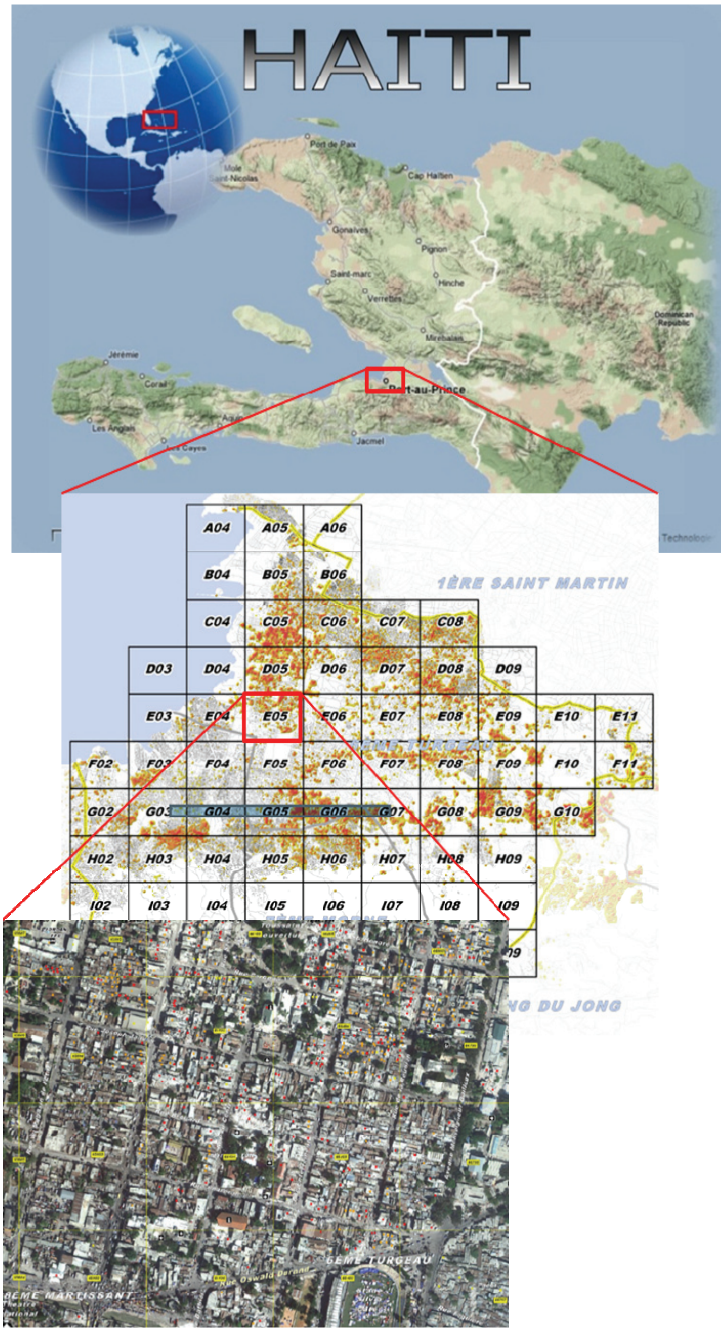


Figure 3-3 Location of Haiti and Port-au-Prince

Also here the centres of created segments are projected in Google earth.

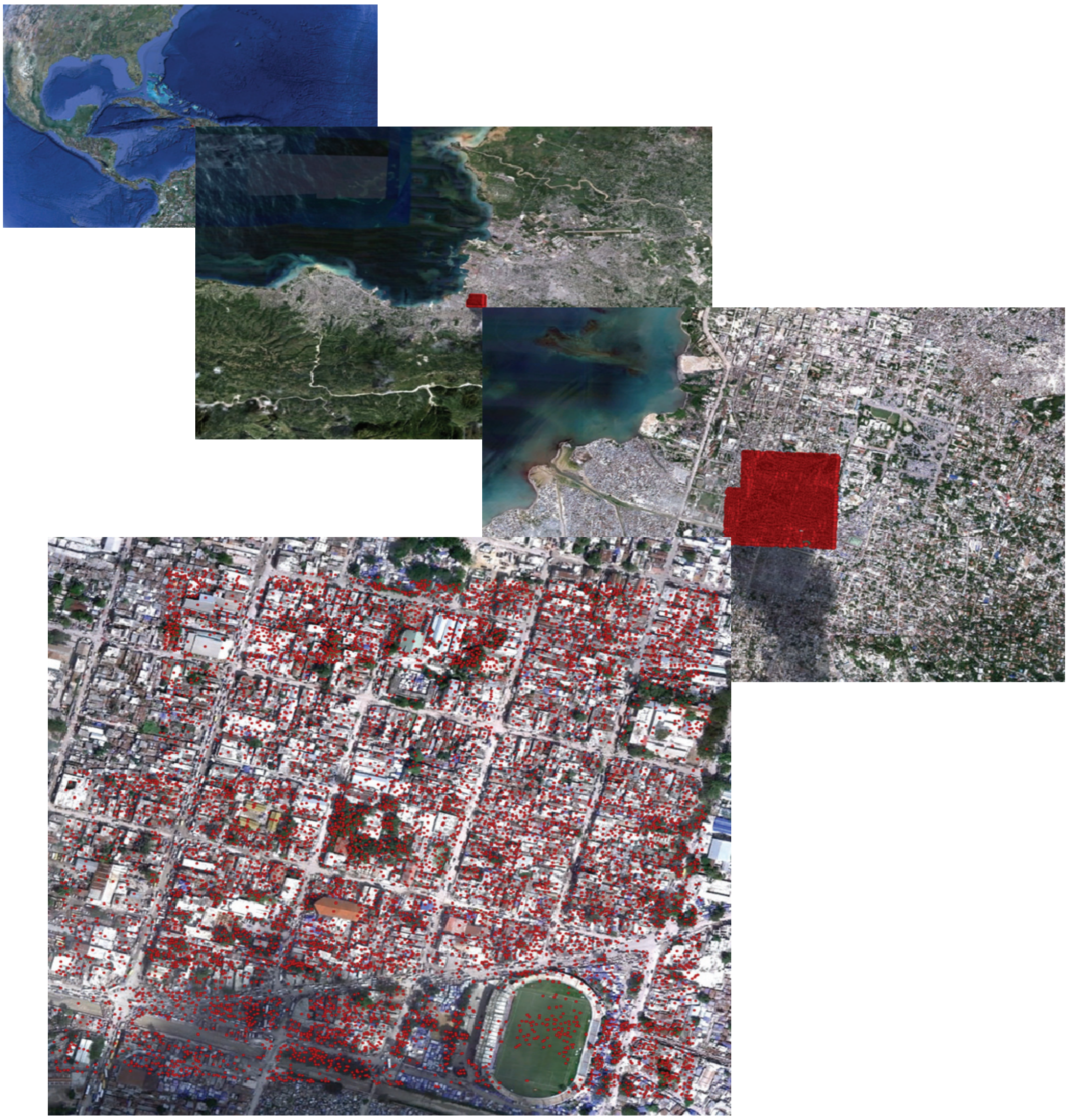


Figure 3-4 The map of Port-au-Prince (centre of segments are projected in Google earth)

3.4. Limits of Damage Detection by airborne Data

Considering existing available source of data, principally there will be some limits to extraction of information and it depends on the specifications of the acquired data as well as on the objects of interest. Damages at buildings are generally described based on damage classes. This may differ from country to country. The way in which a building deforms under earthquake depends on the building type and used material, for instance in our case, damage assessments of individual buildings have been conducted by categorization of damage on the European Macroseismic Scale (EMS-98) Five-level grading system.

Classification of damage to masonry buildings	
	**Grade 1: Negligible to slight damage (no structural damage, slight non-structural damage)** Hair-line cracks in very few walls. Fall of small pieces of plaster only. Fall of loose stones from upper parts of buildings in very few cases.
	**Grade 2: Moderate damage (slight structural damage, moderate non-structural damage)** Cracks in many walls. Fall of fairly large pieces of plaster. Partial collapse of chimneys.
	**Grade 3: Substantial to heavy damage (moderate structural damage, heavy non-structural damage)** Large and extensive cracks in most walls. Roof tiles detach. Chimneys fracture at the roof line; failure of individual non-structural elements (partitions, gable walls).
	**Grade 4: Very heavy damage (heavy structural damage, very heavy non-structural damage)** Serious failure of walls; partial structural failure of roofs and floors.
	**Grade 5: Destruction (very heavy structural damage)** Total or near total collapse.

Figure 3-5 European Macroseismic Scale

Grades 1 to 5 are supposed to represent a linear increase in the strength of shaking for types of masonry buildings.

In a research conducted by Bitelli et al. (2004), has been indicated that by using optical remote sensing imagery, just some levels of damage like collapsed and severely damaged buildings can be detected with a good faithfulness. Also Ogawa & Yamazaki (2000) evaluated the Building damage due to the 1995 Kobe Earthquake by using aerial photographs, they have compared the results with that of ground truth to examine the applicability of aerial photo-interpretation. According to this research, it has been suggested that damage interpretation using aerial photographs of an urban area is only effective to identify collapsed buildings, furthermore, as Weston et al. (2003) indicated, our main interest is to find the relation of fatalities and building vulnerability, about 75% of fatalities attributed to earthquakes are due to the collapse of buildings.

After a disaster not necessarily all buildings in affected areas are damaged, certainly some remain intact. Based on the above damage classification scheme, their description and some geometrical features of the buildings structure after collapse such as reduction in volume, size, shape and structure of debris, it is obvious that there are some differences in the appearance of totally collapsed buildings and non-collapsed one . Collapsed building are fine debris, a coarse heap or rubble, while some of the non-collapsed buildings are small structures that appear rectangular others are bigger, possibly commercial buildings with multi panelled roofing.

In the other hand, for analysing airborne Laser data, roofing is of our interest as often laser data captures roof structures more sharply than the sides of the buildings. Given such result of airborne Lidar data and our described limits to extraction of information needed for our training purpose, we have considered grade 5 of damage classification as total or near to total collapsed building for our class of interest for training purpose.

3.5. Data

Post disaster Lidar data of the study area covering Central Port-au-Prince taken on the 22 January 2010 with average resolution of approximately 2 points per square meters. The data was in WGS84 under a Universal Transverse Mercator Projection System as shown in Table 1 below.

We started to classify segments of area 1 and then the region of study was extended to area 2 in order to have better evaluation of model performance.

UTM 18N - WGS 84		
Area 1	Top: 2052115	Left: 780140
	Bottom: 2051505	Right: 780699
Area 2	Top: 2053652	Left: 780083
	Bottom: 2052403	Right: 781075

Table 1 The area of study

To obtain the statistics data and investigate the feature characteristic, an executable C++ based program was used. It takes a laser file in its .laser format as input and gives output attributes of interest including, X, Y, Z of segment's centre, the number of points per segment, density of points in 2D, density of points in 3D, ratio of segmented points to non-segmented points, standard deviation of intensity, standard deviation of z, distance to DTM and residual to planarity. The neighbourhood relative features like density in 2D and 3D and also ratio of unsegmented points per segmented points are calculated within a cylinder of 5m radius centred at the centre of under investigation segment. The output is primarily in ASCII format where a "csv" format is chosen as this can be uploaded easily into ArcGIS. And also through the use of aerial photos provided by the World Bank, detailed damage assessments of individual buildings have been conducted by comparing pre-earthquake satellite imagery to post Earthquake satellite imagery and aerial photos. The spatial resolution (level of detail) for the satellite imagery used is approximately 50 cm while for the aerial photos it is approximately 15 cm. It is important to say that since damage assessments of individual buildings was recorded based on the centre point coordinate of building thus we assumed a buffer of 4m around the centre point as collapsed area.

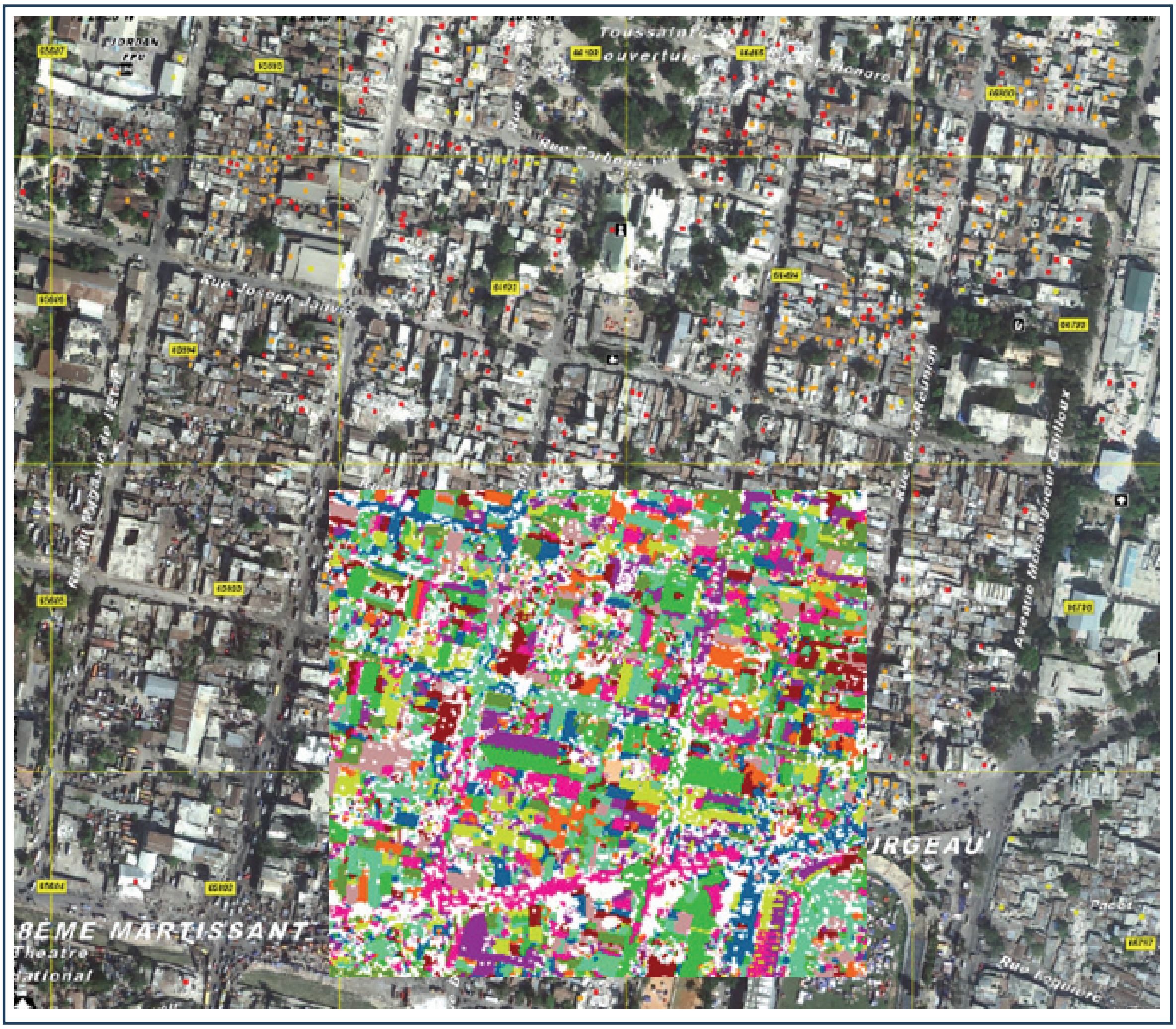


Figure 3-6 Segmented Lidar data Superposed on aerial Image

3.6. Feature Selection

Knowledge about the structure of collapsed buildings is necessary for the interpretation of the data collected by airborne laser scanning. The overall goal is to classify damaged buildings from airborne laser scanning data in the context of disasters, like earthquakes. Classification means to assign unknown patterns to a priori given classes. The patterns to be classified are the result of a segmentation process of the Lidar point clouds. Since our approach for modelling collapsed buildings is based on the assumption that undamaged buildings may be represented by large planar surfaces and in contrast the strong damages will result many small planer surface elements and many non-segmented points. However damaged buildings may show very different pattern types, in order to assign the location to a priori determined classes, features have to be defined and extracted. These features should be chosen in such a way that they cause a high discrimination between the different classes. As geometry is concerned, features could be e.g. planarity, differences in height or distance to DTM as well as changes of the inclinations of the building's surfaces and so on. Therefore, segmentation is a necessary step with respect to the following classification and means division of the point cloud into homogenous regions. Feature selection is important as meaningful features facilitate accurate classification of the data. Therefore, modelling damaged buildings has to take into account such geometrical features which characterize the respective building type very well. In order to detect the collapsed building, following features including some general attributes which could be used for detection of different classes of segmented Lidar data were used for every segment:

- Residual to planarity
- Distance to DTM
- Standard deviation of height
- Standard deviation of intensity
- Number of point per segment
- Density of point in 2D
- Density of point in 3D
- Ratio of unsegmented points per segmented points

3.6.1. Residual to planarity

For segmentation of planar surface in a region growing algorithm, the n assigned points are accepted if they approximately located in a plane, this approximation is a threshold for residual (distance) from points to created plane. Planarity is evaluated by assessing these residuals to form a plane during the segmentation process. Irregular features like vegetation and collapsed area are mathematically likely to have higher residuals. In segmentation process by increasing the maximum distance to planarity the size of segments and their number of points will be increased. Although this feature seems to be meaningful but in parallel it is effecting on two other feature and is correlated with them which needs to be considered carefully.

,

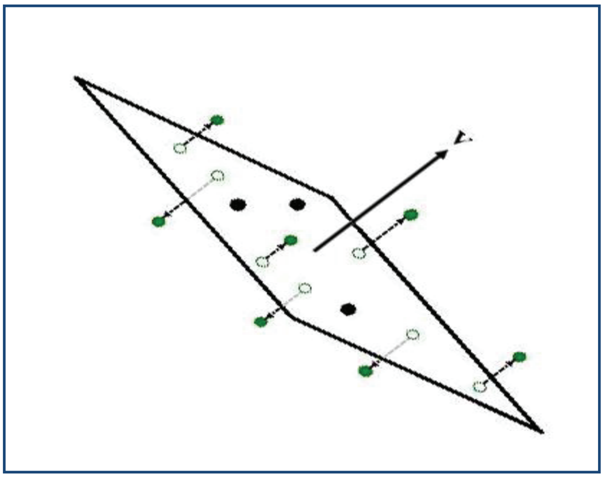


Figure 3-7 Residual to planarity of segments

3.6.2. Distance to DTM

Elevation differences measure local variation and are expected to be a reliable feature to capture the existence of geometry changes. Consequently, height differences enhance the separation of classes. Due to change in geometry after collapse, particularly in the height of building, it is obvious to see some difference in distance to DTM between collapsed building and non-collapsed one, analysing the average distance to DTM of a segment of collapsed buildings compared to that of the non-collapsed one within the Digital Surface Models would be an indicator to determine the characteristic of our class of interest for this feature. Distance to DTM was investigated as it could give us significant indicators useful in classification.

3.6.3. Standard deviation of height

In addition to distance to DTM, due to appearance of collapsed building, as mostly they are heap of debris or partially damaged wall, it won't be surprising to find lots of changes in their heights which will motivate investigation on height changes between the neighbour segments. For such purpose standard deviation of mean height to DTM is the best feature which could show where is likely to have been collapse within the built up area by checking the drop and rise in the heights of segments. Standard deviation of height is normally a very key attribute in building damage assessment as it is a sharp indicator of the nature (e.g. regularity or irregularity) shape of structure.

3.6.4. Standard deviation of intensity

The standard deviation of intensity is a clear indicator on homogeneity of the features reflectance. In the norm one would expect a homogenous area to have a homogenous intensity value. By statistical analysis of the point cloud intensities, the areas that are suspected to be collapsed buildings show changes which seem to be following some rule in their changes, investigating this attribute would shed light on the variation of intensity in the collapsed building classes. The standard deviation of intensity gives us a measure of dispersion from the mean**,** this is a meaningful feature as it is common for the laser and we expect to be of good help to detect our interested area.

3.6.5. Number of point per segment

The number of points per segment for a given set of input parameters and for a given object size, in a segmentation algorithm, can be a useful indicator of geometric characteristic of objects. One expects the planar features like roads and non-collapsed building to have larger segments as a large number of points geometrically grow and fulfil planarity conditions.

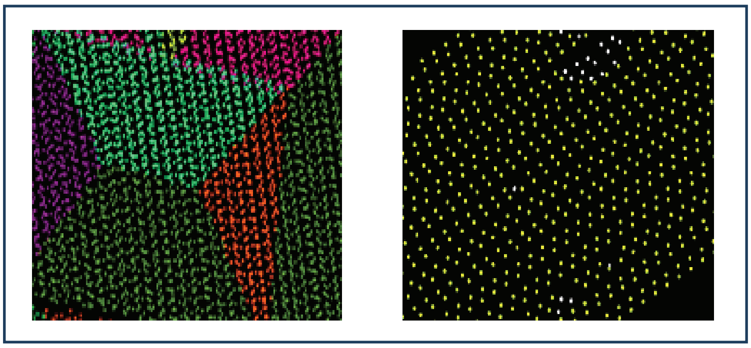


Figure 3-8 Sample of large segments

And in contrast one expects small size of segment as a few numbers of points geometrically fulfil planarity condition for area covered by rough pattern and irregular features, sometimes after fair segmentation (not over or under segmentation) even the number of unsegmented points could indicate of geometric characteristic.

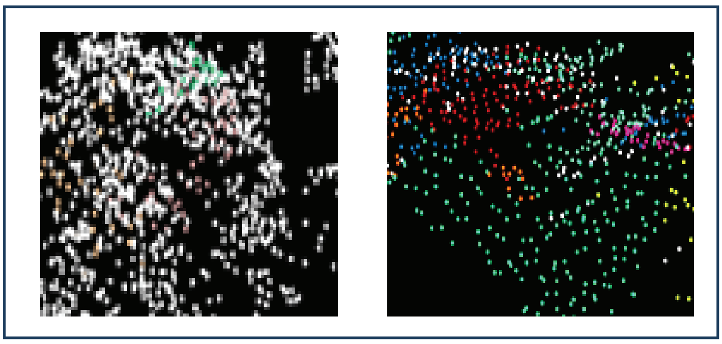


Figure 3-9 Sample of small segments

3.6.6. Density of points in 2D

For investigation of some features we need to consider them in a neighbourhood area to find their behaviour in relation to those neighbour points. For each segment, the Lidar features are computed using the density of points in 2D included in a given cylindrical of 5m radius neighbourhood, centred at the current segment.

3.6.7. Density of points in 3D

As for density of points in 2D, we investigated about Density of points in 3D. Density is considered in a neighbourhood area to find their behaviour in relation to those neighbour points. Difference in geometry and also penetration of pulse inside some area could be indicated by this feature. This feature is computed using the 3D Points included in a given cylindrical of 5m radius neighbourhood, centred at the current segment (Fig 3-10).

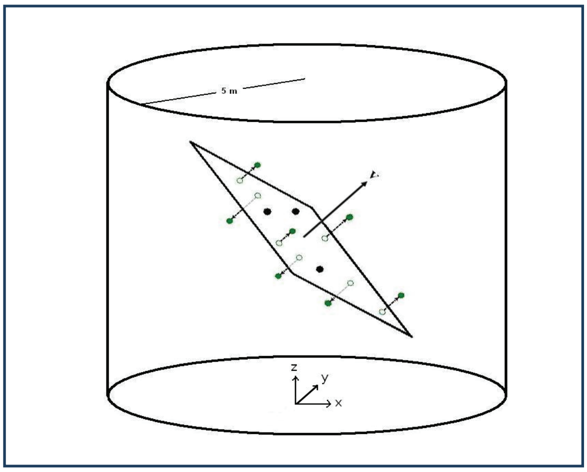


Figure 3-10 Cylindrical of 5m radius neighbourhood

Since our point cloud data are single pulse, so we don't expect significant changes in 2D density but some changes could be seen within trees for density of points in 3D due to penetration of some pulse inside the vegetation area.

3.6.8. Ratio of unsegmented points per segmented points

Depending on the algorithm used for extracting planes from the laser data, some points simply will not fit any of the generated planes and thus remain unsegmented. Due to the fact that damage types like heaps of debris, outspread multi-layer collapses or overturn collapses have a very irregular structure of surface, the assumption can be made that many unsegmented pixels occur in areas affected by these damage types. Therefore the ratio of unsegmented point could be a good indicator of the nature of a collapsed building. In our case of classification algorithm the unsegmented points per segmented points ratio is calculated based on a search within previously defined cylindrical of 5m radius neighbourhood.

3.7. Maxent

The maximum entropy (Maxent) technique is selected for spatial modeling in this study. As far as the author`s knowledge can reach, despite the previously mentioned advantages of Maxent, its applications in Lidar data classifications have been rarely studied so far. Therefore, it was proposed for this study to implement the same method to one-class classification of Lidar data.

3.7.1. Implementation for modelling collapsed building

As illustrated in Fig 3-11, Maxent implementation has a quite easy and user-friendly graphical user interface.

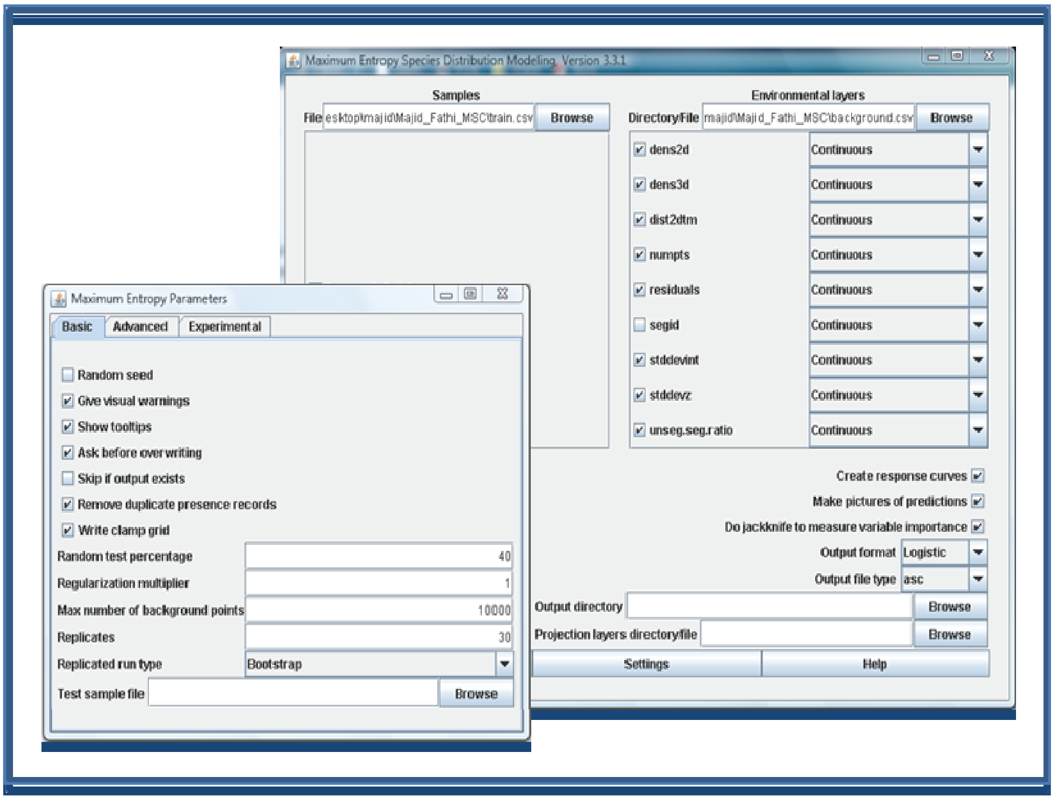


Figure 3-11 User interface of Maxent for modeling collapsed building distributions

We used the Maxent software that is freely available online at:

www.cs.princeton.edu/~schapire/maxent

The inputs are the locations of the positive training and testing reference segments and their corresponding feature values.

3.7.2. The principle

Maxent is a general-purpose method to inference from incomplete information (Philips et al, 2006). As Jaynes mentions in his work, the origins of Maxent method lie in statistical mechanics (Jaynes, 1957). According to the principle of maximum entropy, the distribution of the target class, first should satisfy any given constraints and then should be as uniform as possible (Phillips et al., 2004). Given the constraint values from the train data, this approach can estimate the most uniform distribution that has maximum entropy for the target class for unknown locations. This agrees, to a very good extent, with everything that is known, however carefully avoids assuming anything that is not known (Jaynes 1990). In Maxent, π which is the unknown distribution of probability is calculated over a set of segment's location and assigns a non-negative probability $\pi(x)$ to each segment (x). They are probability, therefore, should sum to1. The constraints on the unknown probability distribution π are represented by a set of features (residual to planarity, distance to DTM, Standard deviation of height, Standard deviation of intensity, number of point per segment, density of point in 2D, density of point in 3D and ratio of unsegmented points per segmented points) denoted as $\mathbf{f_1}, \dots, \mathbf{f_n}$ on $\mathbf{X}$, and the information as measured by averaging of each feature is the expectation of features under π , which is defined as:

$$\pi[\mathbf{f_j}] = \sum_{\mathbf{x} \in \mathbf{X}} \pi(\mathbf{x}) \mathbf{f_j}(\mathbf{x})$$

A set of sample segments $x_1, \dots, x_m$ is drawn independently from X . The corresponding empirical distribution is then:

$$\tilde{\pi}(x) = \frac{|\{1 \leq i \leq m : x_i = x\}|}{m} \tag{1}$$

We define the empirical average of f_j under $\tilde{\pi}$ as:

$$\tilde{\pi}[f_j] = \frac{1}{m} \sum_{i=1}^{m} f_j(x_i) \tag{2}$$

We use $\tilde{\pi}[f_j]$ as an estimate of $\pi[f_j]$. This approach aims to estimate the probability distribution $\tilde{\pi}$ which is an approximation of π , subject to the constraint that the expectation of each feature f_j under $\hat{\pi}$ is the same as its empirical average, stated formally as:

$$\tilde{\pi}[f_j] = \hat{\pi}[f_j] \tag{3}$$

The entropy of $\hat{\pi}$ is defined as:

$$H(\hat{\pi}) = -\sum_{x \in X} \hat{\pi}(x) \ln \hat{\pi}(x)$$

It is clear that this equation includes a natural logarithm (**ln**). The entropy is non-negative and in the specific case of equal probability for all segments will be the most which is the natural log of the number of segments in X. As Anderson et al. (2007) already claimed, Entropy is a fundamental concept in information theory. It was mentioned earlier in this chapter that the distribution model which satisfies any given constraints should be as uniform as possible. In fact, there are many distributions satisfying these constraints but the one maximum-entropy principle suggests is the one with maximum entropy. A software was developed by Philips et al. (2006). In this software a uniform distribution is used as start point. After several iteration, the probability of occurrence of collapsed building for location of samples increases to reach maximum which is called training gain Dudik et al., (2004). Also to explain same informative measure known as test gain in other words, we could say that, this is the average log probability of the sample data used to test the performance of model. For example, if the achieved test gain is 1, it means the likelihood of a test to be true is exp(1) or $e^{(1)}$ (about 2.7) times greater than that of a random classification of segments (is denoted as background data). This value is a measure of goodness of fit and as Philips et al. (2006) suggest, can be used to assess the overall model performance.

3.7.3. Model building

In order to building our model we employed bootstrapping as partitioning method to select 70% of our reference data as occurrence localities for training part and remaining 30% was reserved for testing the resulting models. Models were replicated 30 times and averaged results of these replications were used to analyze and extract the needed information. First step was to find a subset of features which have adequate performance. In order to satisfy this objective of the research, initially an effort was made to identify which features were the most important ones in predicting the occurrence of collapsed building. The full model may be oversized (i.e. some features may have little predictive power) or redundant (some features may be correlated which will be resulting in multi-collinearity). The best approach for feature selection is stepwise method, thereby, after any steps we can analyze the influence of the previous step in the new generated model. Following stepwise approach, the least important feature from the full model was determined by jackknife test (see section 4.2.1) and omitted from features. Then new model including remaining features was built. As in previous step, the weakest feature among the remaining features was determined (again by jackknife test). This stepwise selection of features was continued until the best suitable set of features was achieved. This is particularly useful to make an optimized selection of feature for predictive models. As it was said, we used Jacknife test for identifying weakest feature. However, to achieve such goal, for each individual feature, regularized training gain was calculated by using only under investigation feature isolated while the other features were assumed to be in their empirical average which will turn out by feature's contribution. The Jacknife test was also used for identifying which features are more informative, to reach this goal also, the decrease in the gain when this feature is omitted from the model is calculated. Therefore, by applying these two processes as it is shown in figure 4-2, we could have some information of the most and least predictive features.

In order to achieve first goal for modeling we started with full model including all features. Then by excluding the feature with lowest contribution and re-running the model with remaining features, we continued until we rich the best subset of model. A training data with 70% of the total 281 segments, including , the corresponding segments to the points which are recorded as collapsed, was randomly selected as training or calibration data, and the remaining 30% was used as test or validation data. It should be noted that the performance of models could be varied due to the training data bias. As it was mentioned, to deal with this difficulty, and also, in order to assess the average behavior of the algorithms, instead of one single partition, thirty (30) random partitions by bootstrapping method were made and the averaged output were considered for our research. However, some further settings for other parameters of Maxent were required that we just left them as default for this study. This setting of course, needs more investigation to explore the influence of parameters on performance of the model for classifying Lidar data.

According to parsimonies rule, the logic behind the best choice of model for this study was to find the model, with an average power of prediction not significantly less than other choice by having the least number of features. The following diagram (Figure 3-13) shows the procedure to build the final selected model with best suitable features.

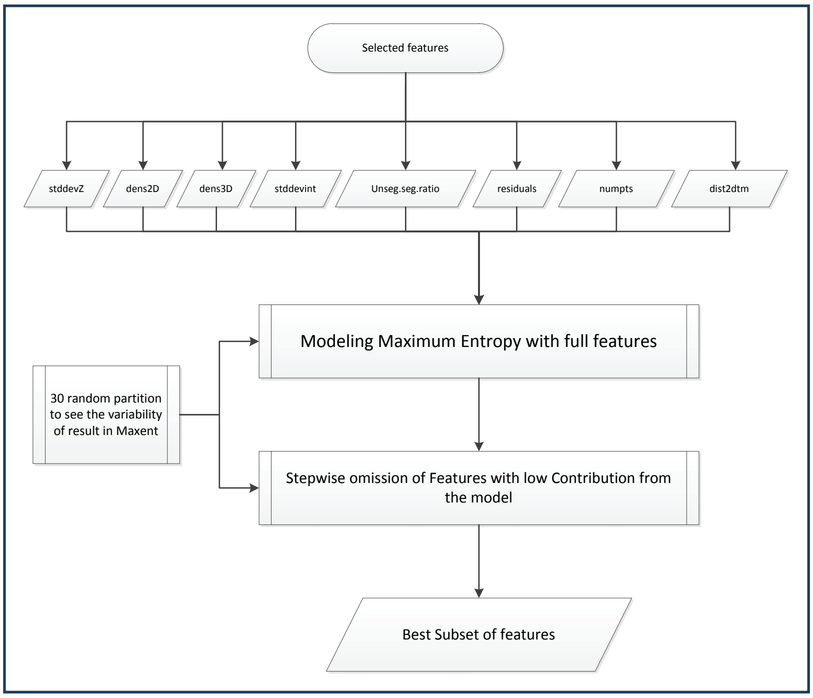


Figure 3-12 Stepwise selection of final features

3.7.4. Model output

Maxent provides output data in three formats of raw, cumulative, and logistic (Phillips et al., 2008). However, since the raw values as primary output for Maxent must sum to 1, they are often very small for each segment and difficult for interpretation. In contrast the second format which is cumulative provides scores for locations of segments. The value of this score is the predicted probability for occurring collapsed at that segment plus all other segments with same or lower probabilities. The provided Scores range at 0–1, therefore, in comparison to the raw format, this output is more easily interpreted.

Third format of output is logistic, it provides estimates of the probability of being a collapsed building for any segment which is predicted by all included features in the model. This means that by having large differences in logistic output, large differences in how likely a building is collapsed are expected. As the cumulative format, logistic format of output ranges from 0–1, which resulting easier and potentially more accurate interpretation. All three explained format of outputs can be imported and mapped into a geographic information system (GIS). All three format of output are monotonically related. This means if we use ranked based statistics, for example ROC curve, all segments will be ranked in same order for all formats, but the result of performance will be different when using statistics that involve the actual values (Phillips et al., 2008).

3.7.4.1. Feature importance

Evaluation of the importance of variables that are used in the model building, gives us the contribution of each feature (variable) in the estimation of the probability of the class (collapsed building). It also gives insights about the features with the lowest and the highest importance for the model. In order to interpret the model, it is crucial to know in what manner these features can influence the class probability. This is specifically important in our case as one of its objectives is to find the most important and suitable features for the distinction of damaged buildings.

In this study, feature importance was determined using jackknife test (leave-one-out). In this approach, the procedure of model building is repeated when each variable is excluded from the variable set and when the variable is the only feature in the model. By comparing the measures (gain and model performance) the importance of variable can be calculated which indicates how important the role of excluded feature in model building is. The big advantage of this procedure is that correlated features can be found, because, in the case of correlation between two or several features, by eliminating any of them, there won't be significant changes in model's gain due to the information provided by the other correlated features which have not been eliminated yet. This procedure was also used as a step-wise procedure to select the appropriate variable set for the model building.

3.7.4.2. Feature profile

Feature profiles (response curves) were produced to illustrate in which range of features the probability of collapsed for a building increases. As an example, a response curve is illustrated in Figure (3-13) in which X-axis is the range of feature values and Y-axis is the predicted probability generated from logistic output. In the case of upward movement, response curve indicates positive effect of feature and for downward movement shows negative effect of feature to predicted probability (Yost et al., 2008).

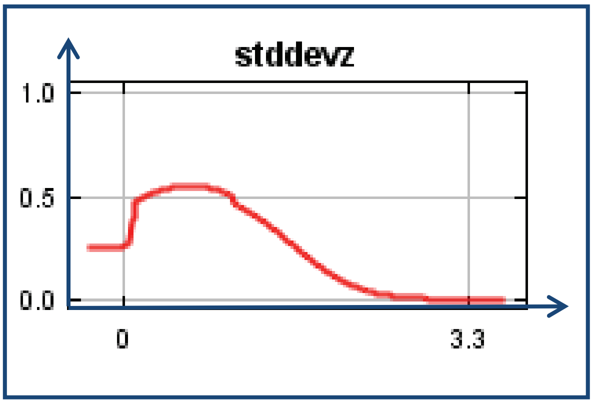


Figure 3-13 Response curve shows the influence of feature and its direction to probability

3.7.4.3. Distribution Maps

Two types of distribution maps were produced as the output. First output gives the probability distribution of being collapsed building for a segment point in a given area. It is, indeed, the logistic output that generally can be used to represent this probability.

The second output is a binary map. By selecting a probability threshold, the segment points were partitioned into collapsed and non-collapsed area. However, selecting that which threshold is the appropriate one is difficult, and some strategies have been introduced by studies to select the threshold. Both output maps were incorporated into a GIS, thereby making it easy to indicate important areas.

3.7.5. Model evaluation

In order to evaluate the performance of the model, two approaches were used: first, area under curve (AUC) of receiver operating characteristic (ROC) plots, which is a threshold- independent method. Second, threshold dependent measures (Kappa & TSS) uses binary distribution considered collapsed and non-collapsed areas.

3.7.5.1. Data partitioning

In order to evaluate the model, an independent test data (validation dataset) is needed. However, in many cases, this dataset is not available. Therefore, existing dataset is partitioned into training dataset and test dataset. Generally, in the statistical modelling, different strategies for data partitioning are used, the simplest method is splitting the available reference data in train and test data, for example we can randomly take 70% of our data as training dataset, and use the rest 30% as test data to validate our model's output. There are some other strategies (i.e. sub-sampling and bootstrapping). In these strategies, the procedure of partitioning repeats several time to avoid the bias that may happen when one-time partitioning is used. In the method of sub-sampling, data are drawn without replacement, but in the bootstrapping the data are drawn with replacement. It means that the same segments (points) could be included in the test more than once. The third partitioning method is cross-validation in which at the beginning of the procedure the reference dataset will be divided into N-fold. The model runs N times, each time one fold is used as test data and the other (N-1) folds are used as training data. Therefore, at the end, all N folds are used as test data.

3.7.5.2. Threshold-independent evaluation

The first approach uses ROC plots in a way that the area under the curve (AUC) is calculated. An ROC is a plot of sensitivity versus 1–specificity and provides an easy way to assess difference. Fawcett (2004) has suggested that ROC graphs are very useful tools for evaluating classifiers. They are able to provide a richer measure of classification performance than accuracy or error rate can, and they have advantages over other evaluation measures such as precision-recall graphs. It is tried to describe ROC curves in a very general and brief terms as following:

Considering a classification result, where each instance is either positive or negative, a classifier assigns a real value to each instance, to which a threshold may be applied to predict class membership in a positive and negative format.

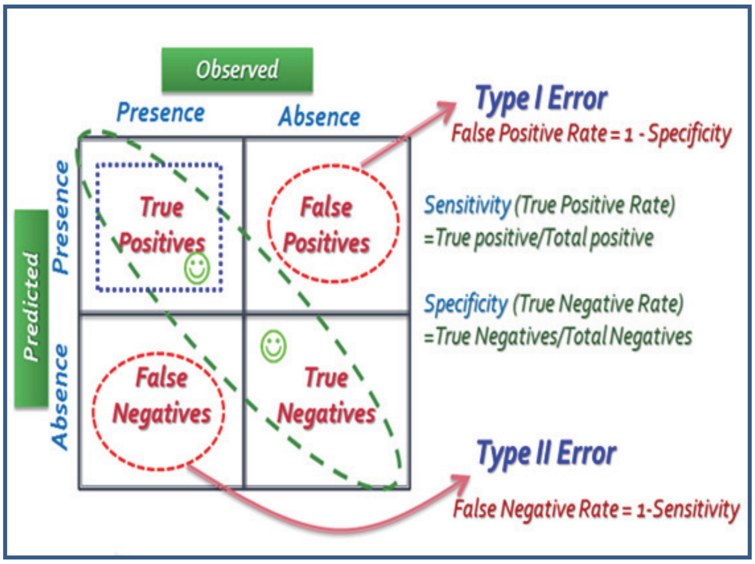


Figure 3-14 Confusion Matrix (Fielding & Bell ,1997)

Sensitivity is known as the true positive rate, and represents omission error. The quantity 1–specificity is also known as the false negative rate, and represents commission error. The ROC curve is obtained by plotting sensitivity on the y axis and 1–specificity on the x axis for all possible thresholds.

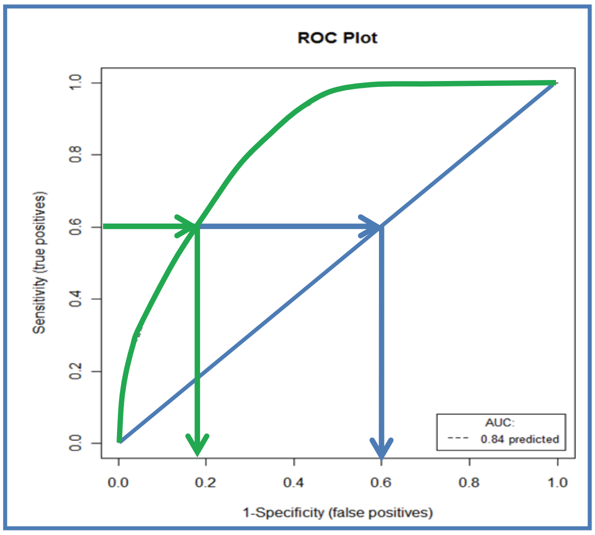


Figure 3-15 ROC curve, comparing model prediction in green and random prediction in blue

Since the **AUC** is a portion of the area of the unit square, its value will always be between 0 and 1. Also the diagonal line between (0; 0) and (1; 1) is the random guessing, which has an area under the curve of 0.5, therefore, no realistic classifier should have an AUC less than 0.5(indeed they are between 0.5 -1). In order to develop a ROC plot, as it was described earlier, a certain percentage of reference data is selected as the training data and the remaining is used for test data. Figure 3-16, below gives a rough interpretation of different result of ROC curves.

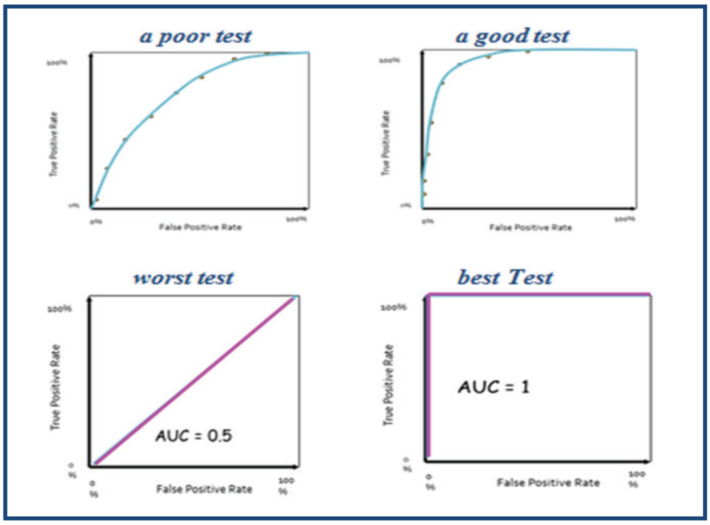


Figure 3-16 Interpretation of different ROC curves Fawcett (2004)

Using AUC as an evaluation method is common in statistical modelling especially for spatial modelling but there are some researches in Lidar which have been using ROC for their evaluation like Blagojevic (2010) for Modeling of Fluorescence Lidar ROC Curves, Fernandes (2004) for Development of neural network committee machines for automatic forest fire detection using Lidar and also Secord (2007) for Tree detection in urban regions using aerial Lidar and image data.

Swets (1988) believes that AUC more than 0.9 is very good; AUC between 0.7–0.9 is good and AUC less than 0.7 are uninformative. The AUC is also closely related to the Gini coefficient (Breiman et al., 1984), which is twice the area between the diagonal and the ROC curve. Care should be taken when using ROC curves. The classifiers cannot be evaluated without a measure of variance. By 30 times replication of models employing bootstrapping, averaging ROC curves can be generated (see section 4.4.1).

3.7.5.3. Threshold-dependent evaluation

Second approach involves selecting thresholds to have positive or negative format of data, it means by selecting a threshold of probability we find the suitable or unsuitable location for the class of interest. The main difficulty with this approach is appropriate threshold selection. Phillips et al. (2006) recommended a method to establish a threshold. They suggested those thresholds which maximize sensitivity while minimizing 1- specificity. However there are some other recommended thresholds like maximum kappa.

In order to determine the accuracy of model by this method, the proportions of correctly classified as collapse segments are compared to the proportion of correctly classified as non-collapsed segments and are presented in a format of matrix called confusion matrix.

We used two indexes (i.e. Kappa & TSS) as threshold-dependent accuracy measures.

Kappa coefficient

In order to evaluate our classifier one of the standard method to be used is Kappa coefficient, by definition of a, b, c, and d in the following table:

		Validation data set	
		Presence	Absence
Model	Presence	a	b
	Absence	c	d

The Kappa coefficient is defined and calculated by:

$$\frac{\left(\frac{a+d}{n}\right) - \frac{(a+b)(a+c)+(c+d)(d+b)}{n^2}}{1 - \frac{(a+b)(a+c)+(c+d)(d+b)}{n^2}}$$

(Allouche et al., 2006). However, we calculated kappa coefficient, but since there are not a balance between collapsed and non- collapsed segments, especially we assumed all roads and trees as non-collapsed, and since Kappa coefficient is sensitive to prevalence (Allouche et al., 2006),therefore, we have used another method to evaluate our classifier is called TSS.

True scale statistics (TSS)

As the Kappa is sensitive to prevalence (proportion of collapsed area to non-collapsed), a measure called true scale statistics (TSS) has been proposed by Allouche et al. (2006), that is not sensitive to prevalence. Based on the earlier definition of a, b, c and d, the following equation can be used to calculate this measure (Allouche et al., 2006):

Sensitivity	$\frac{a}{a+c}$
Specificity	$\frac{d}{b+d}$
TSS	sensitivity + specificity – 1

3.7.6. Effects of sample size on model performance

In order to explore what sample size is appropriate to keep the Maxent as an appropriate classifier, it was tested that how well the model perform when the sample size decreases. Therefore, we used to minimize the size of training data by starting from 70% of data as train and continue to use 60%, 50% and then only 40%of randomly selected data to train the model. The performance of the models were evaluated and compared.

4. RESULTS

This chapter describes the main findings of the research and discusses them briefly.

4.1. Model building

In order to build model first we need to select and finalize our features to have the best suitable combination of them. To achieve this goal, the average behaviour of Maxent in 30 bootstrap simulations using all features was employed. The final result revealed that some features (e.g. density in 2D, density in 3D and residuals to planarity) had the least predictive power. For data partitioning among totally 8797 segments and 281 collapsed records, we employed bootstrapping method by using randomly 70% of data (197 segments) for training and 30% (84 segments) for test which were replicated 30 times and obviously in every replication new training and test data were selected and finally the average of outputs was analyzed. In this test according to the result of Jacknife test (Figure 4-5), excluding these features (i.e. density in 2d, density in 3d and residual to planarity) do not have significant effects on the measures of training gain, test gain and AUC for training and test data.
Regarding this result, the most important feature is the number of points per segment. The training gain, test gain and also AUC of test data will be decreased if this feature omitted from the model, which suggests the size of segments contains useful information that are not existed in other features. After number of points, the unsegmented per segmented points ratio and distance to DTM are the second and third important features, respectively.

4.2. Stepwise feature selection

In order to have the best suitable combination of features, stepwise selection method was employed to find final model. Stepwise procedure was started by analyzing the results of 30 bootstrapping of full model including all features, to identify which features were the most important ones in predicting the occurrence of collapsed building. Therefore, after any steps we can analyze the influence of the previous step in the new generated model. Following stepwise approach, the least important feature from the full model was determined by jackknife and some other extracted information including feature profile and feature contribution.

4.2.1. Jackknife test

One of the well-known standard methods to measure feature importance in order to select the best suitable combination of feature is jackknife test. The following figure shows the results of the jackknife test of feature importance. The feature with highest gain (0.355) when used only is number of points per segment, which therefore appears to have the most useful information by itself. The feature that decreases the gain the most when it is omitted is also number of points per segment, which therefore appears to have the most information that isn't present in the other features. Also for a feature like density in 3D which looks to have minimum gain when is used isolated and also minimum decrease in gain when is omitted, indicates that density in 3D has the least useful information by itself and probably least information that isn't present in the other features. Values shown are averages over 30 replicate runs.

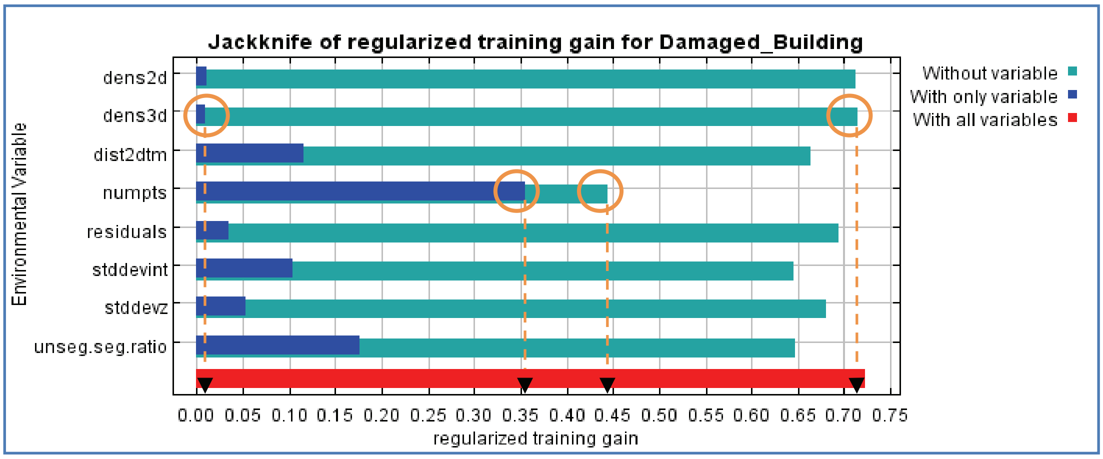


Figure 4-1 Jackknife test of variable importance over replicate runs for full Model training data

The next figure (4-2) shows the jackknife test, using test gain instead of training gain. Note that, conclusions about which feature are most important could be changed, because now we are using test data which are different than training data. However in this case, as usual, there are some changes in the achieved gain but order of features importance are the same and also the most and the least informative features like number of points per segments and density in 3D are the same as well. This can be interpreted as reliable results. As it was explained earlier, by analysing this figure, it is learned that which feature has the highest gain in isolated condition, meaning, by itself only. By having the achieved gain, for example for full model in figure 4-2, it is 0.725, it could be indicated that model can predict collapsed building 2.06 times better than random because:

$$\exp(0.725) = (e)^{0.725} = (2.718)^{0.725} = 2.06$$

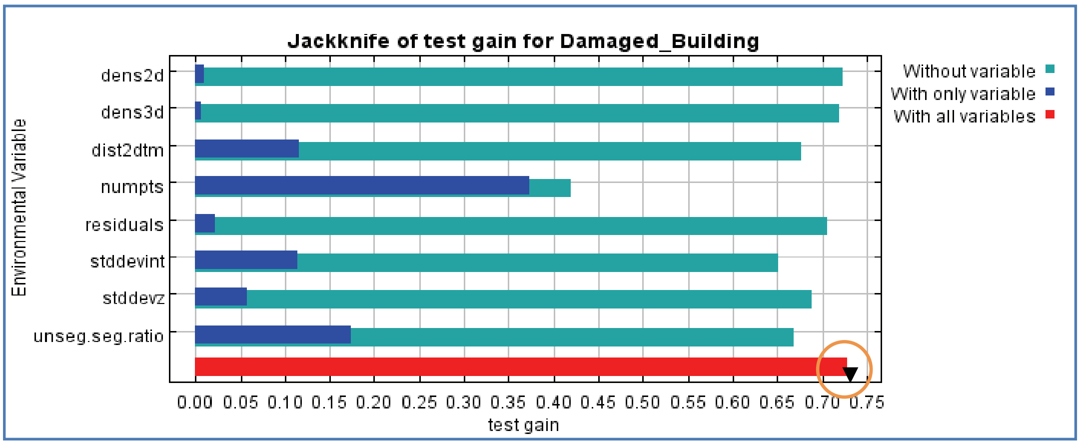


Figure 4-2 Jackknife test of variable importance over replicate runs for full Model test data

Finally in figure (4-3) we have the result of jackknife test, using AUC on test data. It is indicated that number of points is the most informative feature for prediction of collapsed building which, by omitting this feature the achieved AUC is reduced by approximately (0.075) because the gain for full model is equal to 0.82 and the gain for numpts excluded model is equal to 0.745 and by subtracting them we will have: 0.82 – 0.745 = 0.075

This is considerable amount of decrease (almost 10%) for AUC.

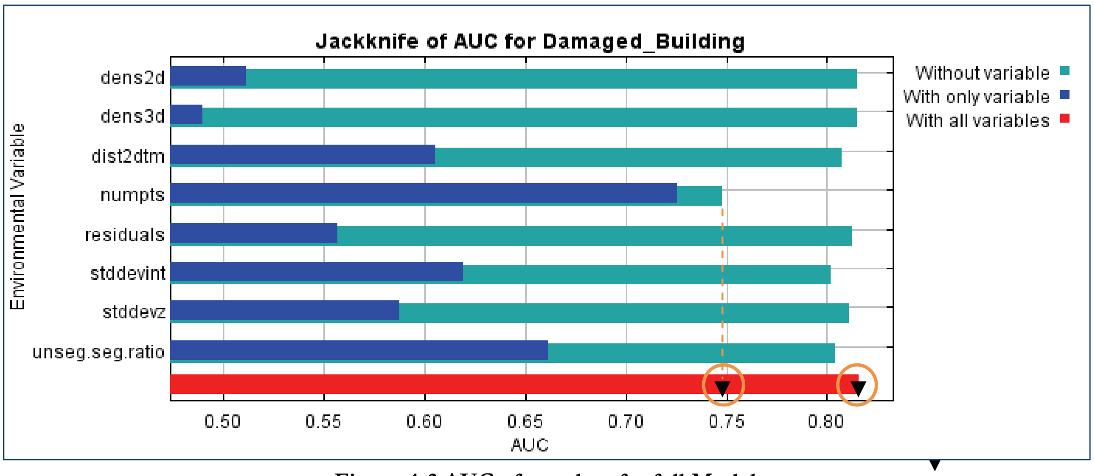


Figure 4-3 AUC of test data for full Model

4.2.2. Feature Profiles

Response curve of features was produced by employing bootstrapping method. As it is shown in figure (4-4) the average probability of collapsed building response to the value of features are presented in red colour and also blue colour shows their variability resulted by 30 times replication. Response curve imply how feature affect the probability, in other word they indicate how probability changes when feature are varied and also in what manner that is. As it can be seen in figure4-4, some classes within some range of their variation have more potential for collapsed building.

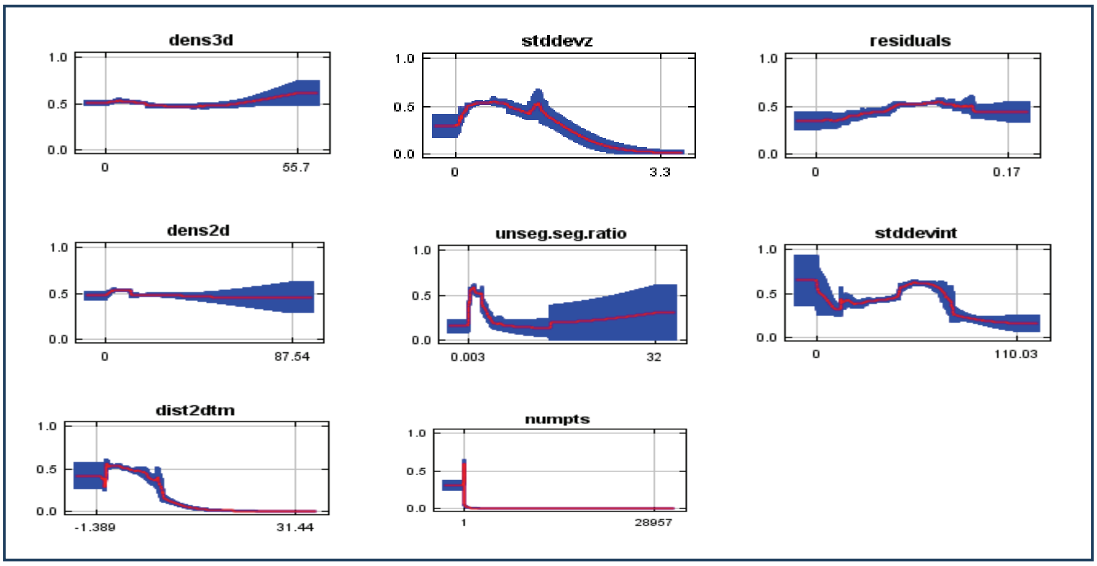


Figure 4-4 Collapsed building response shape to each predictor

As it is shown in figure 4-4, the effects of features in prediction of collapsed building is a function of their value and if they have more potential for occurrences of the collapsed building the response curve will reflect that, for example in a feature like number of points per segments as it is seen for a limited range (0-100) the curve is high and in the rest of area it is low, also for some features like density in 2d, the horizontal red line indicates that by changing in density of 2d there isn't significant change in probability of collapsed building occurrence. Same reasoning is applied for other features profile.

4.2.3. Analysis of Feature contributions

The following table gives an estimate of relative contributions of the features to the Maxent model including full features. To calculate the contribution, increase in regularized gain in the training algorithm is added to the contribution of the corresponding variable, or subtracted from it if the change to the training regularized gain is negative. Values shown are averages over replicate runs.

Features	Percent contribution
numpts	43.3
Unseg.seg.ratio	19.3
Dist2dtm	13.4
Stddevint	11.6
StddevZ	6.6
residuals	3.4
dens2d	1.3
dens3d	1.1

Table 2 Feature Contribution for full Model

4.2.4. Feature selection

The result of model evaluations indicates the power of prediction by model and shows that how much better than random is the performance of model for prediction the class of interest. Such information is really important in selecting the best suitable set of features according to parsimonious rules. In order to have the best suitable combination of features, stepwise selection method was employed to find final model. Stepwise procedure was started by analyzing the results of 30 bootstrapping of full model including all features in which 0.72, 0.725, 0.875 and 0.82 were achieved for training gain, test gain, training AUC and test AUC respectively, the least contribution was 1.1% for density of points in 3D. In the next by stepwise omitting the features with minimum contribution and re-run the model all the steps of first run are repeated to omit the other least informative feature.

The achieved results of this stepwise method were:

I. By omitting the density of point in 3d, training gain was 0.73, test gain was 0.745, training AUC was 0.874 and test AUC was 0.81 and minimum contribution of 1.7% for dens2d (**model 2**)

II. By omitting dens2d, training gain was 0.72, test gain was 0.72, training AUC was 0.873 and test AUC was 0.815 and minimum contribution of 3.6% for residual to planarity (**model 3**)

III. By omitting residual to planarity , training gain was 0.67, test gain was 0.71 training AUC was 0.81 and test AUC was 0.861 and minimum contribution of 8.2% for StddevZ (**model 4**)
Also finally

IV. By omitting StddevZ , training gain was 0.625, test gain was 0.66, training AUC was 0.849 and test AUC was 0.805 and minimum contribution of 12.8% for Stddev of intensity (**model 5**)

As it was mentioned, by omitting residuals (**model 4**) AUC (0.861), test gain (0.71) and training gain (0.67) were showing slightly decrease but the achieved results were still acceptable, Therefore, regarding the parsimonious rule (less is better), this model is preferable especially when In the next step the Stddev of Z was omitted (model 5), the AUC, test gain and training gain were decreased by 1.2%, 5% and 4.5% respectively which are considerable changes. Consequently, based on the result of last step, the final set of selected features will be those which were used in model 4. These features and their contribution **are** presented in the Table 3.

Following these tests in order to compare the influence of number of points per segments for modelling, we re-run the model without this feature which the result as model 6 has been included in the figure (4-5). The comparison of performance measures including gain and AUC of test and training data for different model in stepwise feature selection procedure is presented in Figure (4-5).

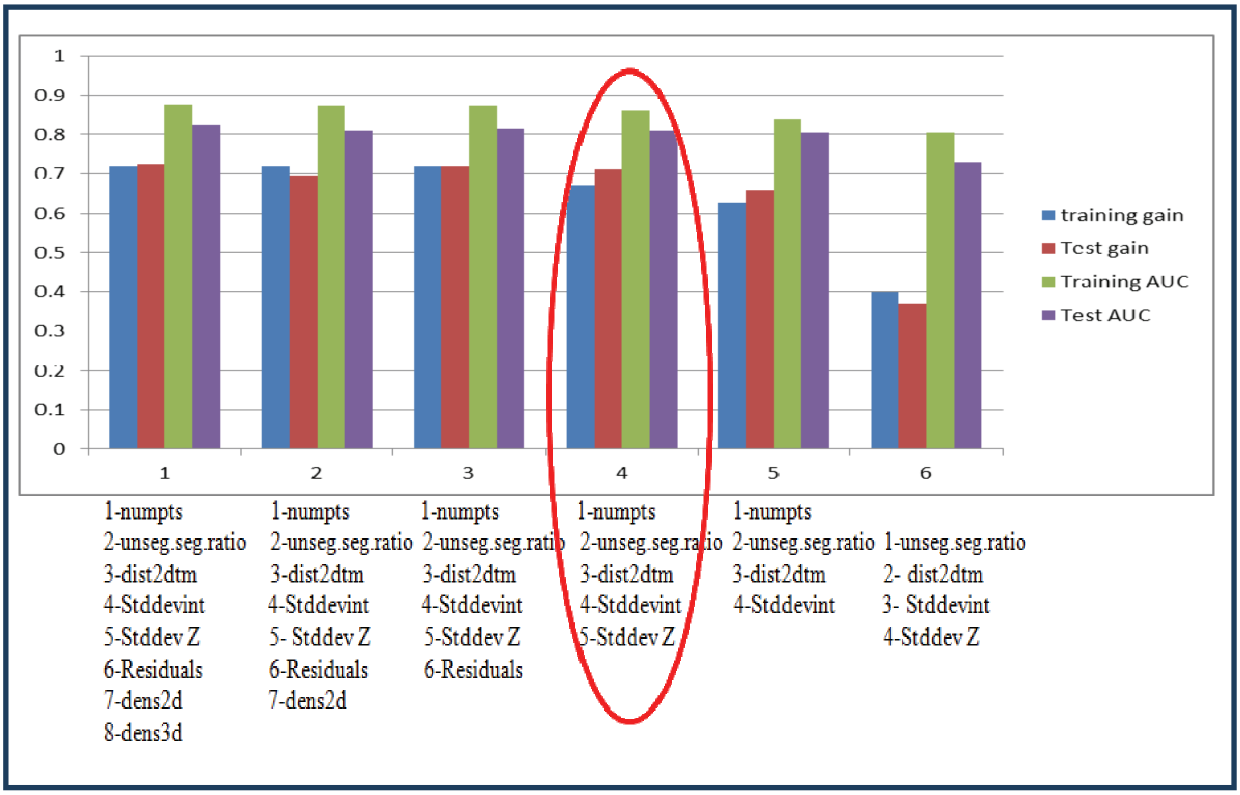


Figure 4-5 Comparing the results of jackknife test for Stepwise feature selection

Model evaluation's information is used to select the best features out of a subset of potential full models. in the figure 4-5 it can be seen that by stepwise omitting the features, the results are changed, however, we are looking for minimum size of feature but after a few step (i.e. model 4), the new step could influence the model relatively more than previous steps, caution should be used to not loss the informative feature in such condition. To find the best and optimize model there are some method suggested in statistic like, Akaike information criterion (AIC) and the Schwartz information criterion (SIC), due to limited time we have only recommended to use those methods in future works.

4.3. Final Model

Based on the results of the Stepwise selection, final set of the predictors have been chosen including number of points per segment, unsegmented per segmented points ratio, distance to DTM, Stddev of Z and Stddev of intensity. Model including new selected features was re-run by replicating 30 times of bootstrapping selection of training and test data and finally the averaged of achieved results have been used to evaluate the final model.

4.4. Evaluation of the final Model

In order to evaluate the performance of the model, two approaches were used: first, area under curve (AUC) of receiver operating characteristic (ROC) plots, which is a threshold- independent method. Second, threshold dependent measures (Kappa & TSS) uses binary distribution considered collapsed and non-collapsed areas.

Table 3 presents an estimate of relative contributions of the features in the Maxent model:

Features	Percent contribution
numpts	45.2
Unseg.seg.ratio	21.1
Dist2dtm	13.1
Stddevint	12.4
StddevZ	8.2

Table 3 Final selected features and their contribution

4.4.1. Threshold-independent evaluation results

The following figure shows the receiver operating characteristic **(**ROC) of training data for final model. As it is illustrated, curve is a plot of sensitivity versus 1–specificity and provides an easy way to assess difference. The red colour indicates the average ROC of training data, and the blue colour represents its variability.

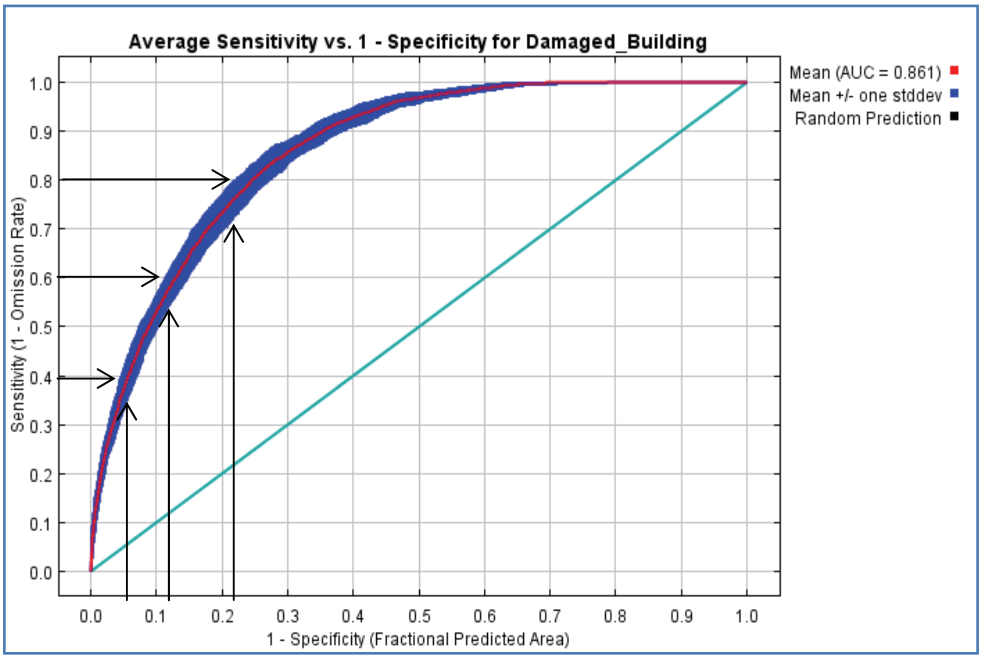


Figure 4-6 Results of (ROC) curve over the replicate runs

As it is presented in figure 4-6, the rate of true-positive (axis Y) is higher than false positive (axis X), thereby, the resulted ROC shows, how much our classifier predicts better than random, as it is shown the averaged AUC resulted from 30 times replication of model is 0.861 with a standard deviation of 0.010 and by comparing to random AUC (i.e. 0.5), it is expected to predict better than random, however, as it was reviewed, according to Swets (1988) who believes that AUC between 0.7–0.9 has a good agreement, could be interpreted to an acceptable classifier with a prediction significantly better than random.

The following figures show the results of the jackknife test for training and test data of variable importance. Again in this run, numpts with a gain about 0.34 has the highest gain when used isolated, which therefore appears to have the most useful information by itself. The feature that decreases the gain the most when it is omitted is numpts as well, which therefore appears to have the most information that isn't present in the other variables. Values shown are averages over replicate runs.

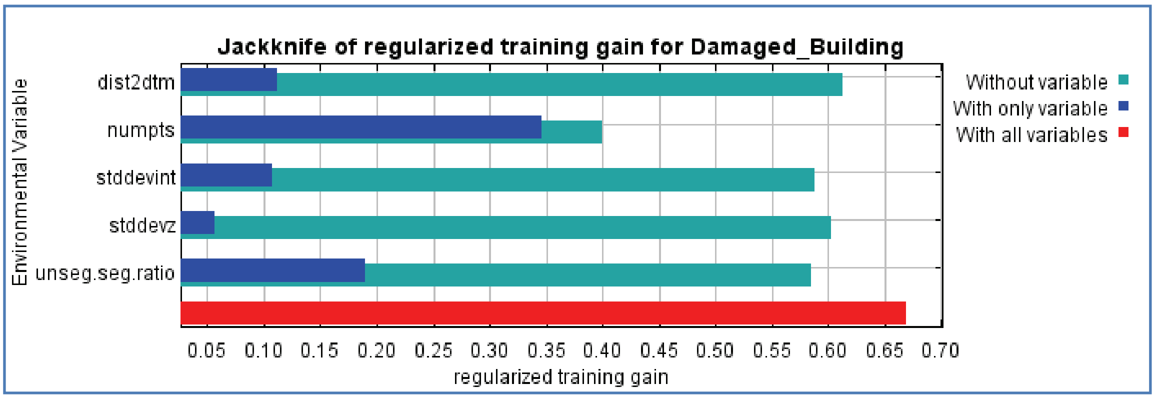


Figure 4-7 Jackknife test of variable importance for final selected Model training data

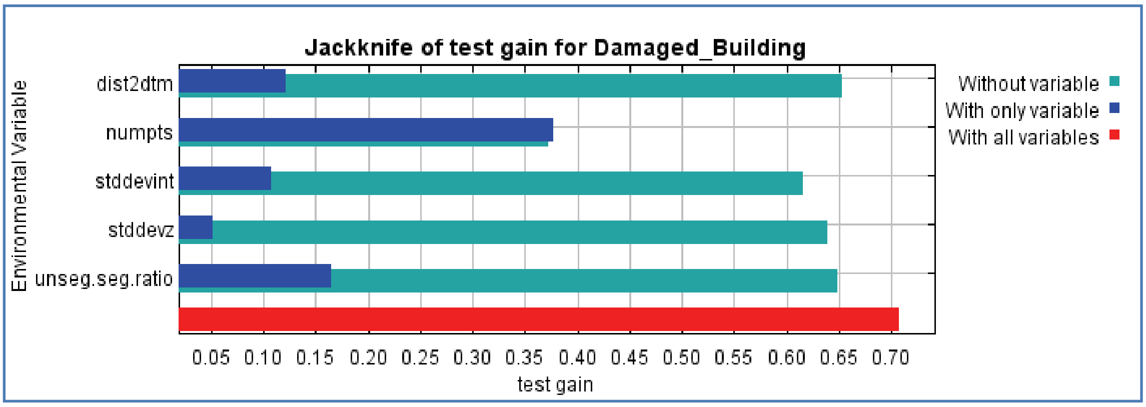


Figure 4-8 Jackknife test of variable importance for final selected Model test data

And finally the jackknife test, using AUC on test data of final Model.

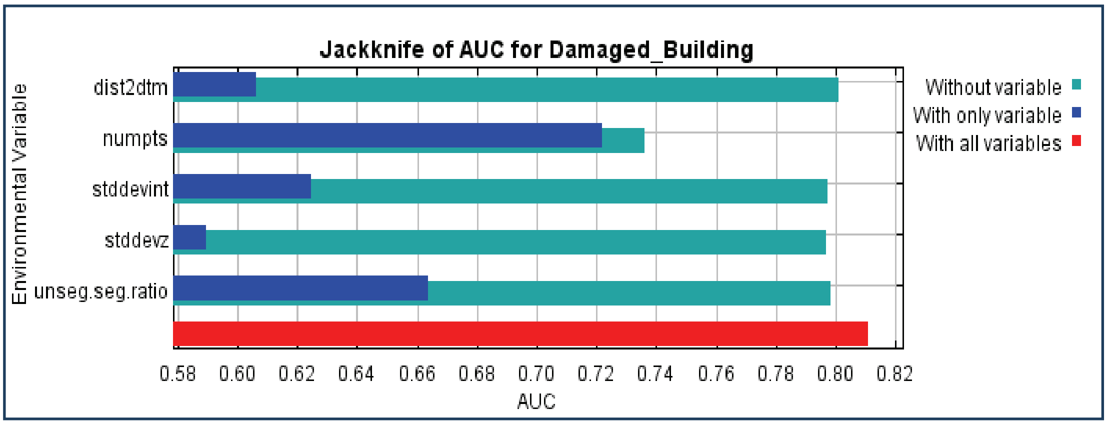


Figure 4-9 AUC of test data for final selected Model

4.4.2. Threshold-dependent evaluation results

4.4.2.1. Kappa coefficient

After applying a threshold in which by such applied value, the output of classification in a format of probability will be divided in two partition of positive and negative (suitable and unsuitable for class of interest). For instance by applying 0.51 all the segments with probability more than 0.51 will be assigned as member of collapsed building class and also all the segments with probability less than 0.51 will be assigned as member of non-collapsed building class. Model performance was investigated using the omission rate; Table 4 shows the confusion matrix of this threshold-dependent evaluation. However kappa coefficients were sensitive to thresholds and by changing the threshold we had different kappa coefficient, by applying 0.51 Threshold the Maxent obtained an average of 0.46224 for kappa coefficient. This test was repeated for results of 30 replication bootstrapping and kappa coefficients are shown in figure 4-10. The kappa coefficient is relatively low which will be compared to results of (Shoko, 2010) and also will be discussed.

observed / predicted	1	0
1	1594	1102
0	475	2917

Table 4 Confusion matrix

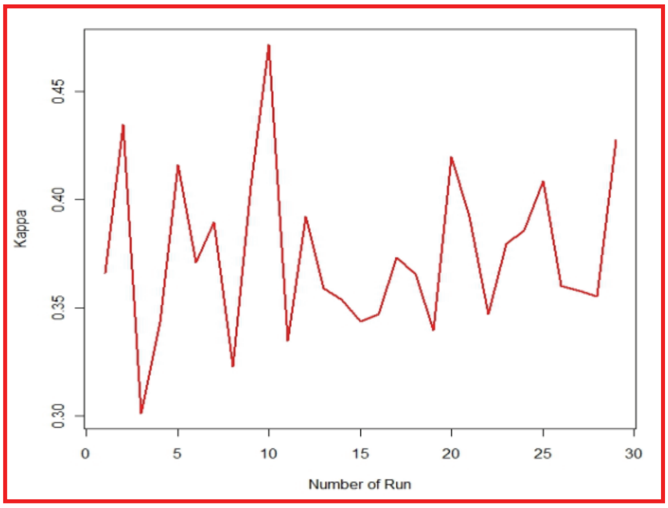


Figure 4-10 Kappa coefficients for different run

4.4.2.2. TSS coefficient

Following Allouche et al. (2006) in order to evaluate our model with a prevalence non-sensitive measure, We tried to employ an alternative measure of accuracy which is called true scale statistics (TSS) , the average of achieved results of 30 bootstrapping was 0.519 and changes are shown in figure 4-12. The achieved results for TSS show promise for predictability of Maxent by Lidar data.

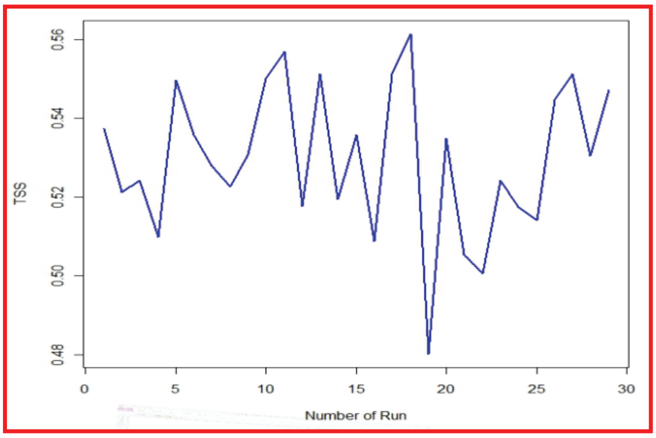


Figure 4-11 TSS coefficients for different run

4.5. Result of test for sample size

Generally by employing bootstrap method 75% of data is randomly chosen to train and the rest 25% is used to test the method, in our experiment we began using 70% as training part and 30% for test and by increasing the percent of test we used 40%, 50% and even 60% of randomly chosen data for test and results were satisfying the less sensitivity of Maxent to small size of training data, the final result of this test showed same important features as the full model and the final AUC calculated for test data was 0.78 which shows acceptable performance of Maxent by small size of sample data.

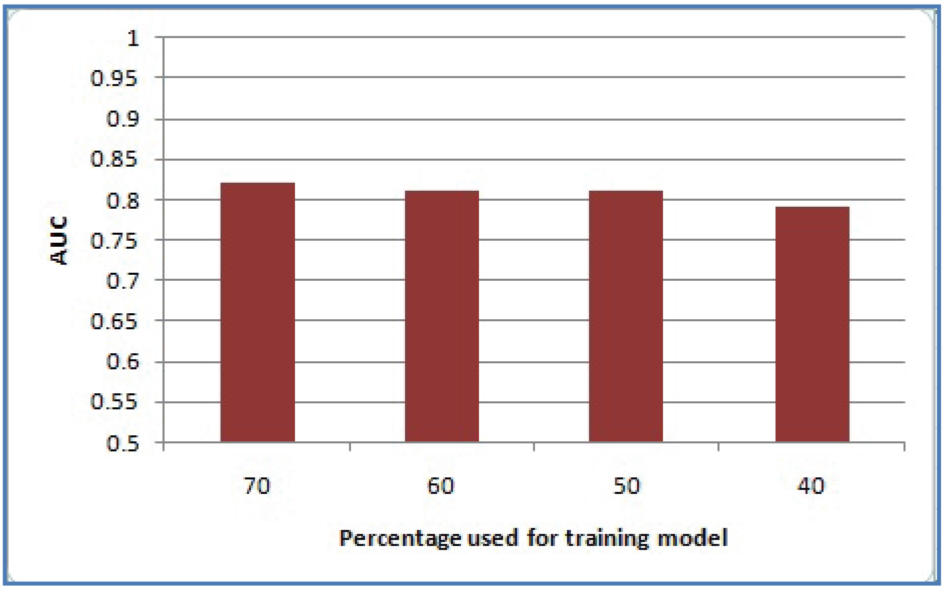


Figure 4-12 Comparing the results of different percentage used for training

4.6. Geographic distributions

By applying a threshold we can have the results of positive and negative points for class of interest, the following maps show predicted potential geographic distributions of collapsed building in which red dot are suitable for collapsed class and green dot are suitable for non-collapsed area.

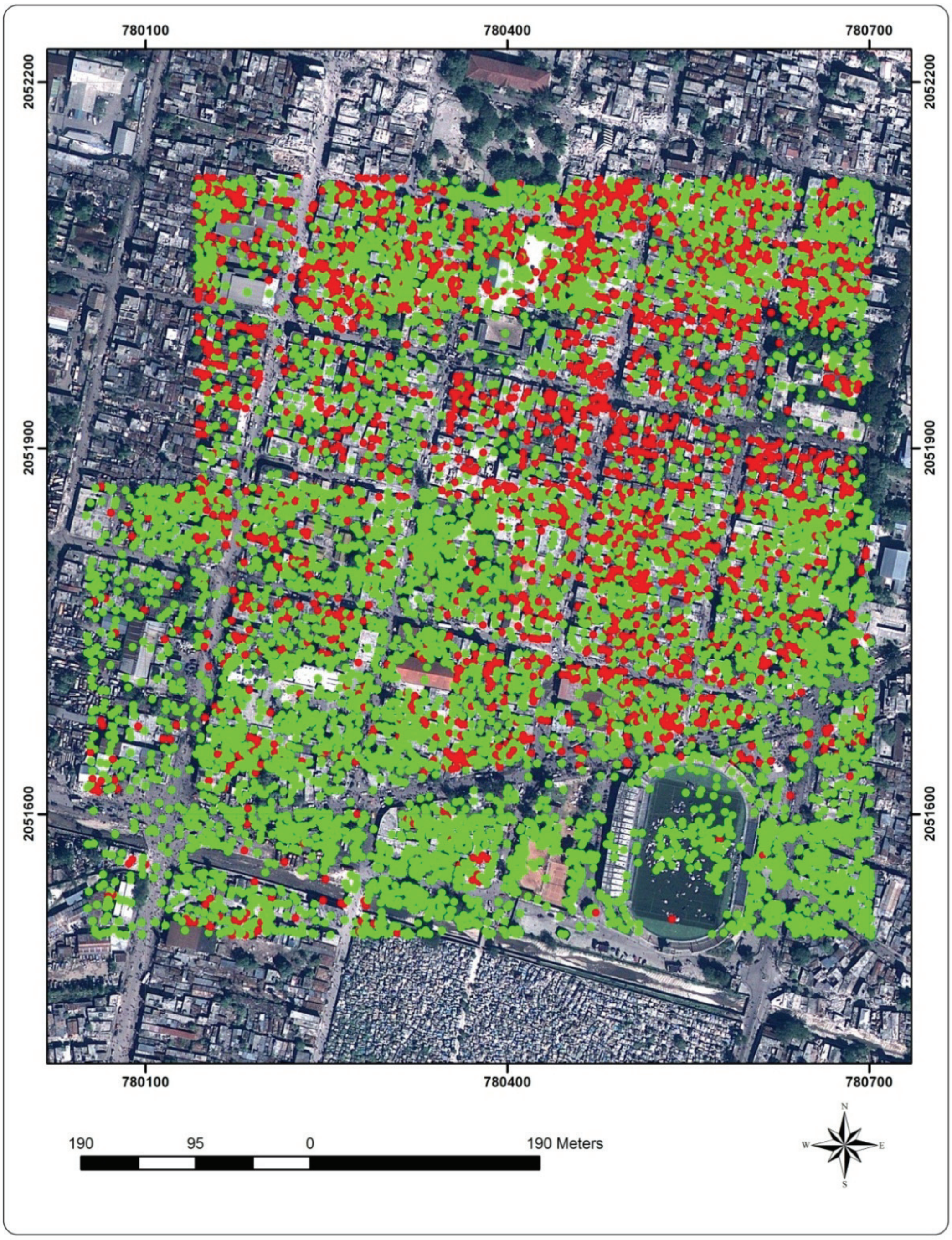


Figure 4-13 Geographic distribution of collapsed area presented by red dot of area 1

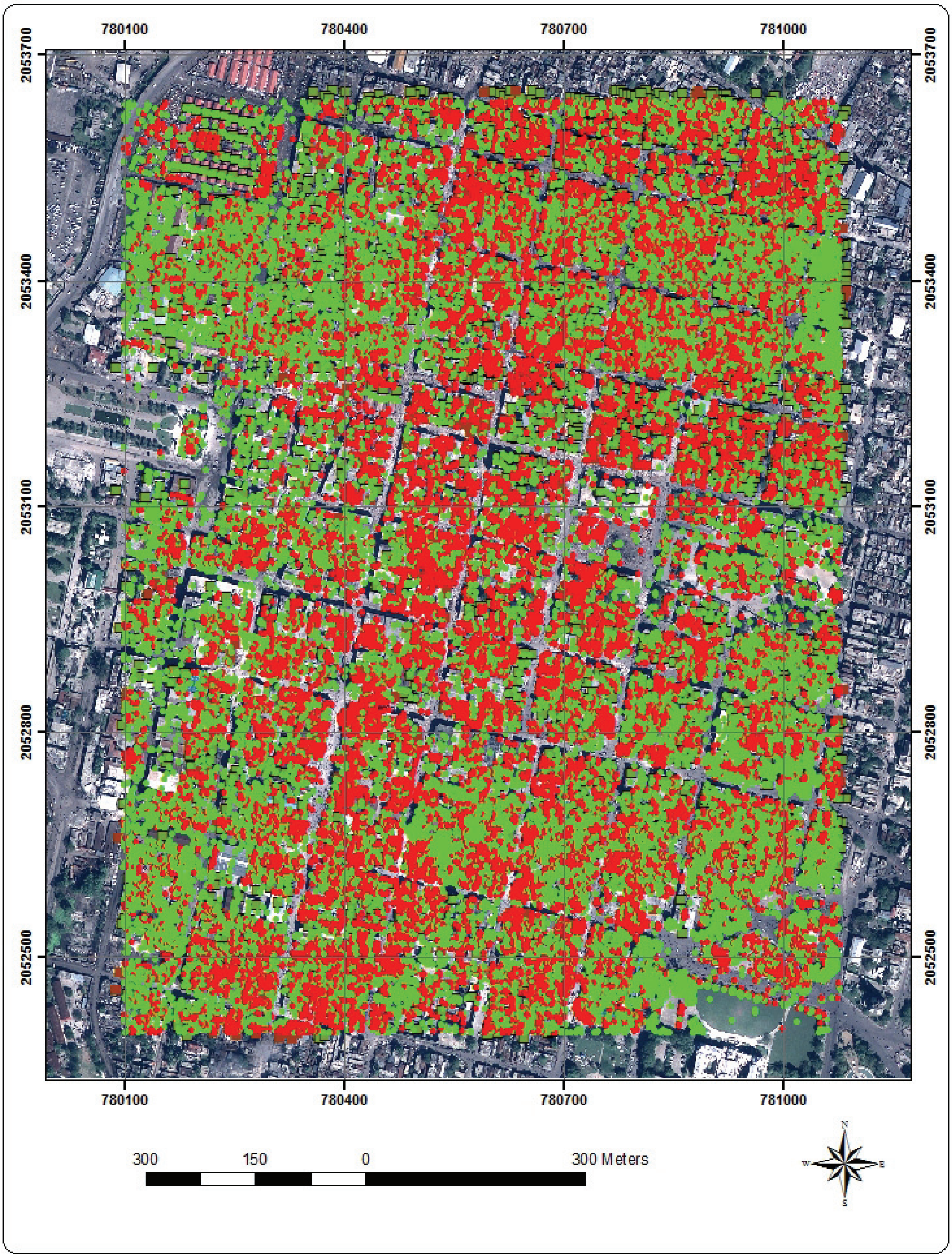


Figure 4-14 Geographic distribution of collapsed area presented by red dot of area 2

In order to have a visual inspection, some images with more details, explaining the involved data are represented here, corresponding points in these two images, show how segments belong to collapsed and non-collapsed area have been detected correctly, in these images segments of trees which have been detected as non-collapsed area, are interesting since almost same layout of trees can be found in corresponding points of other image.

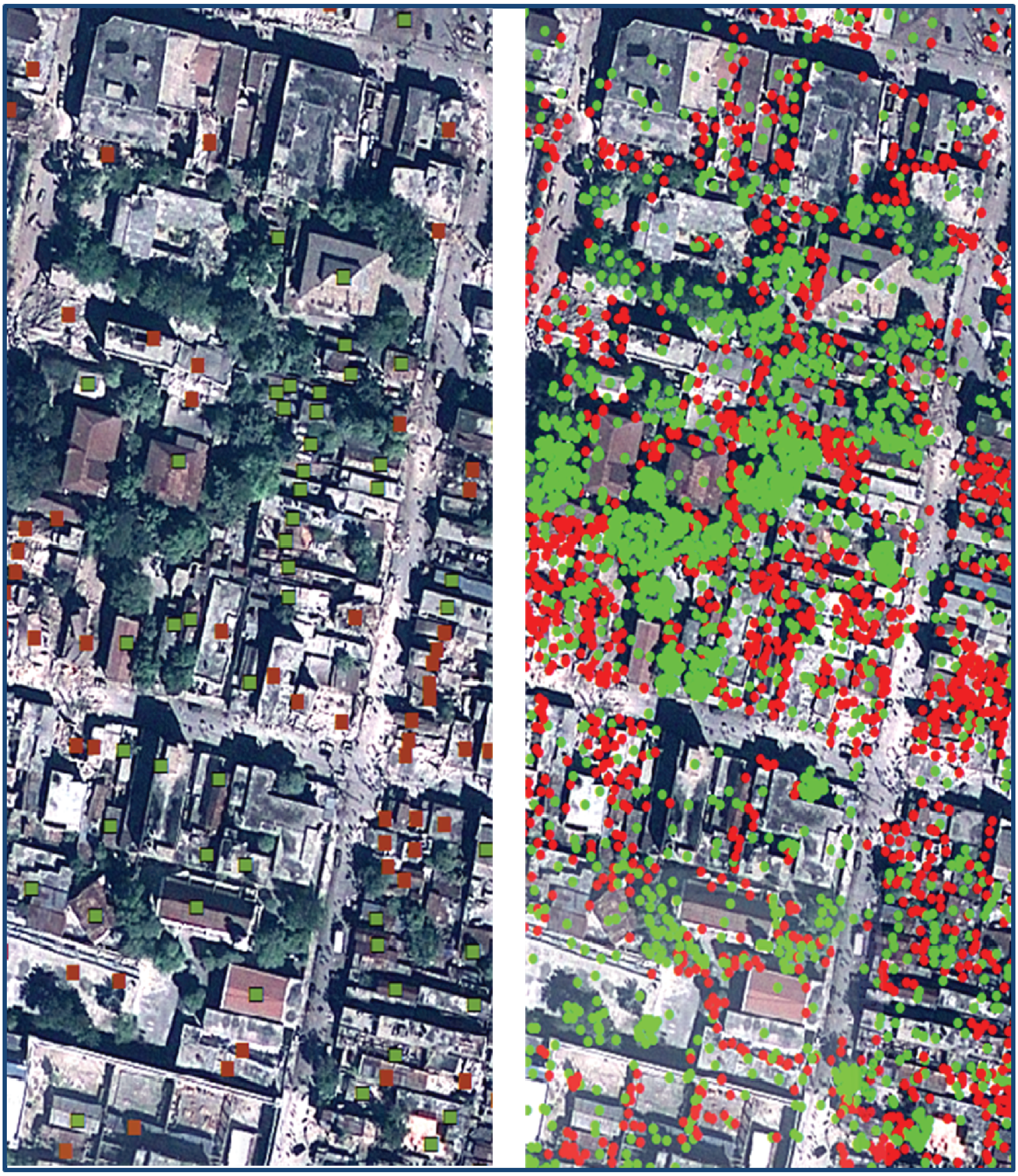


Figure 4-15 Visual inspection of output for corresponding points in two images

- Training points as collapsed
- Predicted points as collapsed
- Predicted points as non-collapsed
- Validated point as non-collapsed by UN (not used for training)

Also in figure (4-17) to (4-18) the predicted suitable for collapsed buildings and non-collapsed buildings are shown with more detail, considering the agreement between detected points and class of ground truth, shows to be a reliable prediction.

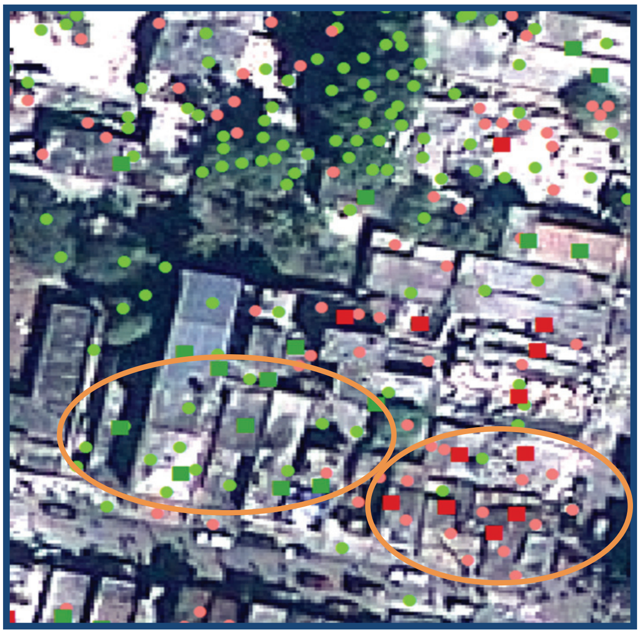


Figure 4-16 Plot of predicted points comparing to UN validated data

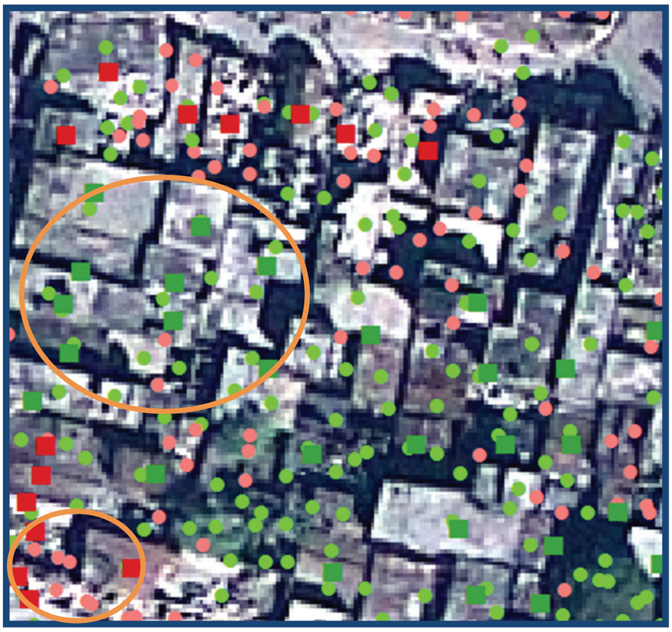


Figure 4-17 Plot of predicted points comparing to UN validated data

- Training points as collapsed
- Predicted points as collapsed
- Predicted points as non-collapsed
- Validated point as non-collapsed by UN (not used for training)

In figure 4-19 the results of some research for collapsed building detection using aerial images are superposed to Maxent's result. The comparison of their results with Maxent's output, as it is shown, indicated that, however, there are some agreements, but they defer is some manner, the main defer are in approaches since, the scale of details for Maxent as it is seen is segments and more details are expected but the other's output are block but no details. It is recommended for future works to use some extra classifier in order to have an interpretation of results, based on a fuzzy logic classification. Also by projecting the corresponding point cloud of detected segments, instead of just centre, real location and shape of segments will be mapped. That will be easier for interpretation and of course more accurate.

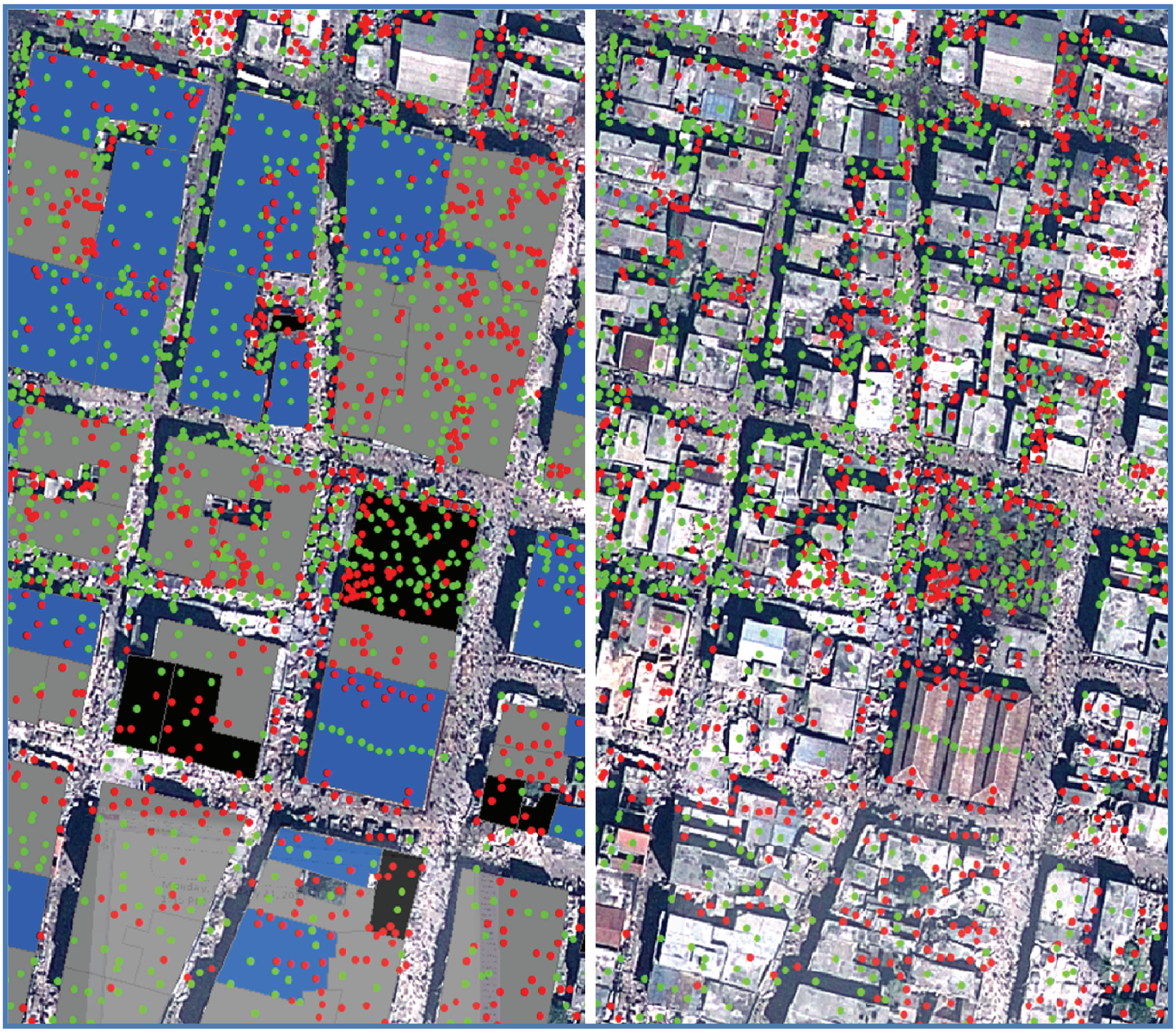


Figure 4-18 Comparing the results of Maxent and the method using aerial images

- No damage detected
- Partly affected area
- Fully demolished area
- Predicted points as collapsed
- Predicted points as non-collapsed

4.7. Type 1 and type 2 errors

Investigation was carried out to find the source of errors. As it is shown in Figure 4-21 we could say, in addition to the wrong reference data which is not impossible, the main sources for error type 1 is the similarity of segment to collapsed area, as it can be seen in figure 4-20 some of segments in the edges of non-collapsed building or in some small size of building have been detected as collapsed segments.

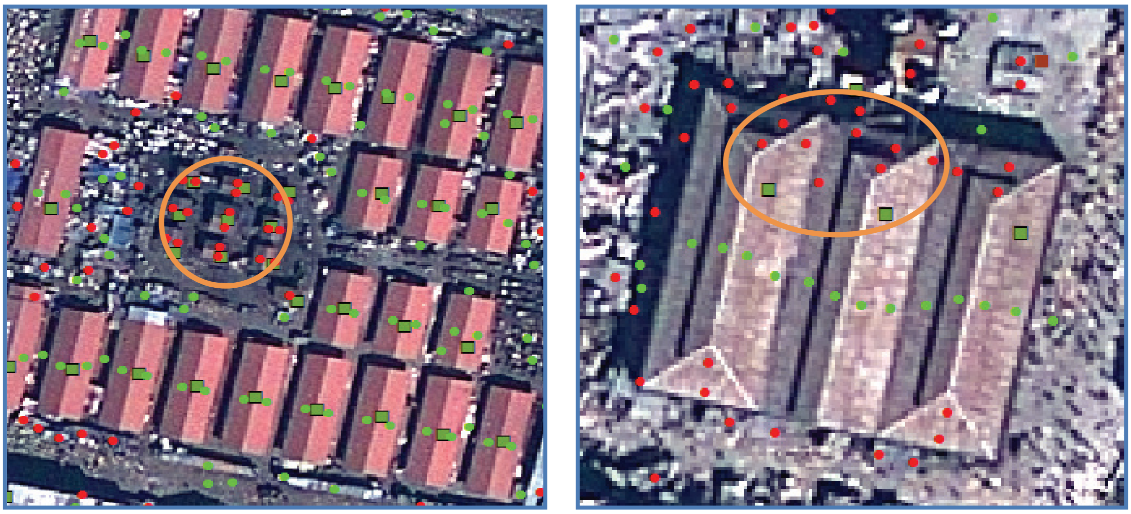


Figure 4-19 Type 1 errors

It is notable that, in addition to the wrong reference, some of type 2 errors (collapsed building predicted as non-collapsed) were due to some neighboring segments, meaning that if there is any tree besides a collapsed building, segments of that tree could be evaluated as type 2 errors. Also pancake collapsed, in which, whole roof is moved to down, could cause such error as well. Finally the centre of large segments like road in some particular shape could be outside of segment and projected in the collapsed area, subsequently detected as non-collapsed, which can be a source of type 2 error.

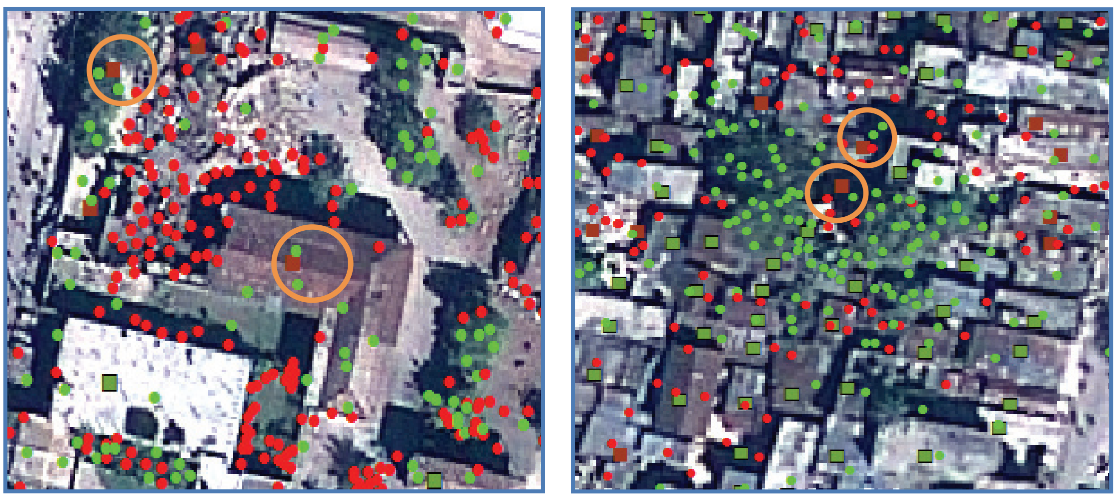


Figure 4-20 Type 2 error

5. DISCUSSION

This study examined the potential use of Lidar data to identify collapsed building. A Maximum Entropy Approach for the classification of damaged building after disasters like earthquakes was presented. It is based on the statistical data distribution model building from Lidar data. However, to our knowledge, its applications in Lidar data classification are rarely studied. Therefore, we propose this approach to one-class classification of Lidar data. As it was mentioned, in addition to several advantages of Maxent, it was a good innovation to test its application in the field of Lidar. It starts with a segmentation of planar surfaces, followed by implementation of Maxent for classification of segments. Finally, these segments are assigned to collapse building according to their calculated features.

The results from this study revealed that the number of points per segment was the most important predictor feature to detect the collapsed building. This concurs with (Rehor, 2007), who used similar feature (segment size) to classify building damaged based on Lidar data. It could be expected that the number of points per segment decreases when the building is collapsed comparing with non-collapsed building or road which mostly have large segments. In contrast, trees mostly have small size of segments including some non-segmented points due to rough surface. However, distinguishing between tree and collapsed area has been the main concern. By using some additional features we have been able to overcome this concern. It could be recommended to use multi pulse Lidar data to have more accurate result, since in that case we will be able to separate vegetation by filtering last pulse.

The results showed that some features have no significant influence to extract the collapsed building class. These features include density in 2D, density in 3D and also residual to planarity. To explain such result if we consider the rigid non- penetrable surface of building and also single return pulse of Lidar, it is reasonable to say that density in 2D is a function of our sensor and flight plan and should be same for all covered area. Same reasoning is applied for 3D density, meaning that 3D density could be an indicator if we had multi returned pulse, in that case penetrated pulse inside trees could make some changes in 3D density. For the third omitted feature (residual to planarity) it could be attributed to the fact that this feature has been limited to a value during the parameter setting of segmentation. This means that, distance to planarity higher than this limit has already influenced our data that creates different number of segments and this might be the reason that these two features (residual to planarity and segment size) are correlated.

Comparing the results of this study with Shoko (2010), revealed that the final selected features in this study concur with the features that has been introduced in her study (i.e. Standard deviation of intensity, mean number of points per segments, height above DTM, unsegmented points ratio) except residual to planarity. She used this feature as an indicator counter for class number. It might be interesting that these two studies had two different approaches but the same results are achieved.

Comparison of the accuracy of Maxent's Threshold-dependent approach used in this study (i.e. confusion matrix table 4) with the results achieved by Shoko (2010) (i.e. 404 true positive (TP), 74 false positive (FP) and 102 false negative (FN)) is Presented here:

	Shoko (2010)	**Maxent**
Completeness/Producer accuracy	79.8%	77%
Correctness	85%	59%
Quality	70%	50%

Table 5 Comparison of accuracy between Maxent and Shoko (2010)

As table 5 shows, Maxent had relatively lower accuracy for completeness, correctness and quality.

However, as it has been addressed in method section, using confusion matrix based measure by applying a threshold is subjected to a challenge as different threshold results different measures of accuracy. It can be recommended using a threshold-independent measure (i.e. AUC).

Maxent algorithm performed significantly better than random classification as has been evaluated by ROC. The area under the ROC curve (AUC) for the final set was 0.81 which can be considered as a useful model. Swets (1988) recommended that the models with AUC ranged between 0.70-0.90 are good and therefore, Maxent proved that has the capability for detection of collapsed buildings based on post-event ALS data.

The threshold independent evaluation achieved through ROC for data of a test area containing real building damages are very promising. In this study, the proposed Maxent shows promise in one-class classification while it does not require negative data for training. The input to Maxent is only a set of positive samples from a target distribution, as well as a set of known constraints on the distribution. Hence, it can significantly reduce the effort of manually collecting training data for classification. In our case by using single pulse Lidar data with approximate density of 2 points per square meter, the Maxent classifier with the final five suitable attributes selected as input for classification has the highest margin while achieving a good accuracy. This confirms that these features are the most discriminating for collapsed building classification. This study also showed that some features are not of significance to extract the collapsed building classes by using the existing single pulse Lidar data, features such as density of points in 2D, density of points in 3D and also residual to planarity. Final Results were satisfactory (AUC = 0.81) for collapsed building Classes prediction. Besides, the algorithm is a user-friendly method with easy adjustable parameters. Furthermore, we successfully used this powerful method to estimate the importance of features to classify collapsed buildings. The permutation accuracy criteria also revealed that the most significant feature is number of points per segment following by unsegmented per segmented point ratio, distance to DTM, Stddev of intensity and finally Stddev of Z. The Maxent started with full model including all features but finally was re-run with the most important features. The fact that variable importance in Figure (4-7) is highly similar to previously achieved one in (Figure 4-1), confirms the reliability of this measure. Moreover, the classification accuracy is enhanced using the five relevant features. The feature importance measure, using a balanced training data, is essential to select the best features. The results of this preliminary study indicate that Lidar data contain a significant amount of information regarding the earthquake-induced damage to buildings, these data will play a role in earthquake reconnaissance efforts, and will aid in identifying damage patterns across affected area.

It is also tested in this study if sample size is a limitation for this approach (i.e. Maxent). The results indicated this approach is not sensitive to sample size, and therefore can be considered as an advantage when there is a limitation for collecting field data.

However, much work can be done to refine the use of Maxent for modelling collapsed building distributions, but considering the difficulty which we had due to the characteristic of used Lidar data I should recommend using multi pulse and full-wave form of Lidar data and also higher density of points per square meters (20-30) as input Lidar data, to achieve better and more accurate Model.

6. CONCLUSION AND RECOMMENDATION

In this study three research questions were supposed to be answered.

The first research question involves the most important suitable features for distinction of damaged buildings. The results that are achieved by employing stepwise selection as part of model implementation and by using statistical method (jackknife test) indicated that most suitable features were the number of points per segments, and then unsegmented per segmented point ratio, distance to DTM, Stddev of intensity and Stddev of Z, respectively.

The second research question was: whether maximum entropy approach can be used as a one-class classifier to detect the damaged buildings based on ALS data?

In order to provide answer to this question investigation was conducted among some of existing statistical classification and their input, output and characteristics and also considering our limits in providing reference data. The Maxent, as a new approach for one-class classification of Lidar data, was used in this study and the results of evaluation indicated that this approach is useful to detect the collapsed buildings.

The evaluation of the chosen approach provided the response to the third research question to find that, which level of accuracy can be achieved by the employed method? A threshold- independent method of ROC as well as threshold- dependent measures (i.e. kappa and TSS) indicated that how this technique is accurate and how it can be optimized by selecting the relevant features.

Although Maximum entropy approach shows promise to become a very useful tool for damage detection from Lidar data, but it is still in early stages and needs to be developed in some manners.

I would recommend several aspects in order to increase Maxent's utility in collapsed building detection from Lidar data. These include:

1. Developing methods for model selection in a manner of using better statistical method like the Akaike information criterion (AIC) and the Schwartz information criterion (SIC) to measure a model's suitability and goodness of fit
2. Developing comprehensive Maxent model in a manner to make it transferable to other regions.
3. Determining the minimum number of occurrence localities of collapsed building needed to make an adequate prediction (as always in disaster situation).
4. Developing protocol for selection of appropriate threshold values.
5. Developing Maxent in a manner to be exported based on a fuzzy logic to have an output which has been interpreted for whole buildings or even whole blocks (as it was shown in

figure4-19), in the same approach we can map the point cloud instead of centre of segments to create an accurate geographic distributions of output in its details.

LIST OF REFERENCES

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology, 43*(6), 1223-1232.

Baldwin, R. (2009). Use of Maximum Entropy Modeling in Wildlife Research. *Entropy, 11*(4), 854-866.

Bitelli, G., Camassi, R., Gusella, L., & Mongnol, A. (2004). Image change detection on urban area: the earthquake case. 692.

Blagojevic, B. (2010). Modeling of Fluorescence Lidar ROC Curves. *Sensors Journal, IEEE, 10*(3), 772-778.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32.

Chehata, N., Li, G., & Mallet, C. (2009). *Contribution of airborne full-waveform lidar and image data for urban scene classification.* Paper presented at the Image Processing (ICIP), 2009 16th IEEE International Conference on.

Darmawati, A.T. (2008) Utilization of multiple echo information for classification of airborne laser scanning data.Master's Thesis, Enschede, ITC, 2008.

Dudik, M., Phillips, S. J., & Schapire, R. E. (2004). Performance guarantees for regularized maximum entropy density estimation. In J. ShaweTaylor & Y. Singer (Eds.), *Learning Theory, Proceedings* (Vol. 3120, pp. 472-486). Berlin: Springer-Verlag Berlin.

Fawcett, T. (2004). ROC Graphs: Notes and Practical Considerations for Researchers.

Fernandes, A. M., Utkin, A. B., Lavrov, A. V., & Vilar, R. M. (2004). Development of neural network committee machines for automatic forest fire detection using lidar. *Pattern recognition, 37*(10), 2039-2047.

Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation, 24*, 38-49

Filin, S. Surface classification from airborne laser scanning data. *Computers & Geosciences, 30*(9-10), 1033-1041.

Guo, L., Chehata, N., Mallet, C., & Boukir, S. Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. *ISPRS Journal of Photogrammetry and Remote Sensing, In Press, Corrected Proof.*

Han, S., Lee, J., & Yu, K. (2007). An approach for segmentation of airborne laser point clouds utilizing scan-line characteristics. *ETRI journal, 29*(5), 641-648.

Hernandez, P. A., Graham, C. H., Master, L. L., & Albert, D. L. (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography, 29*(5), 773-785.

Jaynes, E. T. (1957). Information theory and statistical mechanics. II. *Physical review, 108*(2), 171-190.

Millar, L. (2010, 17 January 2010). Tens of thousands isolated at earthquake centre. Washington Post Newspaper.

Li, W., & Guo, Q. (2010). A maximum entropy approach to one-class classification of remote sensing imagery. *International Journal of Remote Sensing, 31*(8), 2227 - 2235.

Lodha, S., Kreps, E., Helmbold, D., & Fitzpatrick, D. (2006). *Aerial LiDAR Data Classification Using Support Vector Machines (SVM).*

Lodha, S. K., Fitzpatrick, D. M., & Helmbold, D. P. (2007, 21-23 Aug. 2007). *Aerial Lidar Data Classification using AdaBoost.* Paper presented at the 3-D Digital Imaging and Modeling, 2007. 3DIM '07. Sixth International Conference on.

Mallet, C., & Bretar, F. (2009). Full-waveform topographic lidar: State-of-the-art. *ISPRS Journal of Photogrammetry and Remote Sensing, 64*(1), 1-16.

Ogawa, N., & Yamazaki, F. (2000). *Photo-interpretation of building damage due to earthquakes using aerial photographs*.

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing, 26*(1), 217-222.

Phillips, S., Anderson, R., & Schapire, R. (2006). Maximum entropy modeling of species geographic distributions. *Ecological modelling, 190*(3-4), 231-259.

Pu, S., & Vosselman, G. (2009). Knowledge based reconstruction of building models from terrestrial laser scanning data. ISPRS Journal of Photogrammetry and Remote Sensing, 64(6), 10.

Rehor, M. (2007). Classification of building damages based on laser scanning data. *Photogrammetric Journal of Finland, 20*, 2.

Rutzinger, M., Höfle, B., & Pfeifer, N. (2008). Object detection in airborne laser scanning data - an integrative approach on object-based image and point cloud analysis. In T. Blaschke, S. Lang & G. J. Hay (Eds.), *Object-Based Image Analysis* (pp. 645-662): Springer Berlin Heidelberg.

Secord, J., & Zakhor, A. (2007). Tree Detection in Urban Regions Using Aerial Lidar and Image Data. *Geoscience and Remote Sensing Letters, IEEE, 4*(2), 196-200.

Shoko, M. (2010). Detection of collapsed buildings Using Lidar: A Case Study of Haiti, Individual Final Assignment, Enschede, ITC

Sithole, G. (2005). Segmentation and classification of airborne laser scanner data.

Sithole, G., & Vosselman, G. (2005). Filtering of airborne laser scanner data based on segmented point clouds. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 36*(part 3), 12-14.

Tóvári, D., & Pfeifer, N. (2005). Segmentation based robust interpolation-a new approach to laser data filtering. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 36*(part 3), 79-84.

TSO, B., & MATHER, P. (2009). Classification methods for remotely sensed data.

Vapnik, V., & Kotz, S. (2006). *Estimation of dependences based on empirical data*: Springer-Verlag New York Inc.

Vosselman, G., Gorte, B., Sithole, G., & Rabbani, T. (2004). Recognising structure in laser scanner point clouds. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 46*(8), 33-38.

Weston, A., Burton, P., & Lovett, A. (2003). *Imagery and Ground Truth in an Assessment of the Built Environment within an Earthquake Impact Scenario: Colchester, UK.*