Learn to Detect: Improving the Accuracy of Earthquake Detection

Tai-Lin Chin, *Member, IEEE*, Chin-Ya Huang, *Member, IEEE*, Shan-Hsiang Shen, *Member, IEEE*, You-Cheng Tsai, Yu Hen Hu, *Fellow, IEEE*, and Yih-Min Wu

***Abstract*— Earthquake early warning system uses high-speed computer network to transmit earthquake information to population center ahead of the arrival of destructive earthquake waves. This short (10 s of seconds) lead time will allow emergency responses such as turning off gas pipeline valves to be activated to mitigate potential disaster and casualties. However, the excessive false alarm rate of such a system imposes heavy cost in terms of loss of services, undue panics, and diminishing credibility of such a warning system. At the current, the decision algorithm to issue an early warning of the onset of an earthquake is often based on empirically chosen features and heuristically set thresholds and suffers from excessive false alarm rate. In this paper, we experimented with three advanced machine learning algorithms, namely, *K*-nearest neighbor (KNN), classification tree, and support vector machine (SVM) and compared their performance against a traditional criterion-based method. Using the seismic data collected by an experimental strong motion detection network in Taiwan for these experiments, we observed that the machine learning algorithms exhibit higher detection accuracy with much reduced false alarm rate.**

***Index Terms*— Detection accuracy, earthquake detection, machine learning.**

I. Introduction

IN THE recent years, earthquakes occurred much more frequently around the circum-Pacific seismic belt and usually caused severe casualties. In 2011, the 311 Tōhoku earthquake in Japan caused 15 891 deaths and severe damages for the Fukushima Daiichi nuclear power plant. In 1999, more than 2415 deaths and 11 305 severely wounded have been confirmed in the 921 Chi-Chi earthquake in Taiwan. Since earthquake has become a serious threat to human life and property, a significant research effort has been devoted in developing earthquake early warning (EEW) systems in order to prevent severe casualties [1]–[3]. In particular, Taiwan is located on the border of the Eurasian plate and the Philippine Sea plate. The special geological characteristics around the island cause more than 100 felt earthquakes each year on or near the island. The high population density on the island generates a desperate demand for reliable real-time earthquake detection.

Accuracy is one of the most important issues for EEW systems since false alarms may generate unnecessary panic and cause significant economic loss. Unfortunately, sensor readings are usually corrupted by noise. It is very challenging to automatically identify the occurrence of earthquakes in real time without human inspection. Traditional anomaly detection schemes in signal processing are usually based on certain statistical models [4], [5]. However, seismic signals may not follow those theoretical models since earthquakes can occur in locations with different geological conditions. Simple detection schemes could easily misread certain vibrations as earthquakes. Conventional earthquake detection schemes usually monitor certain criteria to determine the occurrence of an event. For example, if the ratio of the short-term average (STA) over the long-term average (LTA) of the shaking acceleration is greater than a preselected threshold, it is regarded as the presence of an event [6]–[8]. In order to reduce false alarms, several other criteria are applied to determine whether the event is a true earthquake. Thresholds for those criteria need to be carefully chosen. However, false alarms could still occur from time to time since the selection of thresholds highly depends on human experiences. Recently, some work develops the detection models using neural networks [9], [10]. Nevertheless, most of the work uses the whole waveform of an event to determine the presence of the earthquake. Those schemes are better for postevent processing rather than real-time detection. Our work aims to determine the presence of the earthquakes within a very short period at the beginning of an event and the criteria for determining an earthquake are learned by machines automatically.

Machine learning is a methodology to classify the monitored objects into different categories according to the features learned from the training data. Through a thorough collection of data from various types of the monitored events, certain invisible properties can be revealed by the learning process. Proper criteria for event classification are formed automatically and the accuracy of correctly identifying the target events can be improved.

Manuscript received January 22, 2018; revised November 11, 2018, March 20, 2019 and May 2, 2019; accepted May 19, 2019. Date of publication July 23, 2019; date of current version October 31, 2019. This work was supported in part by the Ministry of Science and Technology of Taiwan under Grant MOST-107-2221-E-011-129. *(Corresponding author: Tai-Lin Chin.)*

T.-L. Chin, S.-H. Shen, and Y.-C. Tsai are with the Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei 106, Taiwan (e-mail: tchin@mail.ntust.edu.tw; sshen@mail.ntust.edu.tw).

C.-Y. Huang is with the Department of Electrical and Computer Engineering, National Taiwan University of Science and Technology, Taipei 106, Taiwan (e-mail: chinya@gapps.ntust.edu.tw).

Y. H. Hu is with the Department of the Electrical and Computer Engineering, University of Wisconsin–Madison, Madison, WI 53706 USA (e-mail: yhhu@wisc.edu).

Y.-M. Wu is with the Department of Geosciences, National Taiwan University, Taipei 10617, Taiwan (e-mail: drymwu@ntu.edu.tw).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2019.2923453

In this paper, three advanced machine learning schemes, namely, K-nearest neighbor (KNN), classification tree, and support vector machine (SVM), were exploited for earthquake detection. The learning models were built as a binary classifier to identify the presence or absence of an earthquake. To collect the seismic wave data, an experimental high-density strong motion detection network has been built in Taiwan with more than 600 sensors deployed over the main island and several small islands around Taiwan. The relatively low cost of the sensors, which are developed by microelectromechanical System (MEMS) technologies, makes a high-density deployment possible. The motion accelerations in the x-, y-, and z-axes were measured by each sensor. Earthquake events from January 2016 to December 2017 were collected to conduct the simulations. Features extracted from the waveforms were used to train the learning-based schemes. The detection results were compared to those from the traditional criterion-based method. From the comparisons, the accuracy and reliability of the learning-based detector were significantly improved by the learning-based schemes.

This paper is organized as follows. Section II introduces the high-density seismic network that was used to collect the seismic wave data. Section III presents the proposed learning-based detection schemes. Section IV compares the learning-based schemes and the traditional criterion-based method. Section V briefly addresses the related work. Finally, the conclusions of this paper are presented in Section VI.

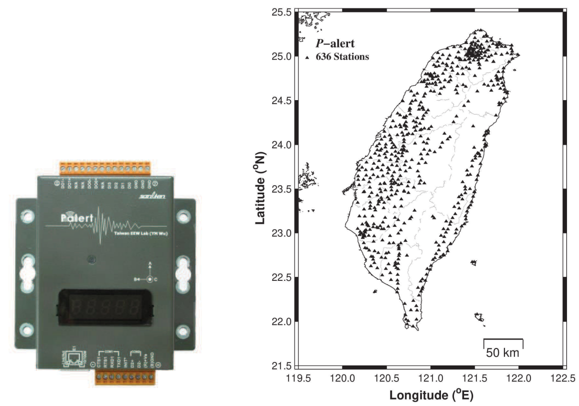
II. Data Collection

This section first describes the seismic network deployed for earthquake data collection in Taiwan. The events and data used to train the learning-based schemes are then presented.

A. Seismic Network

Traditional strong motion monitoring systems usually contain several sensors deployed in a large geographic area far between each other. One sensor has to cover a relatively large area resulting in the low spatial resolution of the collected data. In contrast, an experimental high-density seismic wave collection network using low-cost sensors has been deployed in Taiwan [1]. The sensors, named Palert as shown in Fig. 1(a), are developed by MEMS Technology. The high-density deployment makes it possible to better cover the monitored region of interest. In 2008, there are 636 Palert stations deployed on the island of Taiwan and a couple of nearby islands as shown in Fig. 1(b). All of them are connected through the Internet and have the capability of sampling the motion accelerations in the x-, y-, and z-axes with 100-Hz sampling rate. The acceleration values are transmitted to a main server in Taipei.

Fig. 2 shows the seismic waves collected by the network for the Meinong earthquake in Taiwan on February 6, 2016. The Richter magnitude of the earthquake is 6.6. The earthquake caused very serious damages in the south of Taiwan including more than 100 deaths and many buildings collapse. Fig. 2 shows the vertical acceleration of the seismic waves collected by the Palert stations shown in Fig. 3. The locations and the



(a) (b)

Fig. 1. Palert and the Deployment of the network. (a) Palert. (b) Deployment of the Palert stations.

distance of the Palert stations to the earthquake epicenter are listed in Table I. Note that the sensor density in the network is very high and the magnitude of the earthquake is also large. Thus, almost all of the Palert stations have recorded the seismic wave. Due to the lack of space, only the waves recorded by several selected stations are shown in Fig. 2. The Palert network collects seismic wave data in real time and can be used to implement many applications such as earthquake detection and early warning.

B. Events and Data

This paper used the time series data, i.e., the motion acceleration, collected by the Palert network to conduct the training of the learning-based schemes. The events included in our data set are those announced by the Central Weather Bureau (CWB) in Taiwan from January 2016 to December 2017. The CWB classifies the events into two categories, namely, the numbered earthquakes and the regional earthquakes. The following rules are applied for the numbered events.

The local magnitude of the earthquake is greater than 4.0 and at least one of the following condition holds.

1) When the event occurs, at least one sensor station reports the event with intensity greater than four or at least two stations report the event with intensity greater than three.
2) When the event occurs, at least one sensor station at any county capital reports the event with intensity greater than three or at least two stations at any county capital report the event with intensity greater than two.
3) When the event occurs, at least one sensor station at Taipei or Kaoshiung reports the event with intensity greater than two.

The following rules are used to classify the regional events.

1) Any sensor stations report the event with the intensity of four or above.
2) The Richter magnitude of the earthquake is greater than 3.5 and, at least one sensor station reports the event with intensity greater than three or at least two stations report the event with intensity greater than two.

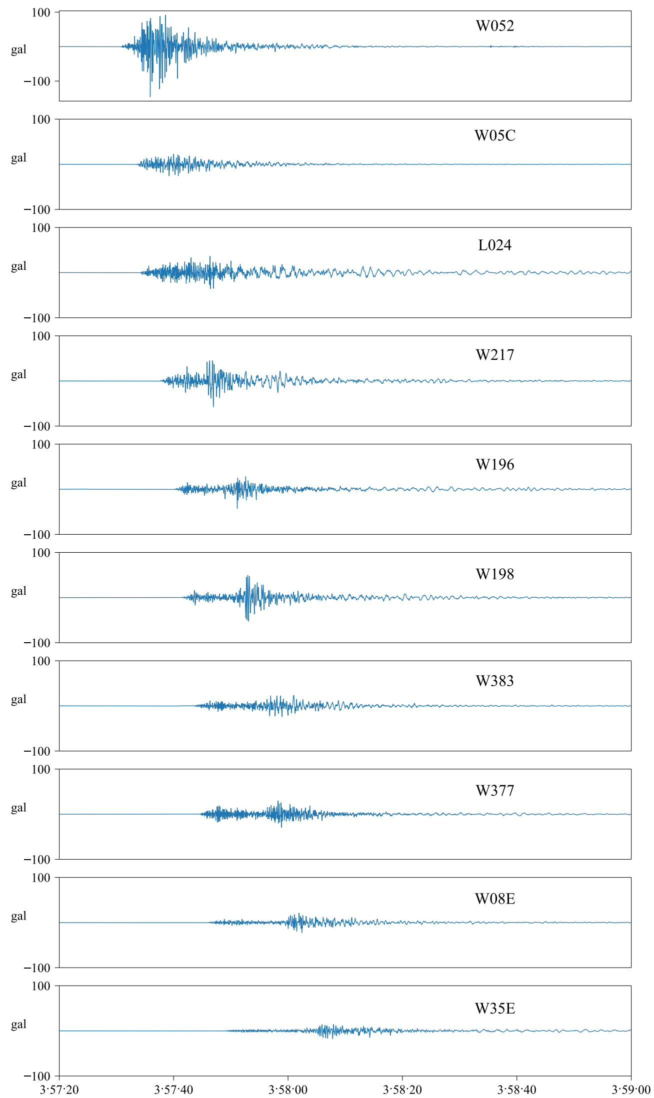


Fig. 2. Waveforms of the selected Palerts for the 2016 Meinong earthquake.

3) Any events that not fulfill the previous two criteria but have been reported as an event that affects the general public.

Note that the above-mentioned rules are used by the CWB in Taiwan to classify the events and the sensor stations used in the rules are the stations in CWB's network not the Palert stations in our network.

For the numbered events, 453 waveforms recorded by the Palert stations from 65 earthquakes with a magnitude between 2.5 and 6.9 were collected in our data set. Fig. 4 shows the epicenters and the magnitude of the CWB-numbered events. The circles represent the magnitudes of the corresponding events. For the regional events, there were 133 waveforms from 63 earthquakes in the data set. Fig. 5 shows the epicenters and the magnitude of the regional events. The magnitude is between 2 and 5. From the figures, one can observe that the CWB-numbered events usually have larger magnitude and most of the regional events are relatively small in terms of magnitude. In addition, most of the events occurred in the east or south of Taiwan. Table II and III list the statistics of the magnitude and depth of the events in the two categories.

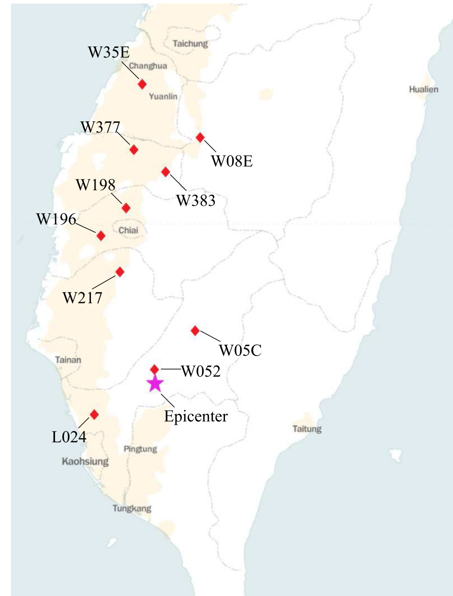


Fig. 3. Selected Palerts for the 2016 Meinong earthquake.

TABLE I
DETAILS OF THE PALERT STATIONS IN FIG. 2

Meinong earthquake with magnitude 6.6 : 02/06/2016 3:57:26.1

Name	Longitude	Latitude	Detection time	Distance(km)
W052	120.540	22.969	57:29.92	5.45
W05C	120.701	23.109	57:32.51	26.73
L024	120.302	22.803	57:33.30	27.66
W217	120.405	23.323	57:36.67	46.94
W196	120.330	23.454	57:38.67	63.20
W198	120.429	23.555	57:40.55	71.59
W383	120.586	23.686	57:42.63	85.40
W377	120.458	23.767	57:43.45	94.65
W08E	120.721	23.809	57:45.12	100.67
W35E	120.492	24.002	57:47.80	120.54

Nonearthquake events were also collected in our data set. This kind of events could be caused by situations such as a vehicle passing by the vicinity of a sensor or other noises generated by the instruments. The purpose to include nonearthquake events is to make the model learn the differences between the earthquake and nonearthquake waveforms. Total 600 waveforms from randomly selected nonearthquake events were included in the data set.

III. EARTHQUAKE DETECTION

In this section, the proposed earthquake detection process is presented. Assume that the acceleration of seismic waves is sampled and collected by seismic sensors. The detection process is divided into two phases. In the first phase, the real-time series of acceleration collected by a sensor is investigated by a fast event screening process to determine the presence of a potential event. In the second phase, the candidate events are further verified by a binary classifier based

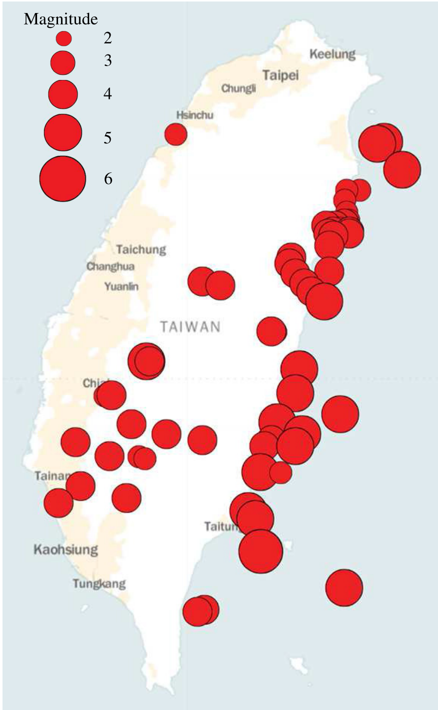


Fig. 4. CWB-numbered events in the data set.

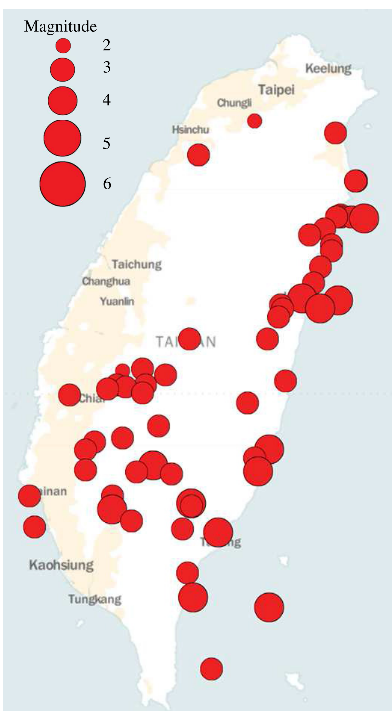


Fig. 5. Regional events in the data set.

on machine learning techniques to check whether the event is a true earthquake or not.

A. Event Screening

In practice, two types of seismic waves are generated by an earthquake, namely, P-wave and S-wave, i.e., primary wave or pressure wave and secondary wave or shear wave, respectively. P-wave is a type of longitudinal wave. In other words, the particles in the medium vibrate along the axis of the propagation. P-waves travel faster than S-waves and, hence, can reach the sensors earlier. Therefore, P-wave is usually used to detect the occurrence of earthquakes in conventional early warning systems. In general, seismic sensors can record the acceleration of seismic waves in x-, y-, and z-axes and the vibration caused by P-wave is usually in the vertical direction. Consequently, in this paper, the time series in the z-axis was used to detect the presence of earthquakes. Since the sampling rate of the sensors could be pretty high, a fast screening scheme is first used to determine potential target events. The screening process should be fast enough to prevent processing lag. False alarms can be tolerated in the screening process since the candidate events will be further verified later in the second phase. In contrast, the missing rate should be as low as possible in the screening phase.

TABLE II

MAGNITUDE OF THE COLLECTED EVENTS

Richter magnitude	CWB-numbered events	Regional events
1 - 2	0	0
2 - 3	0	5
3 - 4	15	49
4 - 5	31	9
5 - 6	17	0
6 - 7	2	0

TABLE III

DEPTH OF THE COLLECTED EVENTS

Depth (m)	CWB-numbered events	Regional events
0 - 10	11	26
10 - 20	29	27
20 - 30	14	7
30 - 40	5	1
40 -	6	2

A simple way to identify an anomaly event in the first phase is to check the fluctuation of the amplitude of the seismic wave. Given a seismic wave, let A_i denote the amplitude of the seismic wave at sample i and D_i denote the difference of the amplitude between sample i and sample $i-1$, i.e., $D_i = A_i - A_{i-1}$. A characteristic function is defined as

$$C_i = A_i^2 + D_i^2. \quad (1)$$

Note that the difference in consecutive samples can be used to indicate the condition of the signal fluctuation and the characteristic function amplifies the variation in the signal amplitude. However, since samples are always corrupted by noise, the characteristic value at a particular sample may not reflect the true condition of the monitored environment. To mitigate the influence of the noise, the STA, and the LTA of the characteristic values are exploited. To evaluate the average values in real time, STA and LTA are implemented based on the running average defined in (2) and (3), respectively,

$$S_i = S_{i-1} + W_s \times (C_i - S_{i-1}) \quad (2)$$

$$L_i = L_{i-1} + W_l \times (C_i - L_{i-1}). \quad (3)$$

The two formulas actually have the same form but different weighting parameters, where W_s should be greater than W_l.

The two weighting parameters control the impact of the current signal on the running averages. If the weight is high, the current signal dominates in the running average, and thus, the result can represent the signal condition in the recent interval. In contrast, if the weight is low, the current status does not affect much of the running average and the result can better characterize the long-term condition of the monitored signal. The initial values, i.e., S_0 and L_0, can be set to be a small value or zero, since, if no events are present, STA and LTA should be similar and the observed signal should be close to the average noise level. To detect a potential event, the ratio of STA over LTA is checked as follows:

$$r_i = \frac{S_i}{L_i} \gtrless \eta. \tag{4}$$

If the ratio is greater than the threshold η, it is recognized as a potential event and the second phase will be triggered to verify whether it is a true earthquake. Otherwise, the screening process will continue to check the next sample. Note that in the first phase, it is allowed to choose a lower threshold so that true events would not be missed. Although the false alarm rate could be high in the first phase, the following verification in the second phase can reduce the false alarm rate by using the learning schemes. In this work, the weights W_s and W_l are 0.6 and 0.015, respectively, and the threshold η is 0.04, which are used in the current Palert network.

B. Feature Extraction

In the second phase, learning-based schemes are used as a binary classifier to verify whether an earthquake is present. Basically, the learning-based schemes have to analyze the features of the collected training data and build rules to characterize those features for the targeted categories. The built rules then form the model for classifying future observations. Theoretically, the precision of the classification can be increased if sufficient information about the features of the data in different categories is provided.

In this work, seven features selected from the collected seismic waves were used to build the classification models. The selected features are defined as follows and illustrated in Fig. 6. Note that the picking sample in a seismic wave is defined as the sample where the STA/LTA ratio is greater than the threshold η in (4) and the second verification phase is triggered.

F1: The first peak amplitude after the picking sample.

F2: The difference between the picking sample and the next sample. It can indicate the condition of fluctuation of the signal. Specifically, the feature is defined as follows:

$$F2 = |A_{i+1} - A_i| \tag{5}$$

where A_i is the amplitude of the picking sample.

F3: The running average of the absolute amplitude of the acceleration at the picking sample. Specifically

$$\bar{A}_i = (1 - W_m) \times \bar{A}_{i-1} + W_m \times |A_i| \tag{6}$$

where W_m is the weight of the current sample in the running average. The initial value for A_0 can be chosen by using any noise sample values of the acceleration. In fact, A_0 can be chosen randomly and the running average will converge to the average noise level in a short period of time if no events occur. In this work, the weight W_m is set to be 0.99, which is used in the current Palert network.

F4: The mean of the absolute amplitude of the acceleration in the two-second window following the picking sample, that is,

$$F4 = \frac{1}{I} \sum_{j=i+1}^{i+I} |A_j| \tag{7}$$

where i is the index of the picking sample and I is the number of samples within two seconds.

F5: The peak amplitude of the acceleration in the two-second window following the picking sample

$$F5 = \max_{i \le j \le i+I} |A_j| \tag{8}$$

where i is the index of the picking sample and I is the number of samples within two seconds.

F6: The peak velocity in the two-second window after the picking sample

$$F6 = \max_{i \le j \le i+I} |V_j| \tag{9}$$

where V_j is the velocity at time j and i is the index of the picking sample. Note that V_j can be calculated by the integral of the acceleration.

F7: The reference level for event checking

$$F7 = L_i \times \eta \tag{10}$$

where L_i is the LTA of the characteristic value at the picking sample. Note that from (4), $F7$ is the reference level for the STA to check for abnormal events. If STA is too small, the event may not be an earthquake even (4) is satisfied.

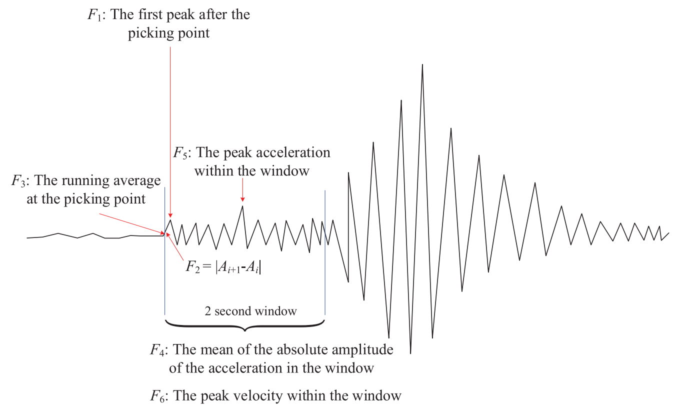


Fig. 6. Feature extraction.

The feature values are also normalized as follows:

$$Z_i = (F_i - \mu_i)/\sigma_i \tag{11}$$

where μ_i and σ_i are the mean and standard deviation of feature F_i, respectively. Note that the data set includes the feature values of both the earthquake and the nonearthquake events.

The training data set is the collection of the feature records of the collected time series.

C. Event Verification

The potential events, which pass the screening in the first phase are verified by the learning-based schemes. Three machine learning schemes were exploited to verify the events, namely, KNN, classification tree, and SVM. In this section, the learning-based schemes are introduced and the experiments are conducted to determine proper parameters for the learning-based schemes when performing the detection.

1) Evaluation Metrics: Before diving into the learning schemes, the metrics used to evaluate the detection performance are first described. Precision and recall are the typical metrics for evaluating the classification performance. For the earthquake detection, precision is the fraction of the reported events that are true earthquake events. Recall is the fraction of the true earthquake events that are reported. Let TP, FP, and FN denote the number of true positive events, false positive events, and false negative events, respectively. True positive events are those events that are, indeed, earthquakes and the detector also reports the events as earthquakes. False positive events are those that the detector reports the events as earthquakes but the events are actually not earthquakes, i.e., the false alarms. False negative events are those that are true earthquakes but the detector does not report, i.e., the missing events. Specifically, precision and recall are defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{12}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{13}$$

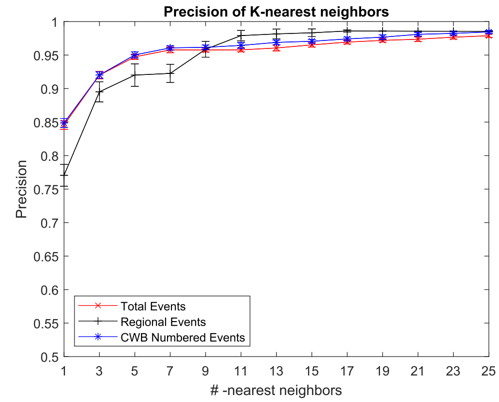
Note that 1-Precision is the false alarm rate and 1-Recall is the missing rate. The above two metrics characterize different aspects of the detection performance. To provide an integrated evaluation, the well-known F-score [11] is used. Specifically, F-score is defined as

$$F_s = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{14}$$

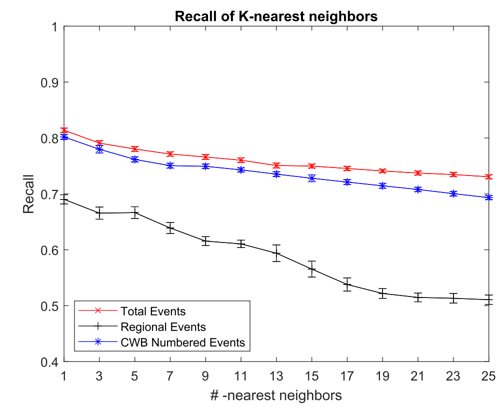
F-score is the harmonic mean of precision and recall. It is a statistical measurement of the accuracy for a binary classifier and usually used to quantitatively characterize the quality of the classifier in the machine learning area. A higher F-score indicates better classification quality. These metrics are used as the metrics to select appropriate parameter values for the learning-based schemes in the following experiments.

2) K-Nearest Neighbors: KNN is one of the fundamental schemes to use the knowledge from historical statistics [12]. In the scheme, the distance between a new observation and each feature record in the training data set is calculated. The top K records with the shortest distance are selected to do a majority vote. The major category of the selected feature records determines the category for the observation. In general, K is an odd number and the parameter K depends on the data collected for a particular application. A simple distance metric is the Euclidean distance, which is also used in this work.

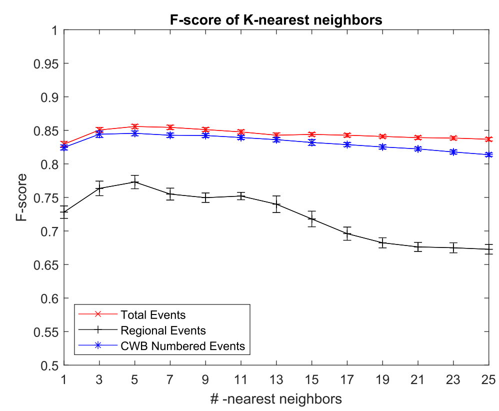
To select the parameter K, three training data sets are built to conduct the experiments. The CWB-numbered event data



(a)



(b)



(c)

Fig. 7. Performance of the KNN scheme. (a) Precision. (b) Recall. (c) F-score.

set contains the feature records of the CWB-numbered events and the nonearthquake events. The regional event data set contains the feature records of the regional events and the nonearthquake events. The total event data set contains all the feature records from the CWB-numbered events, regional events, and the nonearthquake events. The results of the KNN scheme are shown in Fig. 7. Each data point shown in the figure is an average of 50 runs of ten-way cross validation. In the ten-way cross validation, it randomly selects nine-tenth of the data records as the training data and the other one-tenth as the testing data. The results are evaluated when performing the classification on the testing data.

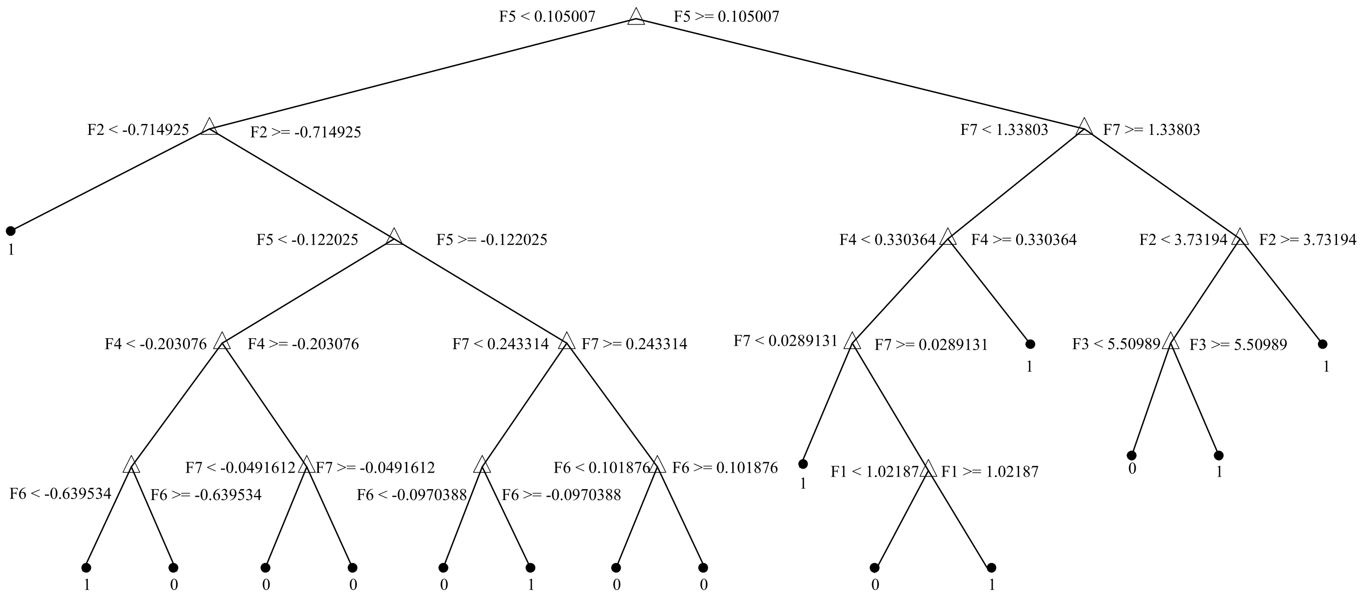


Fig. 8. Example of the built tree model.

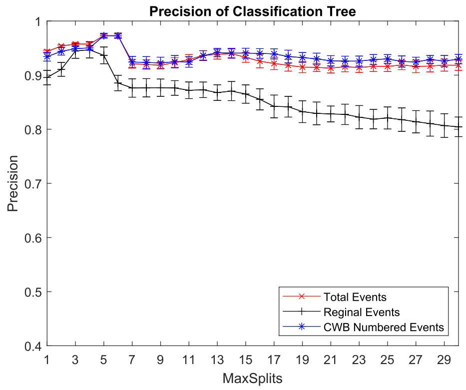
The experiment results show that, in general, F-score decreases as K increases. The error bar at each data point is the interval of one standard deviation. From the figures, it is better to use a small K for the classification. In this work, K is chosen to be five. In addition, the model built by the CWB-numbered event data set has a better performance than that built by the regional event data set. It is because most of the events in the regional event data set have relatively small magnitude and signal-to-noise ratio. Consequently, it could be more difficult to determine the presence of the earthquakes. The total event data set has the best performance since the variety of the events in the training data set provides more information about the features of the earthquakes. Therefore, the classification quality can be improved.

3) Classification Tree: Classification tree is also one of the general models to make a prediction for classification [13]. The training data are used to build a tree, which characterizes the decision rules. Each leaf of the tree represents a particular class for the final prediction. The branch nodes of the tree are the feature cut-points, which lead the decision to the potential class for the input event. Specifically, each branch node contains a rule based on a particular feature to decide the path to the leaf. The tree is usually built by a recursive method from the root that contains all the training data. At each branch node, a feature is selected for the node to best split the data in the node into two subsets and generate two child nodes. The metric for the splitting usually characterizes the homogeneity of the data after the splitting. The process is executed recursively on each child node until all the data in a node belongs to the same class or some predefined rules are met; for instance, the size of the tree.

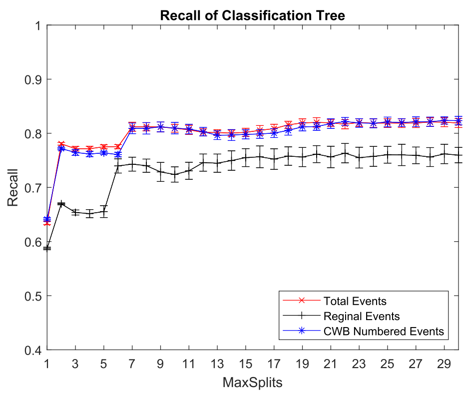
In this scheme, the tree size has a significant impact on the decision quality. If the tree size is too small, the accuracy of the decision may be low. If the tree size is too large, the model may overfit the training data. Let N be the maximum number of splitting the branch nodes when building the classification tree. Experiments are conducted to determine an appropriate value for N. Fig. 8 shows an example of the built tree model for the total data set. Note that the built tree model might be slightly different in each iteration of the ten-way cross validation since it randomly selects nine-tenth of the collected feature records as the training data. In the tree, $F5$ and $F2$ are used twice at different branch nodes, which are close to the root. One can expect that the two features can have higher impact on the detection performance. Fig. 9 shows the impact of N on the detection performance. From the results, the detection quality increases as N grows large. If N is too small, the classification may not be accurate. However, if N is set to be large, the homogeneity level of the leaf nodes may be reached before the number of splitting reaches N. Thus, the detection performance would not change too much when N increases. From the results, the models built by the total event data set and the CWB-numbered event data set have similar performance, and both are better than that built by the regional event data set. In general, the performance has a high correlation to the signal-to-noise ratio of the data in the corresponding training data sets. Finally, in this work, N is chosen to be 15 since the detection performance does not change much if N is greater than 15, but the training time is expected to be longer.

4) Support Vector Machine: SVM is basically formulated as a quadratic optimization problem [14]. The scheme determines a hyperplane with the maximum margin in a high-dimensional feature space, and the training data are classified into two groups by the hyperplane. A reasonable choice of the hyperplane is the one which the distance, or margin, to the nearest data point in each side of the hyperplane is maximized. Since the division of the training data may not be linear, a kernel function is often used to map the data points to a higher dimensional space where the classification can be easily performed.

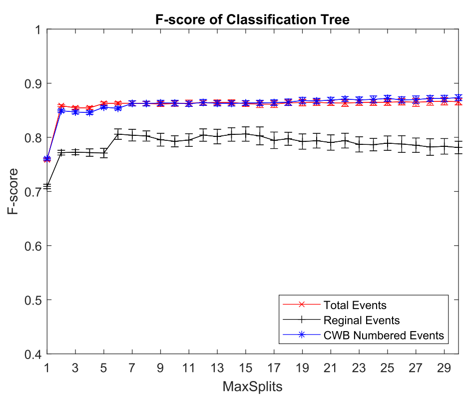
In SVM, the box constraint controls the penalty imposed by the margin-violating data items when calculating the classification hyperplane. Small box constraint allows larger margin



(a)

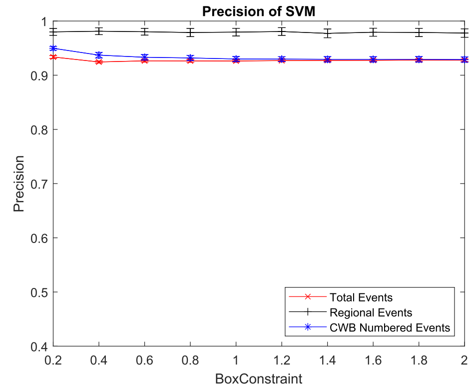


(b)

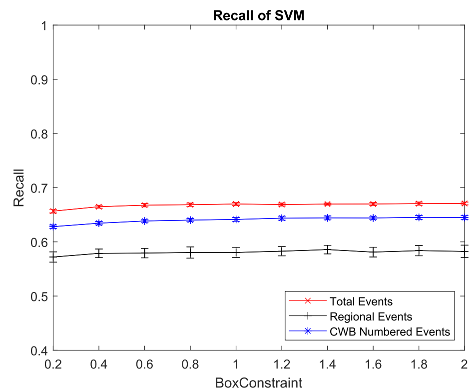


(c)

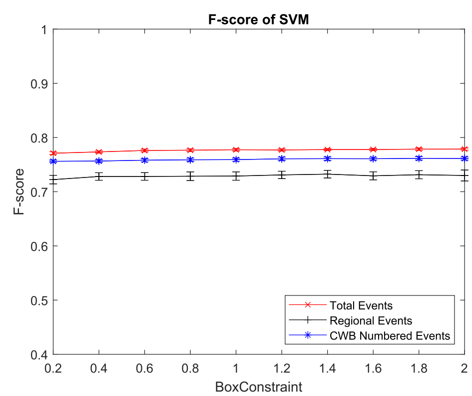
Fig. 9. Performance of the classification tree scheme. (a) Precision. (b) Recall. (c) F-score.



(a)



(b)



(c)

Fig. 10. Performance of the SVM scheme. (a) Precision. (b) Recall. (c) F-score.

violation but spends less training time. In contrast, large box constraint could generate more tight classification hyperplane but leads to longer training time.

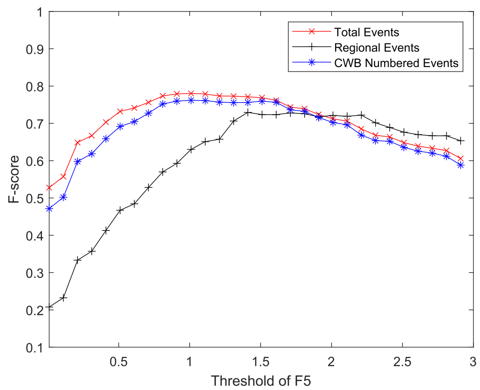
Experiments were also conducted to select an appropriate box constraint for the SVM scheme. From the results in Fig. 10, the detection quality does not change much after box constraint larger than one. For the sake of training time, small box constraint is preferred. In this work, box constraint is set to one.

IV. Experiments

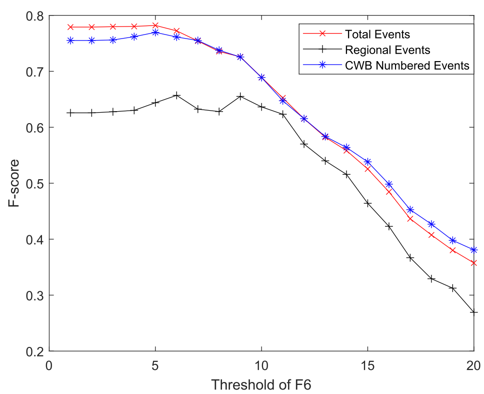
In this section, the detection quality of the learning-based scheme is compared to the traditional criterion-based detection method. An integrated scheme that integrates the three learning-based schemes by majority vote was also investigated. The impact of each selected feature on the performance of the learning-based schemes was further evaluated. Experiments were conducted to compare the performance in terms of the detection quality and the time for training and predicting.

A. Criterion-Based Method

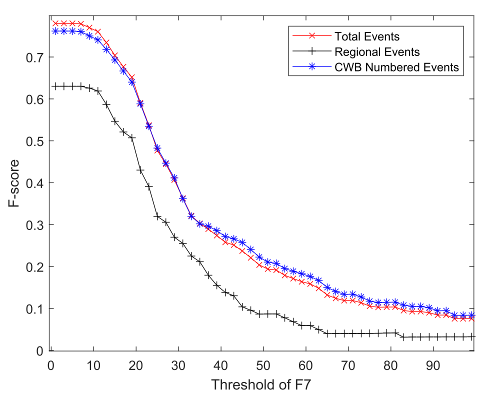
The criterion-based method is the traditional scheme used to detect earthquakes. It is conducted as follows. Once a sample passes the first phase screening, the features, $F5$, $F6$, and $F7$, are used to check the events by thresholds. Experiments to select the thresholds for the features are shown in Fig. 11. In the experiments, the default thresholds of $F5$, $F6$, and $F7$



(a)

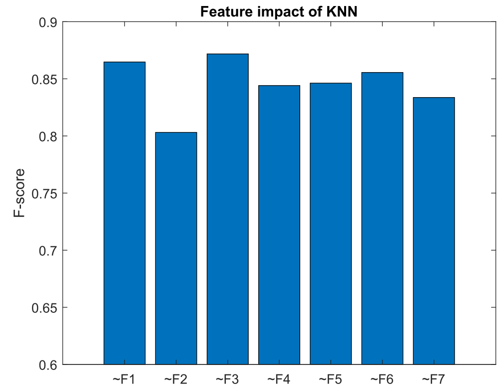


(b)

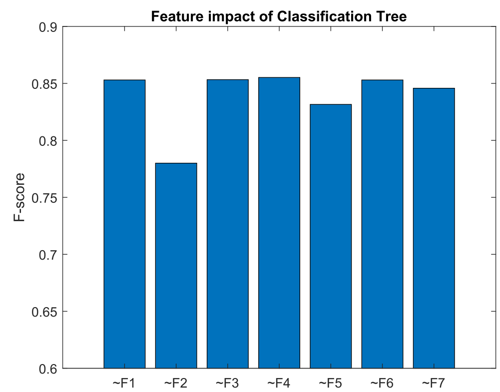


(c)

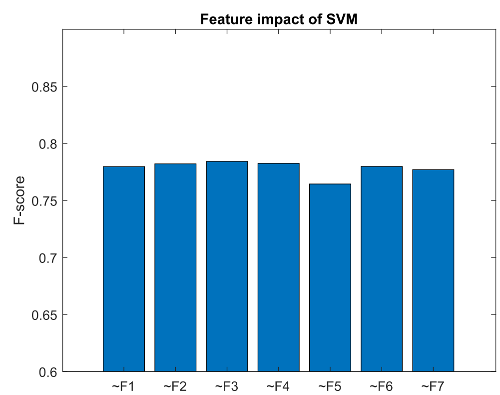
Fig. 11. Thresholds selection for the criterion-based method. (a) F5. (b) F6. (c) F7.



(a)



(b)



(c)

Fig. 12. Feature impact on the learning-based schemes. (a) KNN. (b) Classification tree. (c) SVM.

are set to be 1.1, 4, and 0, respectively, where the F-score is relatively high. The default values are also the final decisions for the associated feature. At each experiment, one of the thresholds is inspected while the other parameters remain the same. After the first phase screening is passed, if the feature values are greater than the thresholds, the criterion-based scheme issues an alarm of the earthquake event. $F7$ adds a further constraint on the STA of the characteristic value. Recall that $L_i \times \eta$ is the reference level for STA from the inequality in (4). Therefore, not only the ratio of STA/LTA should pass a certain level but also the STA should be high enough if it is recognized as an earthquake.

1) Feature Impact: The impacts of the seven features on the learning-based schemes are also studied. In each experiment, one of the features is removed from the training data set and the performance of the detector based on the other six features is evaluated by F-score. The experiments are conducted for all of the three learning-based schemes and the results are generated by ten-way cross validation. Fig. 12 shows the results for feature impact. The label on the x-axis is the removed feature in the corresponding experiment. For example, if the label is $\tilde{}F1$, then the result is generated by the model trained by $F2$ to $F7$. One can observe that $F2$, which is the difference in the acceleration amplitude between the picking sample and the next sample, has the most impact on

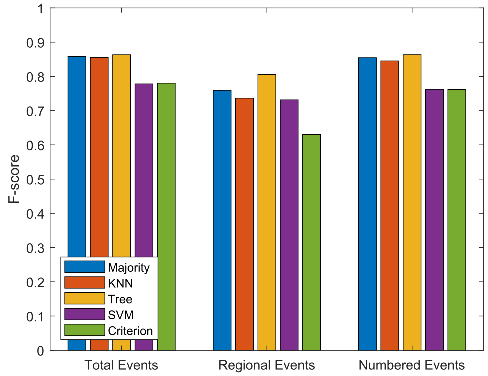


Fig. 13. Performance comparisons.

KNN and classification tree. For classification tree, $F2$ and $F5$ have the most impact, which approves the observation from the tree structure. In addition, it is obvious that $F2$ could highly affect the detection since $F2$ may have substantial change when the seismic wave arrives at the sensor. However, SVM is highly affected by $F5$, which is the peak acceleration amplitude in the two-second window following the picking sample. From the results, one can conclude that $F2$ and $F5$ have the most impact on the detection. In contrast, $F3$, which is the running average of the absolute acceleration at the picking sample, has the least impact.

2) Detection Quality: The detection performance of the learning-based schemes is compared to that of the criterion-based method. In addition, an integrated scheme that integrates the three learning-based schemes by majority vote is built. Again, the results are the average over 50 runs of ten-way cross validation. Fig. 13 shows the comparison results of the detection performance. Obviously, the learning-based schemes outperform the traditional criterion-based method in all the three data sets. Among the three learning schemes and the integrated scheme, the classification tree scheme works slightly better than the others. This is because some of the events are only identified by the classification tree scheme but not the other schemes.

3) Time for Prediction: Table IV shows the average training time and predicting time of the learning-based schemes. The experiments are run on a computer with core i7 CPU and 4-G RAM. Theoretically, training the models spends more time. The predicted decision can be made relatively quick once the models have been built. Note that KNN has no training phase. From the results, classification tree needs relatively shorter training time than SVM. For predicting time, all the schemes can make the detection decision within a very short time. However, KNN spends longer time than the others because KNN actually compares the distance of the observed data to all the feature records in the training data set. The other two schemes can calculate the predict decision within 1 ms. For the criterion-based method, the predicting process is almost finished instantly.

TABLE IV

AVERAGE COMPUTATION TIME

	KNN	**Tree**	**SVM**
Training Time (s)	N/A	0.0066	0.1064
Predicting Time (s)	0.0016	0.0004	0.0008

V. RELATED WORK

Recently, much research effort has been devoted to EEW over the world [2], [3], [15]. Japan has launched a public earthquake warning system since 2007 [16]. The system aims to provide the expected arrival time and seismic intensity before the hit of the strong motion in each subprefectural area. Taiwan has also been developing EEW systems to prevent disasters caused by earthquakes [1], [15]. In [1], a high-density seismic sensor network is built to capture earthquakes and provide early warnings. Earthquake detection has also been studied in Europe [17] and the United States [18]. In general, the average ground acceleration and ground velocity of seismic waves are used to determine the presence of an earthquake. Many further checking criteria that are manually tuned by experienced personnel are also developed to verify the occurrence of an earthquake and to avoid false alarms. In general, a trial-and-error process is inevitable to select appropriate system parameters.

Earthquake detection is highly correlated with anomaly detection. Many anomaly detection techniques have been explored extensively in the context of sensor networks [5], [19]. Conventional schemes for anomaly detection rely on maximum likelihood or hypothesis testing, which require the explicit statistical models for the investigated events [4], [20]. In [5], a collaborative detector based on hypotheses testing is proposed. The goal is to maximize the detection probability subject to a false alarm rate. In [21], the detection problem is formulated for M hypotheses in a large sensor network. It shows that it is asymptotically optimal to divide the sensors into $M(M-1)$ groups as the number of sensors goes to infinity. Fault tolerance for target detection is considered in [22]. The outliers of the observations are dropped when doing the fusion for the final consensus decision. In [23], Gaussian mixture models (GMMs) are used to build anomaly detection models in an on-line manner. The proposed approach can learn the GMM that adapts to nonstationary sources of data. Some other work develops the techniques for anomaly detection using SVMs [24], [25]. In general, the SVMs classify the observations into different categories using the hyperplane developed in the trained model. For most of the anomaly detection studies, the schemes must develop a theoretical model for the monitored signal. However, in practice, the event signals could be complex and do not follow the theoretical models.

Artificial intelligent techniques such as neural networks have also been exploited in analyzing seismic signals [26]–[28]. In [26], a knowledge-based system is developed to automatically interpret the seismic signal such as the time of the beginning and the end of an event. Some studies use artificial neural networks (ANNs) to solve the problem of discriminating natural earthquake signals from other waveforms of man-made events such as underground nuclear explosions. In [27], a simple three-layered ANN is built to determine the earthquake events and nuclear explosions. In [29], ANNs are used to classify seismic events in eastern and northern Europe. Although the previous work uses the real data to train the neural network models, most of the studies are based on the whole waveforms during the events, which may be suitable for postevent processing but inappropriate for real-time detection.

Recently, there is also work that develops earthquake detection systems based on community sensing that uses accelerometers in cell phones held by public community [30]. Each cell phone identifies the presence of an earthquake based on the hypothesis testing inferred from a statistical model. Then, a final decision is made by fusing the reports from the cell phones. The proposed scheme is interesting but difficult to realize in practice. in addition, the stability of the participating cell phones could substantially affect the performance of the detection.

VI. Conclusion

Machine learning is the technique that can make a variety of decisions about the observations based on the extracted knowledge from the historical data. In this paper, the learning-based schemes are exploited to identify the presence of earthquakes. The features of the seismic waves collected from historical events are used to train the classifier for earthquake detection. Three learning-based schemes are built in this paper to perform the verification of earthquake events, namely, the KNN, classification tree, and SVM. From the experiments, the detection performance of the learning-based schemes outperforms the traditional criterion-based method. In particular, one can envision that the reliability of earthquake detection can be dramatically increased if the learning-based schemes are adopted. Further studies about the fusion of local predictions and epicenter localization are worth to be explored.

References

[1] Y. M. Wu *et al.*, "A high-density seismic network for earthquake early warning in Taiwan based on low cost sensors," *Seismol. Res. Lett.*, vol. 84, no. 6, pp. 1048–1054, 2013.

[2] C. Satriano, Y.-M. Wu, A. Zollo, and H. Kanamori, "Earthquake early warning: Concepts, methods and physical grounds," *Soil Dyn. Earthquake Eng.*, vol. 31, no. 2, pp. 106–118, 2011.

[3] R. M. Allen, P. Gasparini, O. Kamigaichi, and M. Böse, "The status of earthquake early warning around the world: An introductory overview," *Seismol. Res. Lett.*, vol. 80, no. 5, pp. 682–693, 2009.

[4] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York, NY, USA: Springer, 2013.

[5] T.-L. Chin and Y. H. Hu, "Optimal detector based on data fusion for wireless sensor networks," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2011, pp. 1–5.

[6] N.-C. Hsiao, Y.-M. Wu, T.-C. Shin, L. Zhao, and T.-L. Teng, "Development of earthquake early warning system in Taiwan," *Geophys. Res. Lett.*, vol. 36, no. 5, 2009, Art. no. L00B02.

[7] D.-Y. Chen, N.-C. Hsiao, and Y.-M. Wu, "The Earthworm based earthquake alarm reporting system in Taiwan," *Bull. Seismol. Soc. Amer.*, vol. 105, no. 2A, pp. 568–579, 2015.

[8] B. Sharma, A. Kumar, and V. Murthy, "Evaluation of seismic events detection algorithms," *J. Geol. Soc. India*, vol. 75, no. 3, pp. 533–538, 2010.

[9] T. Perol, M. Gharbi, and M. Denolle, "Convolutional neural network for earthquake detection and location," *Sci. Adv.*, vol. 4, no. 2, 2018, Art. no. e1700578.

[10] J. Doubravová, J. Wiszniowski, and J. Horálek, "Single layer recurrent neural network for detection of swarm-like earthquakes in W-Bohemia/Vogtland—The method," *Comput. Geosci.*, vol. 93, pp. 138–149, Aug. 2016.

[11] D. M. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 1–28, 2011.

[12] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 21–27, Jan. 1967.

[13] W.-Y. Loh, "Classification and regression trees," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 1, no. 1, pp. 14–23, 2011.

[14] R.-E. Fan, P.-H. Chen, and C.-J. Lin, "Working set selection using second order information for training support vector machines," *J. Mach. Learn. Res.*, vol. 6, pp. 1889–1918, Dec. 2005.

[15] Y. M. Wu and T.-I. Teng, "A virtual subnetwork approach to earthquake early warning," *Bull. Seismol. Soc. Amer.*, vol. 92, no. 5, pp. 2008–2018, 2002.

[16] O. Kamigaichi *et al.*, "Earthquake early warning in Japan: Warning the general public and future prospects," *Seismol. Res. Lett.*, vol. 80, no. 5, pp. 717–726, 2009.

[17] C. D. Saragiotis, L. J. Hadjileontiadis, and S. M. Panas, "PAI-S/K: A robust automatic seismic P phase arrival identification scheme," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 6, pp. 1395–1404, Jun. 2002.

[18] R. M. Allen and H. Kanamori, "The potential for earthquake early warning in southern California," *Science*, vol. 300, no. 5620, pp. 786–789, 2003.

[19] S. Megerian, F. Koushanfar, G. Qu, and M. Potkonjak, "Exposure in wireless ad-hoc sensor networks," in *Proc. Int. Conf. Mobile Comput. Netw.*, Jul. 2001, pp. 139–150.

[20] B. Krishnamachari and S. Iyengar, "Distributed Bayesian algorithms for fault-tolerant event region detection in wireless sensor networks," *IEEE Trans. Comput.*, vol. 53, no. 3, pp. 241–250, Mar. 2004.

[21] J. N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Math. Control, Signals Syst.*, vol. 1, no. 2, pp. 167–182, 1988.

[22] T. Clouqueur, K. K. Saluja, and P. Ramanathan, "Fault tolerance in collaborative sensor networks for target detection," *IEEE Trans. Comput.*, vol. 53, no. 3, pp. 320–333, Mar. 2004.

[23] K. Yamanishi, J.-I. Takeuchi, G. Williams, and P. Milne, "On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms," in *Proc. Int. Conf. Knowl. Discovery Data Mining*, 2000, pp. 320–324.

[24] M. Davy, F. Desobry, A. Gretton, and C. Doncarli, "An online support vector machine for abnormal events detection," *Signal Process.*, vol. 86, no. 8, pp. 2009–2025, 2006.

[25] S. Rajasegarar, C. Leckie, J. C. Bezdek, and M. Palaniswami, "Centered hyperspherical and hyperellipsoidal one-class support vector machines for anomaly detection in sensor networks," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 3, pp. 518–533, Sep. 2010.

[26] C. Chiaruttini, V. Roberto, and F. Saitta, "Artificial intelligence techniques in seismic signal interpretation," *Geophys. J. Int.*, vol. 98, no. 2, pp. 223–232, 1989.

[27] F. U. Dowla, S. R. Taylor, and R. W. Anderson, "Seismic discrimination with artificial neural networks: Preliminary results with regional spectral data," *Bull. Seismol. Soc. Amer.*, vol. 80, no. 5, pp. 1346–1373, 1990.

[28] M. Bevreuther, R. Carniel, and J. Wassermann, "Continuous hidden Markov models: Application to automatic earthquake detection and classification at Las Canadas Caldera, Tenerife," *J. Volcanology Geotherm. Res.*, vol. 176, no. 4, pp. 513–518, 2008.

[29] P. S. Dysart and J. J. Pulli, "Regional seismic event classification at the NORESS array: Seismological measurements and the use of trained neural networks," *Bull. Seismol. Soc. Amer.*, vol. 80, no. 6B, pp. 1910–1933, 1990.

[30] M. Faulkner, M. Olson, R. Chandy, J. Krause, K. M. Chandy, and A. Krause, "The next big one: Detecting earthquakes and other rare events from community-based sensors," in *Proc. ACM/IEEE Int. Conf. Inf. Process. Sensor Netw.*, 2011, pp. 13–24.

Tai-Lin Chin (M'08) received the B.S. degree in computer science and information engineering from National Chiao-Tung University, Hsinchu, Taiwan, in 1995, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Wisconsin–Madison, Madison, WI, USA, in 2004 and 2006, respectively.

Since 2007, he has been with the Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, where he is currently an Associate Professor. His research interests include sensor network technologies, earthquake detection and early warning, mobile computing, and cloud computing.

Chin-Ya Huang (M'13) received the B.S. degree in electrical engineering from National Central University, Taoyuan City, Taiwan, in 2006, the M.S. degree in communication engineering from National Chiao-Tung University, Taiwan, in 2008, and the M.S. degrees in electrical and computer engineering and computer science and the Ph.D. degree in electrical and computer engineering from the University of Wisconsin–Madison, Madison, WI, USA, in 2008, 2010, and 2012, respectively.

From 2012 to 2013, she was a Senior Software Engineer with Qualcomm Tech., San Diego, CA, USA. She was an Assistant Research Fellow with the Information and Communication Technology Lab, National Chiao-Tung University, Taiwan, in 2013. From 2013 to 2015, she was a member of the Technical Staff with Optimum Semiconductor Tech. Inc., Tarrytown, NY, USA. From 2013 to 2018, she was also an Assistant Professor with National Central University, Taoyuan City, Taiwan. In 2018, she was with National Taiwan University of Science and Technology (NTUST), Taipei, Taiwan. Her research interest includes 5G and security and machine learning.

Shan-Hsiang Shen (M'17) received the M.S. degree from National Chiao Tung University, Hsinchu, Taiwan, in 2004, and the Ph.D. degree from the University of Wisconsin–Madison, Madison, WI, USA, in 2014.

He is currently an Assistant Professor with the Computer Science and Information Engineering Department, National Taiwan University of Science and Technology, Taipei, Taiwan. His research interests include software-defined networking, network function virtualization, network security, and cloud computing.

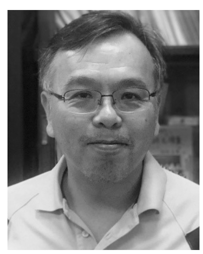
You-Cheng Tsai received the B.S. degree from Tamkung University, New Taipei City, Taiwan, in 2014, and the M.S. degree from the National Taiwan University of Science and Technology, Taipei, Taiwan, in 2017.

His research interests include sensor network and machine learning.

Yu Hen Hu (M'83–SM'88–F'99) received the B.S.E.E. degree from National Taiwan University, Taipei, Taiwan, in 1976, and the M.S.E.E. and Ph.D. degrees from the University of Southern California, Los Angeles, CA, USA, in 1982.

From 1983 to 1987, he was a member of the Faculty with the Electrical Engineering Department, Southern Methodist University, Dallas, TX, USA. Since 1987, he has been with the Department of Electrical and Computer Engineering, University of Wisconsin–Madison, Madison, WI, USA, where he is currently a Professor. In 1999, he was a Visiting Researcher with the Bell Laboratories, Holmdel, NJ, USA, and Microsoft Research, Beijing, China. He has been a Visiting Professor with Graduate Institute of Electronics, National Taiwan University, Taipei, in 2007 and 2015. His research interests include design and implementation of signal processing algorithms, computer aided design and physical design of VLSI, pattern classification and machine learning algorithms, and image and signal processing in general. He has authored or coauthored more than 380 technical papers, edited, and has coauthored four books and many book chapters in these areas.

Dr. Hu has served as an Executive Committee Member and the Secretary for the IEEE Signal Processing Society, a Board of Governor for the IEEE Neural Network Council representing the Signal Processing Society, and the Chair for Signal Processing Society Neural Network Signal Processing Technical Committee and the IEEE Signal Processing Society Multimedia Signal Processing Technical Committee. He has also served as a member for the IEEE Jack S. Kilby Signal Processing Medal Committee and a Steering Committee Member of the International Conference of Multimedia and Expo on behalf of the IEEE Signal Processing Society. He served as an Associate Editor for the IEEE TRANSACTION OF ACOUSTIC, SPEECH, AND SIGNAL PROCESSING, the IEEE SIGNAL PROCESSING LETTERS, *European Journal of Applied Signal Processing*, *Journal of VLSI Signal Processing*, and the *IEEE Multimedia Magazine*.



Yih-Min Wu received the Ph.D. degree in geophysics from National Central University, Taoyuan City, Taiwan, in 1999.

He was involved in earthquake early warning, real-time seismology, and seismotectonics. He is currently a Professor/Chair with the Department of Geosciences, National Taiwan University, Taipei, Taiwan. He is also a Jointed Research Fellow with the Institute of Earth Sciences, Academia Sinica, Taipei. In the past few decades, he has established as one of the leading figures in earthquake early warning. He has authored or coauthored more than 180 SCI journal papers.

Dr. Wu served as a member of the Editorial Board of *Earth, Planets, and Space* (SCI; 2015), and of the Editorial Board of Vietnam *Journal of Earth Sciences* since 2016. He served as an Associate Editor for the *Journal of Terrestrial, Atmospheric and Oceanic Sciences* (SCI; 2009–2015). He also served a Guest Editor for *Seismological Research Letters* (SCI) focus on 2018 Hualien, Taiwan Earthquake.