

A Machine Learning Approach for Asperities' Location Identification

Kostantinos Arvanitakis · Ioannis Karydis · Katia L. Kermanidis · Markos Avlonitis

the date of receipt and acceptance should be inserted later

Abstract Asperities' location is a very important factor in spatiotemporal analysis of an area's seismicity, as they can accumulate a large amount of tectonic stress and, by their rupture, a great magnitude earthquake. Seismic attributes of earthquakes, such as the b -value and seismic density, have been shown to be useful indicators of asperities' location. In this work, machine learning techniques are used to identify the location of areas with high probability of asperity existence using as feature vector information extracted solely by earthquake catalogs (b -value & seismic density), avoiding thus any geo-location information. Extensive experimentation on algorithms' performance is conducted with a plethora of machine learning classification algorithms, focusing on the effect of data oversampling & undersampling, as well as the effect of cost sensitive classification without any resampling of the data. The results obtained are promising with performance being comparable to geo-location information including vectors.

Keywords asperity · density · b -value · seismicity · machine learning

This work is an extension of [1].

Kostantinos Arvanitakis
Department of Informatics, Ionian University, Corfu, Greece 49132
E-mail: k.arvan@ionio.gr

Ioannis Karydis
Department of Informatics, Ionian University, Corfu, Greece 49132
E-mail: karydis@ionio.gr

Katia L. Kermanidis
Department of Informatics, Ionian University, Corfu, Greece 49132
E-mail: kerman@ionio.gr

Markos Avlonitis
Department of Informatics, Ionian University, Corfu, Greece 49132
E-mail: avlon@ionio.gr

1 Introduction

The earth's crust exhibits heterogeneity at variable scales with seismic faults being the thin zone of crushed rock separating blocks of the crust. Asperities [2] are considered to be large and strong patches on a seismic fault that exhibit high stress in relation to surrounding low stress areas. Their dimensions range from less than a kilometer to tens of kilometers.

As asperities are surrounded by weaker fault zones that can sustain lower stress levels, their surrounding areas break easier and generate earthquakes at a faster pace. At the same time, asperities appear to be inactive until their stress limit is reached and they eventually break contributing thus by releasing most of the energy during the eventual earthquake. Due to their aforementioned nature and location, asperities can accumulate a large amount of tectonic stress and accordingly, by their rupture an earthquake of great magnitude is generated.

As was proposed in [3], the complete complex dynamic of asperities of seismic fault evolution can be robust modeled by means of the following stochastic differential equation for local deformation or slip γ within the plate interface (μ arbitrating damping constant),

$$\mu \frac{\partial \gamma}{\partial t} = \tau_{ext} - \tau \quad (1)$$

$$\tau_{ext} = K_L v \quad (2)$$

is the external stress accounting for the evolution of tectonic plates with elastic constant K_L and relative velocity between plates v , and τ_{int} is the internal material stress decomposed to a term

$$K \nabla^2 \gamma \quad (3)$$

modeling internal interaction between material points of elastic constant K , and a stress term

$$\tau_s(\gamma) \quad (4)$$

modeling deformation resistance (for a more detailed discussion of the model see [4]). Within this framework the spatiotemporal evolution of asperities may be robust modeled with the last term, where deterministic and random dynamic properties of asperities interaction can be taken into account, since the exact location between asperities within the fault may drastically affect fault evolution during an earthquake.

Asperities' location identification techniques have so far concentrated on two alternative research directions, the identification of a region's increased stress levels and the measurement of the surface's slip. As by definition, an asperity is a high stress area surrounded by low stress areas, the former calculates the stress levels of a region and points out the areas with higher levels [5, 6, 7]. The latter uses GPS data of the slip's distribution on the earth's surface

and the asperities are located in regions where significant deformations are detected [8, 9, 10, 11, 12, 13, 14].

The purpose of this study is to employ machine learning for the location identification of asperities. In order to further support the proposed methodology, experimentation has been conducted on data collected by Takahashi & Kasahara [15] from the region of Hokkaido, Japan.

1.1 Motivation & Contribution

An asperity's location is information of high value due to the high probability of such a location generating large earthquakes. Thus, this information can help in decision making and strategic planning of the seismic regulations on building construction and also the expansive policy of cities and civil engineering projects, in order to increase safety for human life and avoid loss of infrastructure.

Moreover, existing research on asperities' location identification has focused mainly on fracture engineering and statistical seismology. The former utilise stress levels and surface slip distribution methods while the latter attempt to use stochastic properties of asperities, such as low b -values and low density.

To address these requirements, our previous work [1] presented extensive experimentation with various classification algorithms, on different settings, in order to detect the location of an asperity. Therein, asperities were located in space using attributes that described the seismicity of the examined areas. The vector utilized therein includes the geographical coordinates of the location to be examined. Due to the spatial pattern characteristic of asperities, i.e. the clustering in space, the use of spatial coordinates within the feature vector, may produce highly robust results [1]. On the other hand the use of absolute values of geographical coordinates it is useless when trying to apply the proposed feature vector to other seismogenic faults. This is not the case for the use of the correspondingly absolute values of b -value and earthquake density features, since their range of magnitude within asperities is the same for different faults [2, 15].

This work significantly extends [1] at the process of asperities' location identification with the introduction of a machine learning approach on the information collected by statistical seismology approaches. In detail, the key contributions are:

1. introduction of a better suited vector model that only includes information related to the seismicity of an area, i.e. density and b -value features,
2. proposal for the use of random undersampling with and without the subsequent preprocessing removal of Tomek links [16] from the data,
3. examination of the performance of the proposed methodology by use of Cost Sensitive Classification on the unbalanced data,
4. execution and presentation of promising experimental results with the new feature vector.

Tomek links are pairs of borderline examples from opposite classes. Using the Nearest Neighbor (NN) method [17] pairs of examples belonging to opposite classes are initially identified. Then, examples of the majority class within the Tomek links are removed from the set of examples and thus a more focused undersampling of the majority class is achieved.

The rest of the paper is organised as follows: Section 2 presents background information and related work, while Section 3 discusses the nature and representation of data utilised herein as well as the features selected and extracted from the raw data. Next, Section 4 details with the experimental setup and the experimental results obtained. Finally, the paper is concluded in Section 5.

2 Background & Related research

This Section details necessary background information on machine learning methods as well as related existing research on asperities' location identification.

2.1 Machine Learning Algorithms

Numerous classification algorithms exist that are suitable for the purposes of experimentation on the theme of this work. As classification approaches to asperities' location identification are, to the best of our knowledge, non-existent in the literature, the choice of classification algorithms has, to a large extent, been based on exploratory criteria with the aim to cover varying learner families.

Of the total 39 classification algorithms used to test our hypothesis, the five most effective by means of precision, recall and f1-score are described in this Section.

HyperPipes [18] is a simple algorithm that can handle large volumes of data in short time. The algorithm creates a pipe for each of the available classification classes. During the training phase of the algorithm, the pipes record the instances and their respective classes without keeping record of the number of appearances of each instance in a class. The classification class a new instance will be assigned to, will be selected by the algorithm based on the similarity of the under-classification instance with the already recorded instances of each pipe. The class of the pipe with the higher number of shared recorded instances will be used for the under-classification instance.

ZeroR [19] is even more simplistic than HyperPipes. ZeroR is a rule based algorithm that does not take into consideration the attributes included in the feature vector, but focuses on the frequency of occurrence of samples from the available classification classes. Accordingly, ZeroR classifies in the majority class that is derived from the training set used during the training phase. Despite that ZeroR shows no usefulness in classifying instances, its results can be used as a lower threshold of accepted results for other approaches.

Random Forests is a tree classification algorithm introduced by Breiman [20]. This algorithm creates a forest of random trees. Random vectors are created from the training set and based on them the growth of each tree is made. The algorithm ensures that every random vector will be unique. Finally each tree votes for the class that every testing instance will be registered and the most popular class is selected from the forest. Random Forests is shown [20] to perform favorably when compared to Adaboost [21] while retaining robustness to noise in data.

The Decision Table algorithm [22] creates classification rules using the instances of the training set. Training instances are placed in a two-dimensional array and each instance is used to produce a rule that matches one of the available classification classes. Accordingly, the algorithm identifies for each instance of the testing set which of the rules is verified and assigns the instance in the respective class.

The DTNB algorithm [23] constitutes a Decision TableNaive Bayes hybrid algorithm. The algorithm repeatedly divides the feature vectors attributes in two disjoint subsets by evaluating every time the merit of this process. One subset is used for the Decision Table part of the algorithm and the other one for the Nave Bayes part. In each repetition the selected subsets are modeled from the corresponding algorithm and in addition all the attributes are modeled with the Decision Table. Every time a forward selection search is applied. In any occasion the entirely withdrawal of an attribute of the model is considered.

2.2 Research on Asperities & Seismicity

A prominent related research direction is the identification of asperities' salient characteristics, such as the b -value and seismic density.

The work by Wiemer & Wyss [2], has shown that the b -value, i.e. the slope of the Frequency - Magnitude distribution, is significantly lower in asperities, in comparison to other fault zones which have higher b -values. Therein, the authors mapped the b -value distribution of the Parkfield segment of the San Andreas fault and identified an anomalously low b -value in the same area where the Parkfield asperity is located with the lowest b -value coinciding exactly with the Parkfield asperity while the rest of the rupture area having significantly higher b values. Accordingly, b -value information have been therein shown as useful indicators for asperity location.

In a study made at the region of Hokkaido, Japan, Takahashi & Kasahara [15] proposed a method for locating asperities by means of the earthquakes' density. Therein, asperities were identified as sections of a region with small number of events, i.e. at least one event of high magnitude being surrounded by sections with a plethora of events. To test their method, Takahashi & Kasahara mapped the density of earthquake in the area showing that known asperities of this area coincided spatially with patches of low earthquake density, surrounded by areas with higher seismicity. Accordingly, low earthquake density values have been therein shown as useful indicators for asperity location.

Accordingly, based on the findings of [2] and [15], b -value and seismic density information have been shown as useful indicators for asperity location. It is on these underlying assumptions that our work is based on in order to employ the proposed feature.

Research for asperities' location identification is rather limited while most existing related research has focused on the identification of the temporal aspect of already identified seismic entities.

As far as the former research direction, Yamanaka & Kikuchi in both [24, 25] focused on the examination of characteristic behavior of asperities, by studying the source processes of large interplate earthquakes in various locations of Japan. Therein, information collected seismograms for waveform inversion, while on the basis of derived heterogeneous fault slips, they identified asperities leading thus to an asperity map.

In the latter direction, a plethora of machine learning, data mining, and feature extraction methods have been proposed in seismicity analysis as tools for earthquakes' prediction and hazards' prevention. In [26], the authors therein used co-occurrence cluster mining to identify earthquake swarms and seismic patterns in different regions but with similar properties that were correlated with increased probability. Moreover, data mining methods have been used in [27] for forecasting the month or the year an earthquake will occur. The use of Neural networks has been proposed for earthquake prediction in numerous occasions. Panakkat et al. [28] proposed a recurrent neural network, with training and testing data from the Southern California and the San Francisco Bay, for the prediction of the time and location of seismic events. In another study, Reyes et al. [29] proposed and tested, at the wider region of Chile, a neural network that was shown to predict the probability of earthquake's magnitude being larger than a threshold value as well as the probability of an earthquake's magnitude from a limited magnitude interval.

3 Material & Methods

This Section focuses on the nature and representation of data examined herein as well as the feature vector proposed and its extraction from the raw data.

3.1 Seismic Data

The hypocenter data used in the experiments were determined by Hokkaido University, Sapporo, Japan. The data date from July 1st, 1976 until December 31st, 2002. Every earthquake in the data is a record with information about the time the earthquake occurred (year, month, day, hour, minute), the earthquake's epicenter (latitude, longitude, depth), and the earthquake's magnitude. The area tested (Hokkaido region) is located between 37° - 42.5° latitude and 140° - 145° longitude.

3.2 Data Representation

Due to the complexity of the earthquake phenomenon, there are not many features that can describe an area's seismicity thoroughly. For the presented task, the b -value and the seismic density features were selected, which are widely acceptable characteristics, among seismology researchers, of an area's seismicity.

In our previous work [1], data included the longitude and latitude attributes of the hypocenter (henceforth referred to as *bDensLatLon* dataset). In this work, the dataset employed (henceforth referred to as *bDens*) is the concatenation of the *bDensLatLon* dataset, without the geo-location attributes, together with the records on asperities found in [25] from the south of Hokkaido, Japan.

The removal of the geo-location attributes from the incorporated *bDensLatLon* dataset was necessary due to the natural spatial concentration of asperities, as previously described in Section 1.1.

Thus the *bDens* dataset includes:

- Density of earthquake instances in the corresponding area (Numeric)
- b -value of the Gutenberg-Richter frequency-magnitude distribution (Numeric)
- Asperity indicator (Binary)

where the attribute “Asperity indicator”, is a binary variable (“Yes” or “No”) indicating if an area constitutes an asperity or not, and it was used as the classification class of the vector in the experiments conducted.

3.3 Feature Vector Extraction

For the purposes of this paper the wider area of Hokkaido region was separated in a grid by 0.1 latitude and longitude degrees. In order to ensure the robustness of the estimated b -values, the radius of every cell that had at least 30 events was increased in order to contain 50 events [30], using the data of the surrounding cells. The process of creating the grid was automated by use of software, written in C programming language, that composes a separate catalog for each cell of the grid and also measures the corresponding density.

For all the sections where the number of events (density) was greater than 50, the b -value was calculated. Equation 5 describes the Gutenberg-Richter frequency-magnitude distribution (G-R FMD)

$$\text{Log}(N) = a - b * M \quad (5)$$

where N is the accumulated number of events, M is the events magnitude, a -value indicates the total seismicity rate of the region, and the b -value constitutes the slope of the distribution describing the ratio of small and big earthquakes in an earthquake catalog [31]. The most often used procedures to

calculate the b -value of a G-R FMD is the Maximum Likelihood Estimate of b -value method proposed by Utsu and presented by Aki [32] and the least square technique [33]. For the b -value estimation the Maximum likelihood method was chosen.

The calculations were made using the software ZMAP [34]. The purpose of this application is to determine the quality of seismic data, which are included in earthquake catalogs as well as to calculate and extract useful features. The application combines many basic and useful tools for seismological research.

In our feature vector, every section with density lower than 50 was marked with b -value ‘?’ corresponding to the WEKA’s [35] missing value symbol.

4 Performance Evaluation

In support of the efficiency of the proposed feature vector and the examined machine learning algorithms, this section presents a number of experiments that have been performed. A concise description of the experimentation platform and data sets is also given followed by a performance analysis.

4.1 Experimental Setup

A large number of experiments were conducted using WEKA, a platform that allows experimenting with state-of-the-art techniques in machine learning.

Due to high imbalance of examples between the two classification classes (539 No and 61 Yes) the SMOTE [36] preprocess algorithm was used. This algorithm creates synthetic examples of the minority class. To do so, it uses the k nearest neighbors of every example of the minority class. In the training set the minority class was thus oversampled by 783% in order to even the examples in both classes resulting with 485 “No” and 484 “Yes” examples.

With the two classes evenly matched experiments were conducted using all the available classifying algorithms of WEKA. In each iteration, the 10-fold cross validation technique was used. The available data were randomly portioned in 90% for training and 10% for testing. The training set was oversampled with the SMOTE technique and the whole process was repeated 10 times. In every tested algorithm the results were derived from combining the output of both 10 experiments conducted with randomly created training and test samples.

The following parameters were used:

- The RandomForest algorithm was set to generate 100 trees and every time to use all the vector’s features.
- In the DecisionTable algorithm, the internal cross-validation was made with 1 fold and the BestFirst search method was used in order to identify the best attribute combination for the decision table.

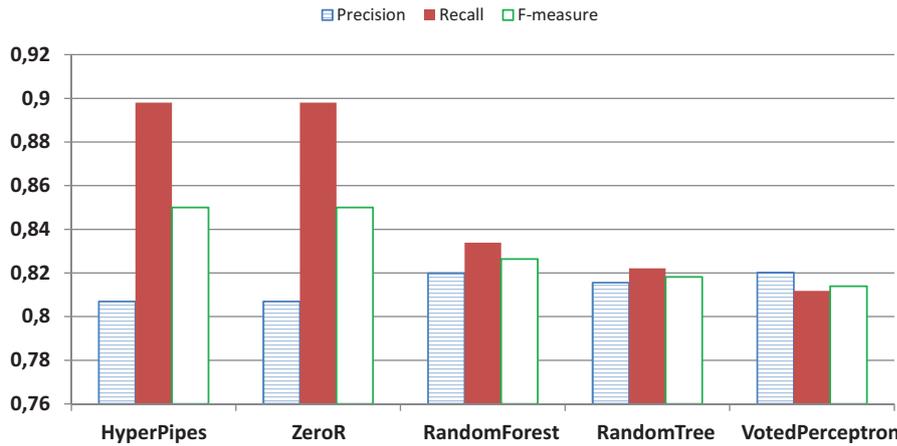


Fig. 1 Performance of the 5 most efficient classification algorithms using the same dataset and preprocessing algorithm (SMOTE) with the *bDensLatLon* dataset.

- For the DTNB classification algorithm, the internal cross-validation was made with 1 fold and the BackwardsWithDelete search method was used in order to identify the best attribute combination for the decision table.

4.2 Experimental Results

The evaluation of the algorithms results is made by means of precision and recall, that was then combined using the F-Measure [37].

The first experiment aimed at repeating the experimental process of our previous work [1], i.e. by use of the *bDensLatLon* dataset while at the same time using the vector proposed herein, that is the attributes of *b*-value and Density. The results, as shown in Figure 1 indicated a slight diminishing of the performance of the algorithms, which was expected given the fact that the original vector also included the geo-location attributed. The HyperPipes algorithm was the best performing, with F-measure 0.85. The top-5 performing algorithms' list, includes HyperPipes, ZeroR, RandomForest, RandomTree and VotedPerceptron, replacing the algorithms SimpleCart, Ridor, BFTree and NBTree, that in our previous work were in the top-5 performing algorithms.

Since in this work the dataset employed is a superset of *bDensLatLon*, as described in Section 3.2, the next experiment is a repetition of the previous experiment aiming at identifying changes in the algorithms' performance due to the newly introduced records. The results, as shown in Figure 2, indicated a marginal change in performance. In detail HyperPipes was the best performing with F-measure 0.837. Similarly, ZeroR presented a minor diminishing with F-measure 0.837, while RandomTree remained at the top-5 performing algorithms with F-measure 0.7942. The top-5 list previous entries of RandomForest and VotedPerceptron were replaced by IBk and MultilayerPerceptron.



Fig. 2 Performance of the 5 most efficient classification algorithms on a dataset that combines data from the Hokkaido region the area where Yamanaka and Kikutchi locate the asperities (SMOTE was used for resampling).

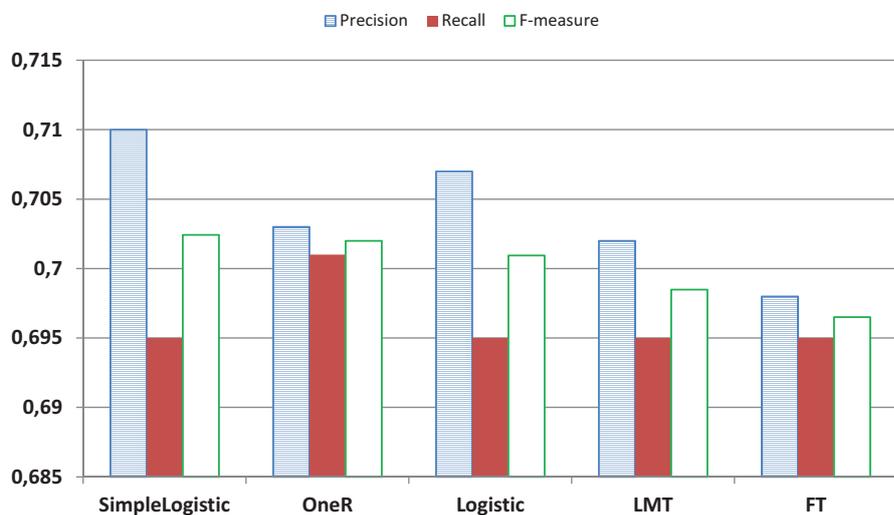


Fig. 3 Performance of the 5 most efficient classification algorithms on the combined areas dataset (random undersampling was used at preprocessing).

Then, using the full *bDens* dataset, the next experimentation was done on the algorithms' change of performance when using random undersampling in the preprocessing in contrast to the previous experiment (where SMOTE oversampling was utilised). The results, as shown in Figure 3, indicated loss of performance while RandomForest and RandomTree did not even made the top-5 performance list. The best performing algorithm was SimpleLogistic with F-measure 0.702. Only LMT managed to remain at the top-5 list with F-measure 0.698 while the rest are OneR, Logistic and FT.

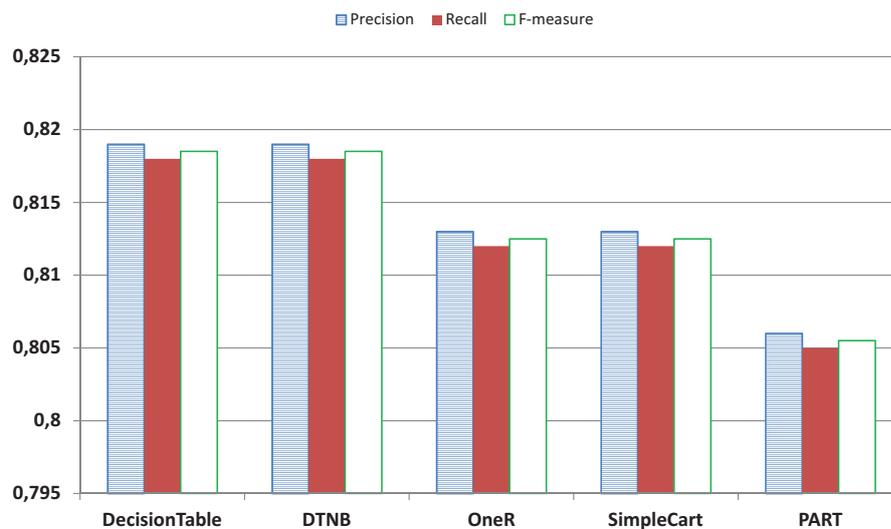


Fig. 4 Performance of the 5 most efficient classification algorithms with use of the combined area dataset (randomly undersampling used after deduction of Tomek links).

In the sequel, the previous experimentation was repeated but instead of solely random undersampling, Tomek links were initially removed from the data and then random undersampling was applied. Results shown in Figure 4 indicated ameliorated performance from the previous results, while most algorithms' performance almost reached the results of the second experiment. In detail, DecisionTable and DTNB performed best with F-measure 0.818 while the top-5 list also includes OneR, SimpleCart, and PART.

The fifth experiment included a preprocessing stem where Tomek links were removed and SMOTE was applied in order to balance the classes with increase of the records of the minority class. As a result, the HyperPipes algorithm performed best with F-measure 0.829 while the rest of the top-5 are ZeroR, IBk, SimpleCart and BFTree, as shown in Figure 5.

Comparing the results from experiments 2-5, it is evident that oversampling with the application of SMOTE (exp 2) performs better than any other preprocessing that includes random undersampling or Tomek links removal. Still, in the case of undersampling, performance was clearly better with Tomek links removal prior to the random undersampling (exp 4) in contrast to the sole application of random undersampling (exp 3).

The next experiment aims to test the effect of cost sensitive classification on the algorithms' performance, without preprocessing of the dataset. The performance of the algorithms was tested solely using the F-measure in relation to the "Yes" class. Figures 6,7,8 and 9 present the algorithms that achieved F-measure greater than zero.

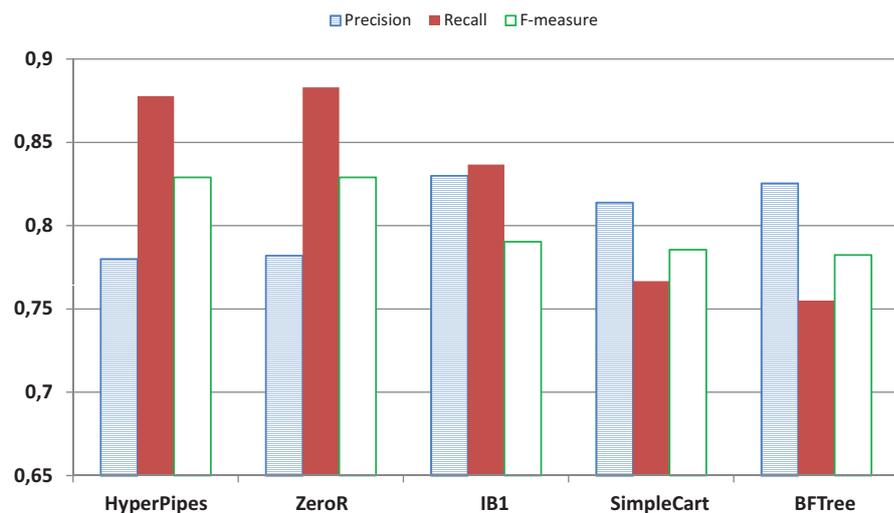


Fig. 5 Performance of the 5 most efficient classification algorithms on the combined area dataset (SMOTE oversampling was used after deduction of Tomek links).

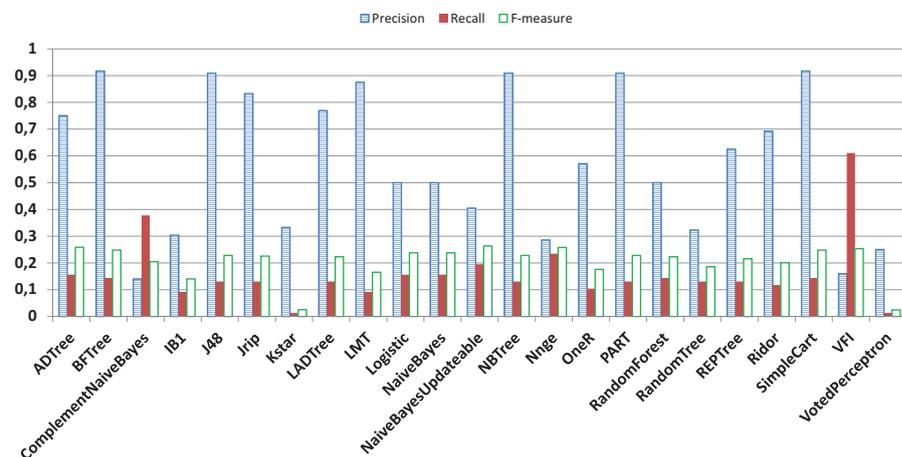


Fig. 6 Performance of all tested algorithms on classifying the Yes class (without data resampling).

In the following experiment, all algorithms were tested without cost sensitive classification in order to verify their baseline performance Figure 6. The best performing was NaiveBayesUpdateable [38] with F-measure 0.263.

Then, cost sensitive classification was applied using the cost matrix $\begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix}$. The results, as shown in Figure 7, indicated diminishing of the performance of the best algorithm NaiveBayesUpdateable into F-measure 0.237. In this case the best performing algorithm was ADTree while retaining the F-measure 0.258 achieved in the baseline experiment.

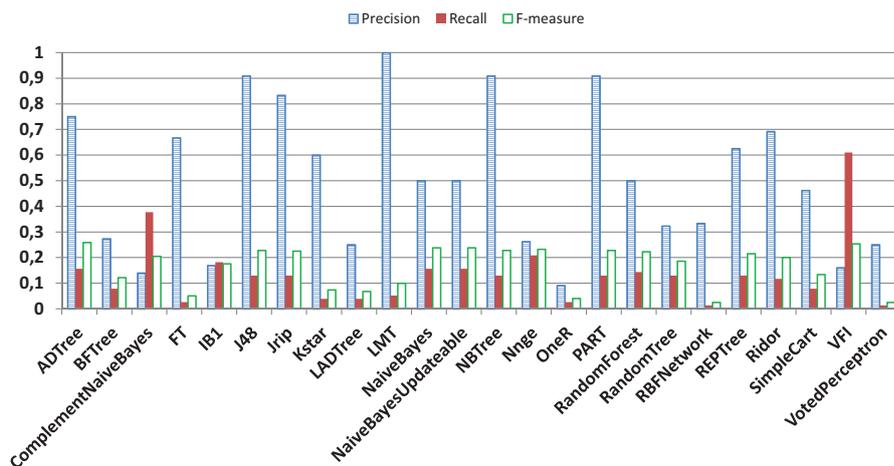


Fig. 7 Performance of all tested algorithms under cost sensitive classification with cost matrix (0,1,1,0), on classifying the Yes class (without data resampling).

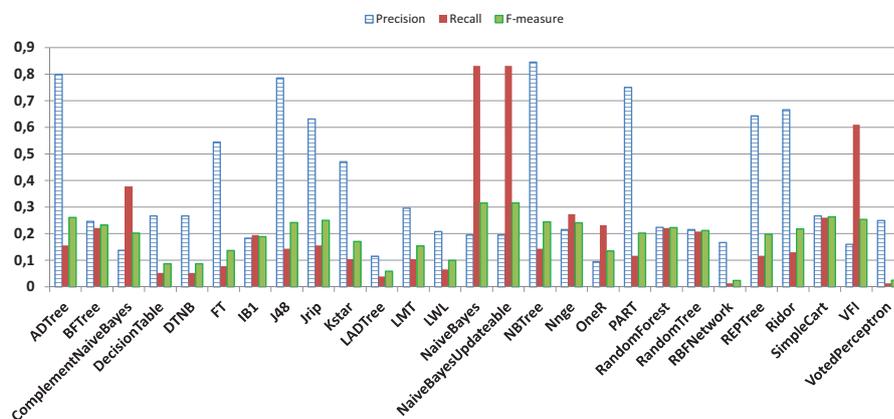


Fig. 8 Performance of all tested algorithms under cost sensitive classification with cost matrix (0,1,2,0), on classifying the Yes class (without data resampling).

In the final experiment, the cost matrix applied was $\begin{matrix} 0 & 1 \\ 2 & 0 \end{matrix}$. The results collected showed NaveBayesUpdateable and NaiveBayes achieving F-measure 0.315, which is the best increase of performance during experiments 6, 7, and 8, as shown in the collective presentation of the combined results in Figure 9.

5 Conclusions

In this work, supervised machine learning algorithms were used to identify areas with asperity properties, in the wider region of Hokkaido, Japan. To this end, the proposed feature vector consisted solely of data attributes referring to the seismic density and *b*-value of the examined region.

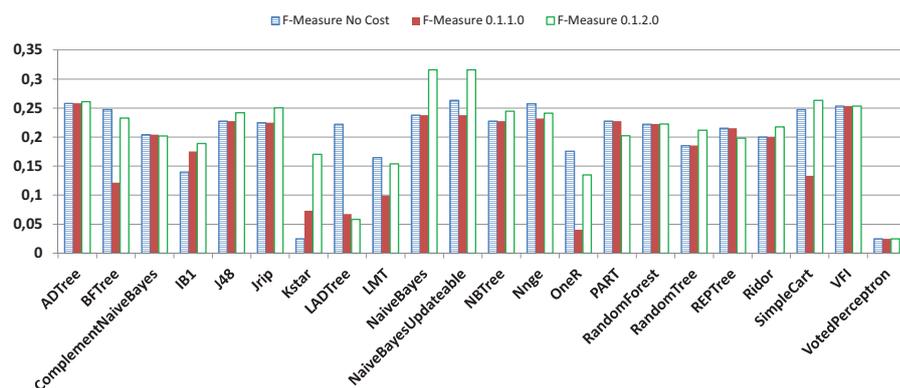


Fig. 9 Combined results of Figures 6, 7, and 8.

Extensive experimentation was conducted focusing on the effect of over-sampling and under-sampling of the data on the classification algorithms' performance. Moreover, the effect of cost sensitive classification without any re-sampling of the data was also tested as to its effect on algorithms' performance.

The results obtained for the aforementioned combination of feature vector and methodologies were promising. The performance of the utilised algorithms was retained at acceptable levels in all cases of resampling, while the best results were achieved with the SMOTE oversampling method. In the cases that no resampling was made, although the resulting performance was lower than the previous method, the use of cost sensitive classification showed promising capability of significantly increasing performance.

Future research is aimed at enlarging the feature vector by addition of more seismicity related attributes, such as the earthquake interval time and the magnitude range. Moreover, customisation and fine tuning of the classification algorithms' parameters for the domain's characteristics is also expected to boost results.

References

1. Kostantinos Arvanitakis and Markos Avlonitis. Identifying asperity patterns via machine learning algorithms. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pages 87–93. Springer, 2016.
2. Stefan Wiemer and Max Wyss. Mapping the frequency-magnitude distribution in asperities: An improved technique to calculate recurrence times? *Journal of Geophysical Research*, 102:115–128, 1997.
3. Markos Avlonitis and GA Papadopoulos. Foreshocks and b value: bridging macroscopic observations to source mechanical considerations. *Pure and Applied Geophysics*, 171(10):2569–2580, 2014.
4. M Avlonitis and K Kalaitzidou. Estimating the real contact area between sliding surfaces by means of a modified ofc model. *Archives of Civil and Mechanical Engineering*, 15(2):355–360, 2015.
5. D Hatzfeld, D Kementzetzidou, V Karakostas, M Ziazia, S Nothard, D Diagourtas, A Deschamps, G Karakaisis, P Papadimitriou, M Scordilis, R Smith, N Voulgaris, S Kiratzi, K Makropoulos, M P Bouin, and P Bernard. The galaxidi earthquake of 18

- november 1992: A possible asperity within the normal fault system of the gulf of Corinth (Greece). *Bulletin of the Seismological Society of America*, 86(6):1987–1991, 1996.
6. Sun-Cheon Park and Jim Mori. Are asperity patterns persistent? Implication from large earthquakes in Papua New Guinea. *Journal of Geophysical Research: Solid Earth*, 112(B3), 2007.
 7. L A Dalguer, K Irikura, and J D Riera. Simulation of tensile crack generation by three-dimensional dynamic shear rupture propagation during an earthquake. *Journal of Geophysical Research: Solid Earth*, 108(B3), 2003.
 8. Nelson Pulido. Broadband frequency asperity parameters of crustal earthquakes from inversion of near-fault ground motion. In *Proceedings of the 13th World Conference on Earthquake Engineering, Vancouver, BC, Canada. Paper*, number 751, 2004.
 9. Kojiro Irikura, Hiroe Miyake, Tomotaka Iwata, Katsuhiko Kamae, Hidenori Kawabe, and Luis Angel Dalguer. Recipe for predicting strong ground motion from future large earthquake. In *Proceedings of the 13th World Conference on Earthquake Engineering*, volume 1341. Citeseer, 2004.
 10. A Arda Ozacar and Susan L Beck. The 2002 Denali fault and 2001 Kunlun fault earthquakes: complex rupture processes of two large strike-slip events. *Bulletin of the Seismological Society of America*, 94(6B):S278–S292, 2004.
 11. Takao Kagawa, Kojiro Irikura, and Paul G Somerville. Differences in ground motion and fault rupture process between the surface and buried rupture earthquakes. *Earth, Planets and Space*, 56(1):3–14, 2004.
 12. Satoko Murotani, Kenji Satake, and Yushiro Fujii. Scaling relations of seismic moment, rupture area, average slip, and asperity size for $m \sim 9$ subduction-zone earthquakes. *Geophysical Research Letters*, 40(19):5070–5074, 2013.
 13. William Spence, C Mendoza, ER Engdahl, GL Choy, and Edmundo Norabuena. Seismic subduction of the Nazca ridge as shown by the 1996–97 Peru earthquakes. In *Seismogenic and Tsunamigenic Processes in Shallow Subduction Zones*, pages 753–776. Springer, 1999.
 14. Nelson Pulido, Shin Aoi, and Hiroyuki Fujiwara. Rupture process of the 2007 Notohanto earthquake by using an isochrones back-projection method and K-net/KIK-net data. *Earth, Planets and Space*, 60(10):1035–1040, 2008.
 15. Hiroaki Takahashi and Minoru Kasahara. Spatial relationship between interseismic seismicity, coseismic asperities and aftershock activity in the southwestern Kuril Islands. *Volcanism and Subduction: The Kamchatka Region*, pages 153–164, 2007.
 16. Ivan Tomek. Two modifications of CNN. *IEEE Trans. Systems, Man and Cybernetics*, 6:769–772, 1976.
 17. Peter Hart. The condensed nearest neighbor rule (corresp.). *IEEE Transactions on Information Theory*, 14(3):515–516, 1968.
 18. Zainab Abu Deeb, Thomas Devine, and Zongyu Geng. Randomized decimation hyperpipes. *Citeseer*, 2010.
 19. C Lakshmi Devasena, T Sumathi, VV Gomathi, and M Hemalatha. Effectiveness evaluation of rule based classifiers for the classification of iris data set. *Bonfring International Journal of Man Machine Interface*, 1:5, 2011.
 20. Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
 21. Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *European conference on computational learning theory*, pages 23–37. Springer, 1995.
 22. Ron Kohavi. The power of decision tables. *Machine Learning: ECML-95*, pages 174–189, 1995.
 23. Mark A Hall and Eibe Frank. Combining naive Bayes and decision tables. In *FLAIRS conference*, volume 2118, pages 318–319, 2008.
 24. Yoshiko Yamanaka and Masayuki Kikuchi. Asperity map based on the analysis of historical seismograms: Tohoku version. *Jpn. Earth Planet. Sci. Joint Meet. (Sy-005)*, 2001.
 25. Yoshiko Yamanaka and Masayuki Kikuchi. Asperity map along the subduction zone in northeastern Japan inferred from regional seismic data. *Journal of Geophysical Research: Solid Earth*, 109(B7), 2004.

26. Ken-ichi Fukui, Daiki Inaba, and Masayuki Numao. Discovering seismic interactions after the 2011 tohoku earthquake by co-occurring cluster mining. *Information and Media Technologies*, 9(4):886–895, 2014.
27. GV Otari and RV Kulkarni. A review of application of data mining in earthquake prediction. *International Journal of Computer Science and Information Technologies*, 3(2):3570–3574, 2012.
28. Ashif Panakkat and Hojjat Adeli. Neural network models for earthquake magnitude prediction using multiple seismicity indicators. *International Journal of Neural Systems*, 17(01):13–33, 2007.
29. Jorge Reyes, A Morales-Esteban, and Francisco Martínez-Álvarez. Neural networks to predict earthquakes in chile. *Applied Soft Computing*, 13(2):1314–1328, 2013.
30. Karen R Felzer. Calculating the gutenbergrichter b value. In *AGU Fall Meeting Abstracts*, 2006.
31. Ota Kulhanek. Seminar on b-value. *Dept. of Geophysics, Charles University, Prague*, pages 10–190, 2005.
32. Keiiti Aki. Maximum likelihood estimate of b in the formula $\log n = a - bm$ and its confidence limits. *Bulletin of the Earthquake research Institute*, 43:237–239, 1965.
33. Beno Gutenberg and Charles F Richter. Frequency of earthquakes in california. *Bulletin of the Seismological Society of America*, 34(4):185–188, 1944.
34. Stefan Wiemer. A software package to analyze seismicity: Zmap. *Seismological Research Letters*, 72(3):373–382, 2001.
35. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.
36. Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.
37. C. J. Van Rijsbergen. *Information Retrieval*. Butterworth, 1979.
38. George H John and Pat Langley. Estimating continuous distributions in bayesian classifiers. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pages 338–345. Morgan Kaufmann Publishers Inc., 1995.