ISSN 2277-2685
IJESR/Sep. 2023/ Vol-13/Issue-3/1-12
Mrs.G.Sumalatha *et. al.*, / International Journal of Engineering & Science Research

Disaster Intensity-Based Selection of Training Samples for Remote Sensing Building Damage Classification

Mrs.G.Sumalatha[1], Kondannagari Sai Prakash[2], Samala Pravalika[2], A Boris [2], Kemmasaram pranay [21assoc] *Professor,Computer Science and Engineering, CMR Engineering College, medchal, T.S, India*[2]*B.Tech, Computer Science and Engineering, CMR Engineering College, medchal, T.S, India*

Abstract: The detection of damaged structures in the wake of a significant catastrophe has been successfully accomplished in the past using machine learning in remote sensing. Standard approaches, however, do not take into account the difficulty and expense of assembling a training data set after a major catastrophe. In this article, we examine catastrophes whose magnitude may be predicted by numerical modelling and/or instrumentation. Two entirely automated approaches are proposed for the identification of seriously damaged structures in such circumstances. The underlying premise is that samples from low-disaster severity locations mostly reflect undamaged houses. Furthermore, both damaged and undamaged buildings are probably present in areas with moderate to strong disaster intensities. The standard support vector machine classifier is learned and calibrated using a method that is based on the automated selection of training data. The second method relies on the definition of the support vectors using two regularisation parameters. These frameworks bypass the time-consuming process of collecting labelled building samples via field surveys and/or visual examination of optical images. Three actual cases—the earthquake and tsunami that struck Tohoku-Oki in 2011, the earthquake and tsunami that struck Kumamoto in 2016, and the floods that hit Okayama in 2018—are used to assess the performance of the suggested strategy. The accuracy of the outcome varies from 0.85 to 0.89, demonstrating that it may be utilised for the quick allocation of damaged buildings.

1. INTRODUCTION

For the purpose of extracting information from remote sensing data, MACHINE LEARNING has emerged as the dominant data processing paradigm. The basic approach is to develop a model using sparse but adequately encoded past information (i.e., training samples) to categorise an instance under investigation (such as a building) with a theme label (in the application context of this article, a damage condition). These techniques are particularly helpful when explicit modelling based on, for instance, mechanical models, is too difficult. In order to attain high predicted accuracy, such systems need both a substantial quantity of past

information and workable descriptors to describe the occurrences under investigation. However, the lack of an adequate supply of training samples affects a lot of applications. Obtaining training samples may sometimes become quite time- and money-consuming. Different strategies have been put forth to address the scarcity of training samples in this situation. Examples include the use of multiobjective sparse representation classifiers [1], the creation of virtual samples [2], and active learning [4] techniques [3][5] in earlier works. Other methods focus on creating a training set entirely automatically from predetermined input data. Such examples include the identification of samples of forest areas using top-of-atmosphere reflectance [6] and the fusion of multisource geodata [7].

A Description Of The Project: The following highlights our contributions.

1) The demand parameter enables narrowing the scope of the change search to just those locations with medium to high demand characteristics.

2) To gather samples that haven't changed, we place a threshold on the demand parameter. The choice of the demand threshold is fairly obvious and doesn't call for any prior processing, unlike unsupervised classification methods, since the demand parameter has a clear physical meaning.

3) The collected nonchanged samples provide a stronger representation of the class nonchanged in the feature space because the demand parameter information is independent of remote sensing data.

4) We combine data from on-site sensors (such as tidal gauges and ground motion sensors), numerical modelling of a natural occurrence, and remote sensing.

2. LITERATURE SURVEY

2.1 Existing System

Previous attempts to use machine learning in remote sensing to identify damaged structures after a significant catastrophe have proved effective. The difficulty and expense of assembling a training data set after a major catastrophe, however, are not taken into account by traditional approaches. The current system analyses catastrophic occurrences whose severity may be predicted using numerical simulation and/or instrumentation. Two totally automated processes for the identification of seriously damaged structures are provided to handle such situations.

The basic premise is that samples from places with low catastrophe severity mostly reflect undamaged structures. Additionally, both damaged and undamaged structures are likely to be found in areas with moderate to strong disaster intensities. In light of this supposition, a method based on the automated selection of training samples for training and calibrating the common SVM classifier is used. In the second method, the support vectors are defined using two regularisation parameters. These frameworks do not rely on time-consuming field research or visual examination of optical images to acquire labelled building sample data. Utilising the 2011 Tohoku-Oki earthquake and tsunami, the 2016 Kumamoto earthquake, and the 2018 Okayama floods as real-world examples, the performance of the suggested method is assessed. Since the accuracy of the result is in the range of 0.85 to 0.89, it can be used to quickly allocate affected buildings.

2.2Proposed System

1) The demand parameter enables narrowing the scope of the change search to just those locations with medium to high demand characteristics.

2) To capture non-changed samples, the system applies a threshold to the demand parameter. The choice of the demand threshold is fairly obvious and doesn't call for any prior processing, unlike unsupervised classification methods, since the demand parameter has a clear physical meaning.

3) The collected non changed samples provide a better representation of the class non changed in the feature space because the demand parameter information is independent of the remote sensing data.

4) The system combines data from on-site sensors (such as tidal gauges and ground motion sensors), numerical modelling of a natural occurrence, and remote sensing.

2.2 DSS (Distance-Based Sample Selection) Method:

The components of a subset B1 B1 that may be utilised to train new samples are found via an iterative process. After that, a common classifier (in this case, an SVM) for binary classification may be developed using training data from both classes. The patterns linked to the change class in the feature space should be present in B1, and they shouldn't clash with the patterns linked to B1's set. In light of this, we suggest that B1 be identified using the

coordinates of its elements in the feature space in relation to B1. Therefore, it is necessary to define B1 in a functional form. A closed border that encompasses B1 may be defined using a number of different techniques [13] In this work, one-class SVM, also known as the support vector of high-dimensional distribution [14], is used.

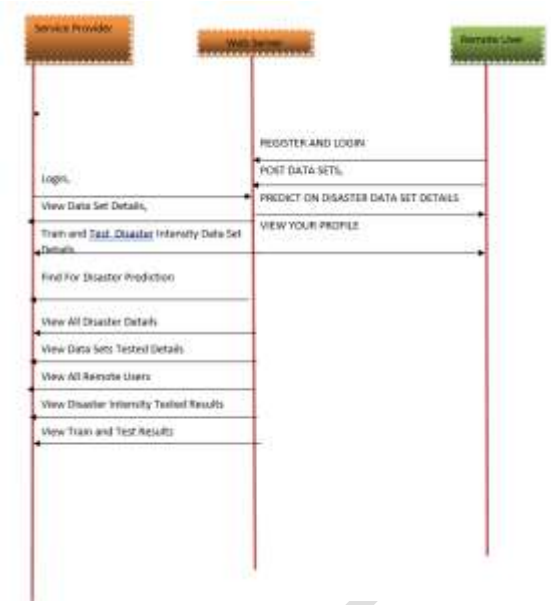

Fig 1: System Flow

3. SYSTEM ANALYSIS&DESIGN

3.1 Approach using Multiple Regularisation Parameters (MRP):

The second method for calibrating the discriminant function g(x) is based on a change to (5) and (6) in the handling of outlier data. The word "soft margin" in the method's name refers to the second term of equation (5), which is also referred to as a regularisation term. It loosens the restriction that the two classes are separated by the hyperplane, g(x) = 0.

When a support vector xi does not fall inside the margin's boundaries, the slack variable, i, measures the degree of disagreement (i.e., g(x) = 1 for set B1 and g(x) = 1 for set B1).

If a sample is not a support vector, then = 0. The training data might be misclassified thanks to the slack variables. The classifier's performance depends heavily on the

regularisation term, > 0. A low value produces a flexible discriminant function that accepts a large number of outliers in the training set. With a high value of, the discriminant function will be stiff and have few outliers. The choice of is a compromise between the discriminant function's ability to generalise and the training data's precision. You may find suggestions for choosing it in [15].

3.2 The 2011 Tohoku earthquake-tsunami is the first case study (CS1).

The Mw 9.0 Tohoku earthquake on March 11, 2011, was one of the most powerful earthquakes ever recorded. The biggest coseismic deformation detected by the Global Navigation Satellite System (GNSS) was 5 m, while the maximum strong-motion acceleration measured was 2.7 g. The tsunami that the earthquake brought forth devastated Tohoku's coastline region severely[15-19]. The highest tsunami height ever recorded was 40 metres. For the coast of Miyagi Prefecture, the Ministry of Land, Infrastructure, Transport, and Tourism (MLIT) conducted field surveys and provided a building damage inventory, as shown in Fig. 6(a). The report specified seven different categories of harm, from no damage to washed away. The houses that are situated inside the floodplains along the coast of Miyagi Prefecture that were the subject of TerraSAR-X pictures are the main focus of this research.

Demand Parameter: The tsunami was mostly responsible for the damage. The demand parameter in this case is the depth of the tsunami inundation. The MLIT gave the actual inundation depth, which is employed in this empirical assessment [see Fig. 6(b)].

Feature Space Two microwave images taken by TerraSAR-X of the coastline region of Miyagi Prefecture were used to build the feature space [see Fig. 6(c) and (d)]. Before and after the earthquake-tsunami, on October 12, 2010, and March 13, 2011, respectively, the photographs were taken. The synthetic aperture radar (SAR) pictures underwent coregistration, radiometric calibration, speckle filtering, and terrain correction as preprocessing. 1.25 m is the photos' resolution. The averaged difference in backscattering

intensity, xi1, and the correlation coefficient between the pictures, xi2, are two characteristics that are calculated from the imaging. At the site of each building that the MLIT assessed, the features were calculated. The building footprint's rectangle box's pixels were used to calculate both characteristics [see Fig. 6(a)]. The data collection consists of 31 262 samples in total. The feature vectors were standardised as part of the preprocessing such that, for all j > 1, 2 >, (1/M)

I x I j = 0 and (1/M)

i x 2 i j = 1. 3) Results: D = 0.15 m served as the definition for the subsets. 498 structures from the MLIT's inventory had flood depths of less than or equal to 0.15 m, while 30 764 buildings had inundation depths more than 0.15 m. Consequently, #B1 = #B1 = 498. The sets B1 and B1 are shown in the bidimensional feature space in Fig. 7. The predicted outcomes are shown in Fig. 8, where the non-changed and altered samples are denoted by the colours blue and red, respectively. Using v = 0.1 and the RBF kernel function, k(xi, x j) = exp(xi x j 2), with = 0.1, the one-class SVM classifications were calculated from B1. It is important to note that samples with the highest correlation coefficient, xi1, are incorrectly categorised as modified samples.

The degree of change between the two photographs is measured by the correlation coefficient, a hand-engineered characteristic.

Prior to standardisation, one, which is related to nonchange samples, had the biggest value. However, a correlation coefficient of one is uncommon in real-world applications.

Such samples appear to be outliers, and the one-class SVM may mistakenly classify them as damaged.

As a result, using a classifier that is calibrated just using B1 is not ideal.

3.3 System Architecture:Demand Parameter: The PGV is represented by the demand parameter in this case study. For earthquake events, additional demand parameters, such as

the PGA and seismic intensity, may be used. However, field observations have demonstrated a strong correlation between the PGV and damages to wood-framed buildings in Japan, where wood is the most popular building material. QuiQuake, a platform that offers numerous demand parameter maps in almost real-time gives the PGVs at the sample locations [see Fig. 10(c)]. The K-NET and KiK-net networks' strong motion recordings are used by QuiQuake to construct the PGA, PGV, and JMA seismic intensity scales. Together, the K-NET and KiK-net have around 2000 accelerometers, and they evenly cover the whole island of Japan. The method used to calculate the demand parameter maps is shown in Fig. 11. The demand parameter is first calculated by QuiQuake at the position of each accelerometer. Next, the demand parameter values at the bedrock level are predicted using soil amplification factors. The demand parameter values are calculated about every 250 m inside a grid format using an interpolation approach. The interpolation takes into account the demand parameter's attenuation laws in relation to its proximity to the earthquake source. To prevent the impacts of the underlying layers on the soil, the interpolation is done at the bedrock level.

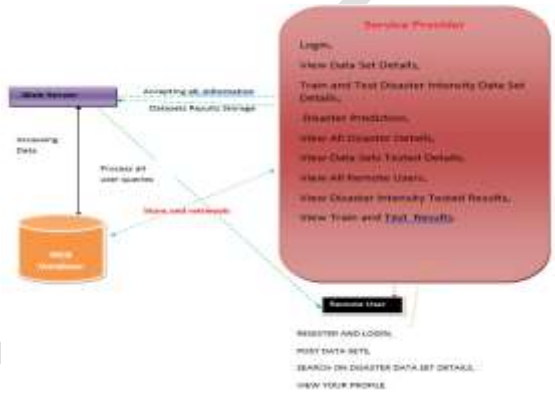

Fig 2 : System Architecture

3.4 Data Flow Diagram : Whenever a new system is developed, user training is required to educate them about the working of the system so that it can be put to efficient use by those for whom the system has been primarily designed. For this purpose the normal working of the project was demonstrated to the prospective users. Its working is easily understandable and since the expected users are people who have good knowledge of computers, the use of this system is very easy.

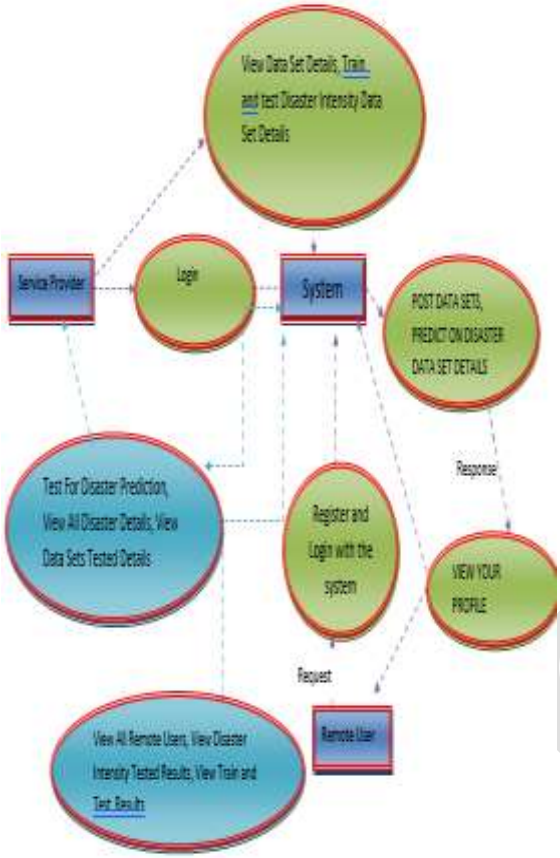

Fig 3: Data Flow Diagram

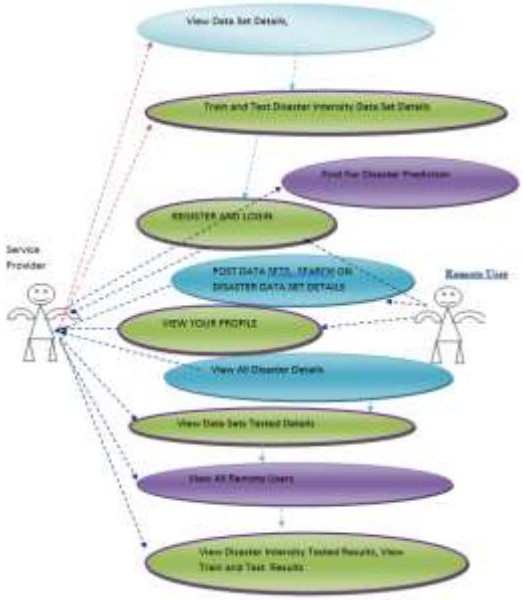

Fig 4 Use Case UML Diagrams

5. CONCLUSION

The demand parameter, which assesses catastrophe severity, is used for the first time in this work to systematically extract samples from remote sensing pictures and utilise them to calibrate a change detection classifier. Instrumentation and/or numerical simulation, which may be performed in real-time or very close to real-time, are used to estimate the demand parameter for each sample. We suggest using a demand parameter map to divide the samples into two subsets, one of which is made up of samples with low demand for the geo locations and the other of which is made up of samples with medium and/or high demand. We assumed that the first group was mostly made up of samples that had not been altered, and that the second subset was made up of both altered and unaltered samples. Two techniques are given for calibrating a discriminant function under these restrictions. The first approach consists of two key phases. The discriminant function is first calibrated using the one-class support vector machine (SVM) on the subset with low demand. Second, the second subset is used to enhance the discriminant function. A soft margin SVM with two regularisation parameters is used in the second approach. The discriminant function of the SVM with two regularisation terms can have different levels of tolerance for the subsets than the standard SVM, which only uses one regularisation parameter. In particular, the discriminant function will accept few outliers from the subset made up of samples with low demand while being highly flexible and accepting many outliers from the subset made up of samples with high demand.

Three disasters—the 2011 Tohoku earthquake tsunami, the 2016 Kumamoto earthquake, and the 2018 western Japan floods—were used to assess the suggested approaches. Each case study's feature space was also built using a variety of remote sensing data sources. In the first case study, backscattering intensities from microwave imaging were utilised, in the second, Lidar-based DSMs, and in the third, backscattering complex values from microwave imagery. The outcomes were roughly at the same level of accuracy as those that were reported in earlier investigations using conventional machine learning techniques. But unlike the other studies, our methods can be applied in almost real-time.

The usual method for gathering training samples after a major catastrophe is what slows down the development of a machine-learning-based damage map. An unsolved issue in the use of machine learning for rapid disaster response is the automated extraction of training

samples. Our research is pertinent because it helps find answers to problems from which the magnitude of a catastrophe might be inferred.

6. FUTURE SCOPE

We suggested a multilayered deep convolutional neural network for natural disaster identification and intensity classification to overcome these issues. The suggested technique consists of two blocks: the first block is used to identify natural disasters, and the second block is used to address concerns with unequal class representation. The suggested model performs noticeably better at identifying and categorising natural disasters, but it can be used to a variety of natural disaster detection procedures in the future.

7. REFERENCES

[1] B. Pan, Z. Shi, and X. Xu, "Multiobjective-based sparse representation classifier for hyperspectral imagery using limited samples," IEEE Trans. Geosci. Remote Sens., vol. 57, no. 1, pp. 239–249, Jan. 2019.

[2] C. Geiß, P. A. Pelizari, L. Blickensdörfer, and H. Taubenböck, "Virtual support vector machines with self-learning strategy for classification of multispectral remote sensing imagery," ISPRS J. Photogramm. Remote Sens., vol. 151, pp. 42–58, May 2019.

[3] C. Geiß, M. Thoma, M. Pittore, M. Wieland, S. W. Dech, and H. Taubenbock, "Multitask active learning for characterization of built environments with multisensor Earth observation data," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 10, no. 12, pp. 5583–5597, Dec. 2017.

[4] C. Persello, A. Boularias, M. Dalponte, T. Gobakken, E. Næsset, and B. Schölkopf, "Cost-sensitive active learning with lookahead: Optimizing field surveys for remote sensing data classification," IEEE Trans. Geosci. Remote Sens., vol. 52, no. 10, pp. 6652–6664, Oct. 2014.

[5] C. Geib, M. Thoma, and H. Taubenbock, "Cost-sensitive multitask active learning for characterization of urban environments with remote sensing," IEEE Geosci. Remote Sens. Lett., vol. 15, no. 6, pp. 922–926, Jun. 2018.

[6] C. Huang et al., "Use of a dark object concept and support vector machines to automate forest cover change analysis," Remote Sens. Environ., vol. 112, no. 3, pp. 970–985, Mar. 2008.

[7] C. Geiß, P. A. Pelizari, S. Bauer, A. Schmitt, and H. Taubenböck, "Automatic training set compilation with multisource geodata for DTM generation from the TanDEM-X DSM," IEEE Geosci. Remote Sens. Lett., vol. 17, no. 3, pp. 456–460, Mar. 2019.

[8] E. Booth, K. Saito, R. Spence, G. Madabhushi, and R. T. Eguchi, "Validating assessments of seismic damage made from remote sensing," Earthq. Spectra, vol. 27, no. 1, pp. 157–177, Oct. 2011.

[9] A. Cooner, Y. Shao, and J. Campbell, "Detection of urban damage using remote sensing and machine learning algorithms: Revisiting the 2010 haiti earthquake," Remote Sens., vol. 8, no. 10, p. 868, Oct. 2016.

[10] M. Janalipour and A. Mohammadzadeh, "A fuzzy-GA based decision making system for detecting damaged buildings from high-spatial resolution optical images," Remote Sens., vol. 9, no. 4, p. 349, Apr. 2017.

[11] H. Gokon et al., "A method for detecting buildings destroyed by the 2011 Tohoku earthquake and tsunami using multitemporal TerraSAR-X data," IEEE Geosci. Remote Sens. Lett., vol. 12, no. 6, pp. 1277–1281, Jun. 2015.

[12] M. Wieland, W. Liu, and F. Yamazaki, "Learning change from synthetic aperture radar images: Performance evaluation of a support vector machine to detect earthquake and tsunami-induced changes," Remote Sens., vol. 8, no. 10, p. 792, Sep. 2016.

[13] Y. Bai et al., "A framework of rapid regional tsunami damage recognition from post-event TerraSAR-X imagery using deep neural networks," IEEE Geosci. Remote Sens. Lett., vol. 15, no. 1, pp. 43–47, Jan. 2018.

[14] L. Moya, H. Zakeri, F. Yamazaki, W. Liu, E. Mas, and S. Koshimura, "3D gray level co-occurrence matrix and its application to identifying collapsed buildings," ISPRS J. Photogramm. Remote Sens., vol. 149, pp. 14–28, Mar. 2019.

[15]G.sumalatha, Software Defect Prediction Survey Introducing Innovations with Multiple Techniques. In: Kumar, A., Mozar, S., Haase, J. (eds) Advances in Cognitive Science and Communications. ICCCE 2023. Cognitive Science and Technology. Springer,Singapore. https://doi.org/10.1007/978-981-19-8086-2_76

16. Suriya Begum, Farooq Ahmed Siddique, **Rajesh Tiwari**, "A Study for Predicting Heart Disease using Machine Learning", Turkish Journal of Computer and Mathematics Education, Vol. 12, Issue 10, 2021, pp 4584-4592, e-ISSN: 1309-4653

17. **Rajesh Tiwari**, Manisha Sharma and Kamal K. Mehta, "Performance Analysis of various High-Performance Computing Models" , Journal of Advanced Research in Dynamic and Control System, Vol. 10, special issue 2, July 2018, pp 2192 – 2200 , ISSN: 1943 – 023X

18. Dr.Sheo Kumar, **Dr. Rajesh Tiwari** and NeethuChoudhary, "Detection of Phishing Attacks in Web Environment using Unsupervised Machine Learning" National Conference on Computational Methods, Data Science and Applications (NC-CMDSA 2021) held at Maulana Azad National Urdu University, Hyderabad on 24th – 25th May 2021.

19. Mrutyunjaya S Yalawar, K Vijaya Babu, Bairy Mahender, Hareran Singh, A Brain-Inspired Cognitive Control Framework for Artificial Intelligence Dynamic System,2022/4/29,International Conference on Communications and Cyber Physical Engineering 2018,735-745,Springer Nature Singapore. https://doi.org/10.1007/978-981-19-8086-2_70