Citation: Ahn, H.; Kim, S.; Lee, K.; Choi, A.; You, K. Imbalanced Seismic Event Discrimination Using Supervised Machine Learning. Sensors 2022, 22, 2219. https://doi.org/10.3390/s22062219

Abstract: The discrimination between earthquakes and artificial explosions is a significant issue in seismic analysis to efficiently prevent and respond to seismic events. However, the discrimination of seismic events is challenging due to the low incidence rate. Moreover, the similarity between earthquakes and artificial explosions with a local magnitude derives a nonlinear data distribution. To improve the discrimination accuracy, this paper proposes machine-learning-based seismic discrimination methods—support vector machine, naive Bayes, and logistic regression. Furthermore, to overcome the nonlinear separation problem, the kernel functions and regularized logistic regression are applied to design seismic classifiers. To efficiently design the classifier, P- and S-wave amplitude ratios on the time domain and spectral ratios on the frequency domain, which is converted by fast Fourier transform and short-time Fourier transform are selected as feature vectors. Furthermore, an adaptive synthetic sampling algorithm is adopted to enhance the classifier performance against the seismic data imbalance issue caused by the non-equivalent number of occurrences. The comparisons among classifiers are evaluated by the binary classification performance analysis methods.

Keywords: seismic discrimination; artificial explosion; oversampling method; supervised machine learning

1. Introduction

Seismic signal analysis is one of the significant problems in geology. In particular, several studies have been conducted to discriminate between earthquakes and artificial explosions, and the results have exhibited the importance of seismic discrimination on the grounds of preparation and damage minimization caused by an earthquake. As earthquakes show aperiodic and non-stationary characteristics, the work of seismic event classification should be immediately performed. One factor that interferes with the classification is an artificial explosion. Artificial explosions, e.g., quarry blasts, explosive bomb tests, and underground nuclear tests, accompany seismic tremors, and the surface oscillations are similar to earthquakes [1].

Due to the analogous characteristics between earthquakes and artificial explosions, seismic stations that record and analyze various seismic waves can misidentify seismic events. Therefore, seismic signal discrimination needs to be performed before the precision analysis and initial response to the event. Typical seismic discrimination is performed by visually analyzing the recorded signals of earthquakes and explosions or extracting the characteristics of each record [2].

The process consumes a substantial amount of time and requires a large number of seismic data. To reduce the analysis time and to discriminate precisely between earthquakes and artificial explosions, state-of-the-art machine-learning methods have been
applied to design classifiers from seismic datasets. Many machine-learning-based discrimination methods have been introduced to discriminate between earthquakes and artificial explosions [3–6].

Li [3] conducted a classifier to discriminate earthquakes and several noises using a generative adversarial network for early warning system. Lyubushin [4] proposed a linear Bayesian discriminator. The discriminator was based on the properties of multi-fractal singularity spectrums that represent the fractal dimension of time moments. Lindenbaum [5] suggested a neural-net-based deep canonical correlation for automatic discrimination. A signal-to-noise ratio estimator and a short-time average/long-time average detector were introduced to identify triggers for seismic events. Bergman [6] induced an array-based seismic discrimination method using diffusion maps to discriminate between earthquakes and explosions. To handle the diffusion map, this work included a pre-processing step in which P-wave and S-wave analysis was done under a time-frequency representation.

However, seismic events do not occur evenly. In particular, the incidence rate for artificial explosions becomes lower than for earthquakes in regions where the earth’s interior and plate tectonics are activated. Conversely, artificial explosions often occur more frequently than earthquakes in certain regions [7]. Furthermore, the small-scale seismic events that are measured under local magnitudes ($m_l$) with less than 3.0 make it difficult to analyze features linearly. Therefore, a significant incidence discrepancy implies that data imbalances occur between earthquakes and artificial explosions.

The imbalance problem results in the overfitting or underfitting problem. Therefore, the discrimination accuracy of seismic signals decreases [8]. In this paper, support vector machine (SVM), naïve Bayes (NB), and logistic regression (LR) are used as seismic classifiers to discriminate binary classification.

In addition, SVM, NB, and LR-based discrimination methods are executed using limited seismic data recorded from a single station as a training dataset for the rapid response. Furthermore, to overcome the nonlinear data distribution caused by the similarity between earthquake and artificial explosion signals, the regularized logistic regression (RLR) and kernel methods for SVM are applied to construct seismic discrimination system.

For supervised machine-learning methods to improve the seismic discrimination accuracy, the feature vector is obtained from the time- and frequency-domain-based amplitude ratio. Furthermore, to prevent errors caused by domain change between time and frequency, fast Fourier transform (FFT) and short-time Fourier transform (STFT) are applied to reduce the conversion time and visualize the frequency band of the seismic signal, respectively. However, earthquakes and artificial explosions exhibit different seismic stroke incidence rates. Differences in the incidence rate lead to the imbalances in the dataset that can bring unintended error in designing a machine-learning-based seismic classifier.

To handle the imbalanced dataset, in this paper, the adaptive synthetic sampling (ADASYN) algorithm is applied to machine-learning algorithms in order to reduce the misclassification. To maintain a balance between majority class and minority class, the synthetic dataset is generated from the minority class using the ADASYN algorithm. The supervised training techniques derived in this paper are compared and evaluated to identify an optimal machine-learning-based seismic classifier between earthquakes and artificial explosions.

To determine the optimal machine-learning method for imbalanced seismic dataset, this paper is organized as follows. Section 2 analyzes the seismic signal in the time and frequency domain to form feature vector of machine-learning classifiers. Section 3 presents seismic discrimination methods using four different ADASYN-based machine-learning algorithms. The results of machine-learning classifiers are compared using evaluation indexes. Furthermore, the ADASYN algorithm demonstrates the solution of the imbalanced dataset problem and increases classification performance in Section 4. Finally, our conclusions are presented in Section 5.
2. Seismic Signal Discrimination

A seismic wave refers to the energy flow through the surface. Seismic events occur due to various causes, such as earthquakes caused by natural phenomena and artificial explosions caused by man-made situations. Although the causes of seismic events are different, the signals of seismic waves exhibit similar characteristics.

In seismology, several approaches have been attempted to demonstrate the difference between earthquakes and artificial explosions to determine the inherent characteristics of each seismic event [9–11]. To distinguish between earthquakes and artificial explosions, the amplitude ratio \( A_r \) and spectral ratio \( S_r \) for P- and S-waves that represent seismic characteristics are used to derive seismic type.

Many studies have shown that earthquakes and artificial explosions have distinct peaks in amplitude. The \( A_r \) method in [12,13] is based on the amplitude of P-wave and S-wave in the seismic signal in the time domain. The earthquake with a specific magnitude band \((1.8 < m < 3.0)\) and local distance from 50 km to 200 km, in general, exhibits a peak amplitude of P-wave \( (A_p) \) less than the peak amplitude of S-wave \( (A_s) \). In contrast, the artificial explosion has a larger \( A_p \) than \( A_s \). Based on the distinct peak values, the \( A_r \) method is used to obtain the peak amplitude ratio. Figure 1 shows the different peak amplitudes of earthquakes and artificial explosions, respectively.

\[
S_r = \frac{\int_{f_{l1}}^{f_{l2}} Amp(f) df}{\int_{f_{h1}}^{f_{h2}} Amp(f) df}, \quad (1)
\]

where \( Amp(f) \) represents the amplitude of the frequency domain and \( f_{h1}, f_{h2}, f_{l1} \) and \( f_{l2} \) denote the high-frequency and low-frequency bands. In general, the Fourier transform (FT) is used to represent the signal in the frequency domain. However, the time-frequency data is difficult to obtain from FT, since the FT is a function of frequency \( f \) and does not include the continuity property over time. Furthermore, when the signal data contain discontinuity and high-frequency components, the analysis of the signal’s features becomes more complicated.

![Figure 1. Peak amplitudes of P- and S-waves.](image-url)
The FFT algorithm is adopted to prevent data loss caused by the conversion of the domain and overcome the weakness of the FT. In addition, STFT algorithm is applied to the seismic signal to acquire the optimized frequency band in the seismic signal. The STFT of the seismic signal \( f(t) \) is expressed as follows \[15\]:

\[
\text{STFT}\{f(t)\}(\tau, f) = \int_{-\infty}^{\infty} f(t) \rho(t - \tau) e^{-j2\pi ft} dt,
\]

where \( \rho(t) \) represents a window function of the analysis. The resolution of the STFT is affected by the shape and size of the window function. When the STFT is applied, the transformed data are represented as a spectrum in the time-frequency division since the STFT is based on the FT in each time interval. Therefore, we can show the frequency change of the given data as a frequency spectrum. Based on the frequency spectrum, the optimized frequency bands are determined to obtain the \( S_r \) data of seismic signals.

3. Implementation of ADASYN-Based Classification Algorithms

To implement the machine-learning algorithms to distinguish between earthquake as the majority class and artificial explosions as minority class, four separate machine-learning algorithms are adopted as illustrated in Figure 2. As a first step, an imbalanced seismic dataset is collected as a training dataset. To determine out the feature vectors, the \( A_r \) method in the time domain and the \( S_r \) method in the frequency domain converted using FFT and STFT are applied, respectively. The training dataset can be expressed as,

\[
D = \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \\ \vdots & \vdots \\ d_{n1} & d_{n2} \end{bmatrix}, \quad l = \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_n \end{bmatrix}, \quad l_i \in \{+1, -1\},
\]

where the element \( (d_i) \) of \( D \) is a two-dimensional feature vector that describes \( A_r \) and \( S_r \) value. \( l_i \) is a label with +1 that denotes an earthquake and −1 that represents an artificial explosion for \( i \in \{1, 2, \cdots, n\} \). \( n \) denotes the number of training data. Second, the ADASYN algorithm is used to generate the synthetic dataset on minority class for the balance between majority class and minority class.

As a third step, classifiers based on SVM, NB, LR, and RLR with ADASYN are applied. Finally, a new seismic dataset that includes \( A_r \) and \( S_r \) is used to evaluate the performance of the model. For the evaluation of the classification model, nine methods, such as the specificity, sensitivity, receiver operating characteristic (ROC) curve, area under the curve (AUC), \( F1 \)-score, accuracy, Matthews correlation coefficient (MCC), Youden’s index (YI), and Fowlkes–Mallows index (FMI), are used as performance indicators of binary classification and a comparison value of each seismic discrimination performance.
3.1. Support Vector Machine

SVM is a supervised and non-probabilistic classification model for the machine learning. Several studies have used SVM to solve the classification problem and distinguish the dataset as true or false in various fields [16].

To compose the classifier using an SVM, a hyperplane in the feature space between two classes of data is determined by a weighting vector $\omega$. The weighting vector is derived with the maximum margin between the support vector and hyperplane to minimize the measurement uncertainty caused by the noise components in the measured dataset. The hyperplane is written as $\omega \cdot d_i + b = 0$, where $b$ is the unregulated bias term. The separating hyperplane can be defined as two classes, which are $l_i = +1$ and $l_i = -1$, respectively,

$$l_i = \begin{cases} +1, & \text{for } \omega \cdot d_i + b \geq 1, \\ -1, & \text{for } \omega \cdot d_i + b \leq -1. \end{cases}$$

(4)

$l_i(\omega \cdot d_i + b) - 1 \geq 0$ is derived by combining the inequality condition in Equation (4). The margin $2/\|\omega\|$ between the hyperplanes needs to be maximized to solve the constrained optimization problem under inequality constraints. To derive the maximum margin, the constrained optimization problem can be written as

$$\begin{align*}
\text{maximize} & \quad \frac{2}{\|\omega\|}, \\
\text{subject to} & \quad l_i(\omega \cdot d_i + b) \geq 1.
\end{align*}$$

(5)

Lagrange multipliers in the Karush–Kuhn–Tucker (KKT) condition are applied to maximize the objective function. Equation (5) can be transformed into $\|\omega\|^2/2$, which represents the converted form of the minimization, and the Lagrange function is

$$L(\omega, b, \alpha) = \frac{\|\omega\|^2}{2} - \sum_{i=1}^{n} \alpha_i [l_i(\omega \cdot d_i + b) - 1],$$

(6)

where $\alpha_i$ is the Lagrange multiplier, that is nonzero for support vectors [17]. According to the complementary slackness and stationarity in the KKT condition and dual solution, the decision function is obtained as

$$f(d) = \text{sgn} \left( \sum_{i=1}^{n} \alpha_i l_i(d_i \cdot d) \right).$$

(7)
In this paper, the $n$-th order polynomial kernel ($\chi(d_i, d_j) = (d_i^T d_j + c)^n, c > 0$) and radial basis function (RBF) kernel ($\chi(d_i, d_j) = \exp(-\|d_i - d_j\|^2/(2\sigma^2)), \sigma \neq 0$) is applied to handle the non-linearly separated dataset. Considering that the training dataset is mapped into high dimensional space, non-linear transformation of the input vector $\phi(d) : R^2 \rightarrow F$ should be applied, by which an optimal hyperplane can be obtained in the mapped high-dimensional feature space $F$. With the kernel functions in which $\chi(d_i, d_j) = \phi(d_i) \cdot \phi(d_j)$ the decision function of SVM after the nonlinear transformation is $\phi(d)$ written as [18]

$$
 f(d) = \text{sgn} \left( \sum_{i=1}^{n} a_i l_i \chi(d_i \cdot d_j) + b \right). \quad (8)
$$

### 3.2. Naïve Bayes Classifier

The naïve Bayes classifier is based on Bayes’ theorem with the independence assumption that all features are uncorrelated. The Bayes’ theorem is expressed as follows:

$$
 P(l \mid d) = \frac{P(d \mid l) \cdot P(l)}{P(d)} , \quad (9)
$$

where $P(l \mid d)$ denotes likelihood, $P(d \mid l)$ is the posterior probability, and $P(d)$ and $P(l)$ represent the evidence and predictor prior probability, respectively [19]. Equation (9) can be rewritten as Equation (10) using the chain rule and conditional independence assumptions,

$$
 P(l_j \mid d) = \prod_{i=1}^{n} P(d_i \mid l_i) \cdot P(l_i) P(d) . \quad (10)
$$

By applying the maximum posterior rule, the probability can be obtained by maximizing only the numerator equation. In binary classification using the NB model, $l_j$ is considered to be $+1$ or $-1$. The data label is determined as follows,

$$
 \hat{l} = \arg \max \prod_{i=1}^{n} P(d_i \mid l_i) P(l_i) , \quad (11)
$$

under the maximum posterior rule. $\hat{l}$ denotes the maximum a posteriori class that is determined by maximizing the classification performance [20].

### 3.3. Logistic Regression

As a regression method, LR calculates the probability that the realization of the output variable falls into a proper category. LR considers the conditional mean to obtain a probability value. The regression model $\beta^T d = \beta_0 + \beta_1 d_1 + \beta_2 d_2 + \cdots + \beta_n d_n$, where $\beta$ is a parameter vector, is to be established. Following the regression model and log-odds, the specific equation of the logistic function is as follows,

$$
 \theta(\beta, d) = \frac{1}{1 + e^{-\beta^T d}} = y , \quad (12)
$$

where $y$ is the probability that the label indicates $+1$ in binary dataset. $y$ considers a rational value between 0 and 1. To estimate the optimal parameter $\hat{\beta}$ that minimizes the regression model error caused by the noise components in measured dataset, the objective function is derived with the log-likelihood method that is the logarithm of the maximum likelihood method [21]. The objective function based on the log-likelihood estimation of $L(\beta)$ is expressed as follows:

$$
 \ln L(\beta) = \sum_{i=1}^{n} [y_i \ln(\theta(\beta, d_i)) + (1 - y_i) \ln(1 - \theta(\beta, d_i))] \\
 = \sum_{i=1}^{n} [y_i (\beta^T d_i)] - \sum_{i=1}^{n} \left[ \ln \left( 1 + e^{\beta^T d_i} \right) \right] . \quad (13)
$$
Based on Equation (13), the objective function is derived to find the optimal parameter $\beta$ that induces Equation (13) to have the largest value. To determine $\beta$, several optimization algorithms, e.g., the gradient descent, Broyden–Fletcher–Goldfarb–Shanno algorithm (BFGS), limited-memory BFGS, quasi-Newton, and Newton–Raphson method were applied [22–24]. Despite using the complete optimization method and general dataset, the classification performance with new data is often below expectations. Moreover, overfitting can occur in the case of high-dimensional regression models.

Many studies have been conducted to improve the classification performance and overcome the weakness of the LR algorithm caused by a small training set and low-feature dataset. To reduce misclassification, model selection algorithms, such as the Akaike information criterion and Bayes information criterion, are combined with the LR model [25]. However, earthquakes and artificial explosions occur irregularly, and are difficult to obtain a large number of dataset for training and selection.

In this paper, the RLR method is used to discriminate between earthquakes and artificial explosions and to solve the overfitting problem of the regression model. Moreover, the regularized term of the RLR method improves the classification performance to overcome the small dataset problem that is relatively susceptible to noise components in a dataset compared with a large dataset.

By inserting the regularized term into the objective function in Equation (13), the objective function becomes penalized. With the regularized term, the objective function $J(\beta)$ with constant $p$ induces appropriate results by adjusting the optimal parameter $\lambda$ to control the trade-off between the training data adjustment and misclassification avoidance caused by overfitting and underfitting. As represented in Equation (14), the RLR is designed with $L_2$-regularization that is known as a ridge-regularized term.

$$J(\beta) = -\sum_{i=1}^{n} \left( y_i (\beta^T d_i) \right) + \sum_{j=1}^{n} \left( \ln \left( 1 + e^{\beta^T d_j} \right) \right) - \lambda \sum_{k=0}^{p} (\beta_k^2).$$  \hspace{1cm} (14)

3.4. ADASYN

Earthquake events occur more frequently than artificial explosion events. Differences in the incidence rate result in the imbalanced data collected at seismic stations. Despite the mathematical evidence of machine-learning algorithms, machine-learning classifiers exhibit limited discrimination performance based on an imbalanced dataset. Moreover, data imbalance issues to overcome are studied in seismology and various fields, such as medical science, computer vision, and economics [26–29].

In this paper, the ADASYN method is adopted for machine-learning algorithms to overcome classification error caused by data imbalance. The ADASYN algorithm is an oversampling method to balance between majority class and minority class that represent earthquakes and artificial explosions, respectively [30]. ADASYN algorithm generates synthetic data according to the imbalance ratio between the majority class and minority class. The generated synthetic data belong to the minority class to balance the majority class data [31].

In this paper, we implemented the ADASYN algorithm for four machine-learning methods. Algorithm 1 describes how to generate the synthetic dataset to balance between majority class and minority class using ADASYN pseudocode. The input of the suggested algorithm is a training dataset $(x_i, y_i)$ that represents the feature vector and label, respectively. In the first step, the variable of $S$, that is the amount of synthetic dataset, is calculated based on the difference in data between the majority class ($A_{maj}$) and minority class ($A_{min}$) data. $\nu$ is used to determine the balance of the data generation required with $\nu \in [0, 1]$.

Steps 2–4 are processed to obtain the number of synthetic data ($g_i$) induced by the product of $\beta \in [0, 1]$ and normalized ratio ($\gamma_i$), that is calculated according to the $K$-nearest examples ($\gamma_i$) in the minority class. To generate the synthetic dataset, the loop from step 5 to 8 is performed. The function $\text{randomlySelect}()$ is used to randomly select $x_{zt}$, that
is $K$-nearest neighbor of $x_i$. The output $u_i$ of Algorithm 1 is the generated data for the minority class.

**Algorithm 1 ADASYN**

**Input:** Imbalanced training data $(x_i, y_i)$  
**Output:** Synthetic data in minority class $(u_i)$

1: $S = (A_{maj} - A_{min}) \cdot v$
2: $\rho_i = \gamma_i / K$
3: $\bar{\rho}_i = \frac{\rho_i}{\sum \gamma_i}$
4: $g_i = \beta \cdot \bar{\rho}_i$
5: for $i < g_i$ do
6: $x_{zi} = \text{randomlySelect}(x_i)$
7: $u_i = x_i + \text{rand}(0, 1) \cdot (x_{zi} - x_i)$
8: end for
9: return $u_i$

4. Classification Performance Evaluation

The simulation comprises three different procedures to discriminate between earthquakes and artificial explosions and compares the performance of each machine-learning method. As the first step, an imbalanced seismic dataset was obtained from the United States Geological Survey (USGS) and the Incorporated Research Institutions for Seismology (IRIS), which recorded from the station on the Pacific Northwest Seismic Network (PNSN) during 2017 and 2020 [32].

The feature vectors were derived using the $A_r$ and $S_r$ methods, respectively. The $A_r$ method was applied to an imbalanced training dataset to configure the feature vector. As a preprocessing step before the implementation of the $S_r$ method, the time domain was converted into the frequency domain using FFT, and optimized frequency bands were selected from the spectrogram obtained by STFT. As a second step, to overcome data imbalances, we applied the ADASYN algorithm to maintain the balance of the data between the majority class and minority class by generating synthetic data for the minority class. Machine-learning-based classifiers were applied using the balanced training dataset. Four classifiers were trained using a supervised model: SVM with RBF kernel function, NB, LR, and RLR. As a third step, to discriminate the seismic data, each classifier was applied to the new seismic dataset that was driven by the $A_r$ and $S_r$ methods.

Figure 3 shows the results of STFT and FFT of the earthquakes and artificial explosions, respectively. To simulate seismic discrimination, an imbalanced seismic dataset in which the ratio of earthquake to artificial explosion is 7 to 3 was used to train the classifiers. The training dataset was obtained from seismic events, which exhibited a magnitude range of 1.8–3 and a local distance of 50–200 km. Earthquake and artificial explosion seismic data were measured at the same station under identical conditions of geological characteristics and measurement environment.

$A_r$ and $S_r$ method with the spectral frequency bands $f_{11}$, $f_{12}$, $f_{h1}$ and $f_{h2}$ as 1, 5, 6, and 10 Hz was adopted, respectively. Using the ADASYN method, the training dataset becomes a balanced dataset in which the ratio of earthquake to artificial explosion is 5 to 5. Various classifiers were built according to the same dataset by selecting different machine-learning methods. The RBF kernel function used in the SVM classifier has the parameter of $\sigma = 0.3$. In addition, the fourth order polynomial kernel was used to configure the SVM classifier. The quasi-Newton method was used to optimize the objective functions of LR and RLR.

The value of $\lambda = 0.01$ was selected as the optimal parameter of the RLR to avoid overfitting. Figure 4a shows the test dataset to verify the performance of the ADASYN-based classifiers. Figure 4b–f represent the classification results of the SVM classifier with RBF kernel with $\sigma = 0.3$, SVM classifier with fourth order polynomial kernel, NB classifier, LR classifier, and RLR classifier with $\lambda = 0.01$, respectively. Table 1 shows the numerical parameters used for the realization of seismic event classification models.
Figure 3. The FFT and STFT results for the seismic signals in Figure 1.

Table 1. Numerical parameters for the classification models.

| Parameter | Method          | Value |
|-----------|-----------------|-------|
| $\lambda$ | RLR             | 0.01  |
| $K$       | ADASYN          | 5     |
| $\sigma$ | SVM with RBF    | 0.3   |
| $f_{h1}$  | $S_r$           | 1     |
| $f_{h2}$  | $S_r$           | 5     |
| $f_{l1}$  | $S_r$           | 6     |
| $f_{l2}$  | $S_r$           | 10    |
| $\alpha_l$| SVM             | 0.7   |

To compare the classification performances of SVM, NB, LR, and RLR, nine performance indicators (the ROC curve, AUC, sensitivity, specificity, F1-score, accuracy, MCC, YI, and FMI) were computed. The discrimination results are represented as $+1$ and $-1$. The result of $+1$ indicates that the test data are determined as an earthquake by the classifier. In contrast, $-1$ shows that the test data are identified as artificial explosion using the classifier. The classification outcomes have four possible results using actual labels and prediction results. The true positive (TP) and false negative (FN) indicate that the classifier accurately predicted the test data. On the contrary, the false positive (FP) and true negative (TN) imply that the classifier incorrectly predicted the test data. The confusion matrix for the binary classification is shown in Table 2.
The ROC curve is an effective measure of discrimination performance and a graphical analysis tool that was initially proposed in the field of signal detection [33]. The AUC value indicates the space under the ROC curve. The AUC value is 1 when the classifier is perfectly performed, whereas 0 describes random distinguishing. Sensitivity and specificity
are the ability of the classification to identify correctly the binary class. The sensitivity represents the conditional probability that is positive. On the contrary, the specificity means the conditional probability that indicates negative. The two methods are defined as follows:

\[
\text{Sensitivity} = \frac{TP}{TP + FN},
\]

\[
\text{Specificity} = \frac{TN}{TN + FP}.
\]

F1-score is used to show the balance between specificity and sensitivity. A higher F1-score indicates that the specificity and sensitivity of the classifier represent good performance. However, the F1-score is 0 when the specificity or sensitivity is zero, which indicates the fault classification result. Moreover, the MCC is used to evaluate the measurement of the binary classifier. The MCC value is always between $-1$ and $+1$, indicating a perfect negative correlation and a perfect positive correlation [34]. The accuracy is the proportion of the dataset that is correctly classified.

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}.
\]

\[
\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}}.
\]

\[
\text{YI} = \frac{TP}{TP + FN} - \left(1 - \frac{TN}{TN + FP}\right).
\]

\[
\text{FMI} = \sqrt{\frac{TP}{TP + FP} \cdot \frac{TP}{TP + FN}}.
\]

All performance metrics represent that closeness to the value of 1 indicates the best classification performance. The performance comparison for each ADASYN-based machine-learning method to discriminate between earthquakes and artificial explosions is listed in Tables 3 and 4 and Figure 5. Table 3 and Figure 5a show the performance comparison results when ADASYN is not applied. As a proof of the datasets balancing performance, Table 4 and Figure 5b are the results after applying ADASYN to input datasets.

In Figure 5, the black solid, black dotted, blue solid, magenta dotted, and red solid lines denote the SVM with RBF kernel function, SVM with polynomial kernel, NB, LR, and RLR trained based on imbalanced and balanced data with the ADASYN algorithm, respectively. In Tables 3 and 4, the fourth order polynomial kernel and RBF kernel with inverse variance ($\sigma = 0.3$) were used for SVM realization. Based on the results in Table 3, the NB classifier demonstrated the best discrimination performance. The NB classifier trained using the imbalanced seismic dataset showed that it is better than other classifiers in six performance metrics as shown in Table 3.

However, this shows that the NB classifier performs worse than some other classifiers with respect to the balance between sensitivity and specificity. The NB classifier’s MCC and YI, which represent binary classification performance are evaluated as the same with RLR classifier’s MCC and YI. However, the NB classifier is not the only appropriate method for seismic classifier since SVM with RBF and LR exhibited the highest scores for sensitivity and specificity, respectively. Table 4 represents the performance of the ADASYN-based machine-learning methods.
The best performance of discrimination between earthquakes and artificial explosions was identified as an SVM classifier with RBF kernel \((\sigma = 0.3)\) based on the sensitivity, specificity, and AUC results that exhibit the classification rate. In addition, MCC, YI, and FMI of SVM with RBF kernel function score that are the performance indexes to verify the binary classification ability show the highest performance compared with other classifiers. Based on the results of comparing the discrimination ability and performance measurement score of classifiers with the imbalanced seismic dataset and the balanced seismic dataset derived by the ADASYN algorithm, the ADASYN-based SVM classifier with RBF kernel function \((\sigma = 0.3)\) showed outstanding performance.

Table 3. Performance comparison of SVM with different kernel functions, NB, LR, and RLR trained based on an imbalanced seismic dataset.

|                  | SVM          | NB          | LR          | RLR          |
|------------------|--------------|-------------|-------------|--------------|
|                  | Polynomial   | RBF         |             |              |
| Sensitivity      | 0.9615       | 1.0000      | 0.9615      | 0.8462       |
| Specificity      | 0.8125       | 0.5000      | 0.8750      | 0.9375       |
|                  | 0.9259       | 0.8667      | 0.9434      | 0.8980       |
|                  | 0.9048       | 0.8095      | 0.9286      | 0.8810       |
|                  | 0.7974       | 0.6183      | 0.8478      | 0.7646       |
|                  | 0.7740       | 0.5000      | 0.8365      | 0.7837       |
|                  | 0.9266       | 0.8745      | 0.9336      | 0.8896       |

Figure 5. ROC curve of machine-learning methods and ADASYN-based machine-learning methods.
Table 4. Performance comparison of ADASYN-based SVM with different kernel functions, NB, LR, and RLR.

|                | SVM     | RBF     | NB     | LR     | RLR    |
|----------------|---------|---------|--------|--------|--------|
| **Polynomial** | 0.9615  | 1.0000  | 0.8846 | 0.4615 | 0.9615 |
| **RBF**       | 0.8125  | 0.8750  | 0.9375 | 1.0000 | 0.8750 |
| **AUC**       | 0.8870  | 0.9375  | 0.9111 | 0.7308 | 0.9183 |
| **F1-score**  | 0.9259  | 0.9630  | 0.9200 | 0.6316 | 0.9434 |
| **Accuracy**  | 0.9048  | 0.9524  | 0.9048 | 0.6667 | 0.9286 |
| **MCC**       | 0.7974  | 0.9014  | 0.8067 | 0.4961 | 0.8478 |
| **YI**        | 0.7740  | 0.8750  | 0.8221 | 0.4615 | 0.8365 |
| **FMI**       | 0.8686  | 0.9354  | 0.8839 | 0.7184 | 0.9037 |

5. Conclusions

In this paper, the main contribution of this research was the outstanding performance of seismic discrimination using machine learning in specific circumstances. The machine-learning models were trained by data of the time-frequency domain. In particular, the frequency domain data, \(S_r\), was converted using the FFT method. Machine-learning methods were applied to discriminate between earthquakes and artificial explosions on the imbalanced dataset.

Machine-learning methods were proposed using feature vectors obtained from the \(A_r\) method in the time domain and the \(S_r\) method in the frequency domain. To convert to the frequency domain for the \(S_r\) method, the FFT method was used. Moreover, to induce the optimal \(S_r\) method, the best frequency bands with low-frequency range (1–5 Hz) and high-frequency range (6–10 Hz) were derived using the STFT method. To overcome the performance degradation owing to the imbalanced seismic dataset, the ADASYN algorithm was used to keep a balance between datasets of earthquakes and artificial explosions. Based on the balanced dataset, the SVM, NB, LR, and RLR models were designed.

Using the four ADASYN-based machine-learning methods, the discrimination of the new seismic dataset that comprised earthquakes and artificial explosions was executed. In order to confirm and compare the discrimination performance of the four machine-learning models, the fitness evaluation was verified using various performance metrics of ROC, AUC, sensitivity, specificity, F1-score, accuracy, MCC, YI, and FMI, respectively. Using the various performance indexes, SVM with RBF (\(\sigma = 0.3\)) was proven to be the best classifier for seismic event discrimination.

Through comparisons of each model, we proved which machine-learning method had an edge on seismic discrimination. The improved performance of seismic discrimination was able to reduce a number of false alerts about whether a seismic event is from nature or human activity regardless of small sample rates. Therefore, the discrimination ability can create a quantitative discrimination criterion, and an immediate response can be used for the realization of an early warning system.

Author Contributions: This research was accomplished by all the authors. S.K. and K.Y. conceived the idea, performed the analysis, and designed the simulation; K.L., A.C. and H.A. conducted the numerical simulations; and S.K. and H.A. co-wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (NRF-2019R1A2C1002343, NRF-2020R1I1A1A01061632) and the BK21 FOUR Project.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.
Conflicts of Interest: The authors declare no conflict of interest.

References

1. Koper, K.; Pechamann, J.; Burlacu, R.; Pankow, K.; Stein, J.; Hale, J.; Roberson, P.; McCater, M. Magnitude-based discrimination of man-made seismic events from naturally occurring earthquakes in Utah, USA. Geophys. Res. Lett. 2016, 4, 10638–10645. [CrossRef]

2. Meier, M.; Ross, Z.; Ramachandran, A.; Nair, S.; Kundzicz, P.; Li, Z.; Andrews, J.; Hauksson, E.; Yue, Y. Reliable real-time seismic signal/noise discrimination with machine learning. J. Geophys. Res.-Solid Earth 2019, 124, 788–800. [CrossRef]

3. Li, Z.; Meier, M.; Hauksson, E.; Zhan, Z.; Andrews, J. Machine learning seismic wave discrimination: Application to earthquake early warning. Geophys. Res. Lett. 2018, 45, 4773–4779. [CrossRef]

4. Lyubushin, A.; Kaláb, Z.; Lednická, M.; Haggag, H. Discrimination of earthquakes and explosions using multi-fractal singularity spectra properties. J. Seismol. 2013, 17, 975–983. [CrossRef]

5. Lindenbaum, O.; Rabin, N.; Bregman, Y.; Averbuch, A. Seismic event discrimination using deep CCA. IEEE Geosci. Remote Sens. Lett. 2019, 17, 1856–1860. [CrossRef]

6. Bergman, Y.; Lindenbaum, O.; Rabin, N. Array based earthquakes-explosion discrimination using diffusion maps. Pure Appl. Geophys. 2020, 178, 2403–2418. [CrossRef]

7. Miao, F.; Carpenter, N.; Wang, Z.; Holcomb, A.; Woolery, E. High-accuracy discrimination of blasts and earthquakes using neural networks with multiwindows spectral data. Seismol. Res. Lett. 2020, 91, 1646–1659. [CrossRef]

8. Kotsiantis, S.; Kanellopoulos, D.; Pintelas, P. Handling imbalanced datasets: A review. GESTS Int. Trans. Comput. Sci. Eng. 2006, 30, 25–36.

9. Shang, X.; Li, X.; Morales-Esteban, A.; Chen, G. Improving microseismic event and quarry blast classification using artificial neural networks based on principal component analysis. Soil Dyn. Earthq. Eng. 2017, 99, 142–149. [CrossRef]

10. Kahabasi, A.; Moradi, A. Earthquake-explosion discrimination using waveform cross-correlation technique for mines in southeast of Tehran. J. Seismol. 2016, 20, 569–578. [CrossRef]

11. Hartse, H.; Taylor, S.; Phillips, W.; Randall, G. A preliminary study of regional seismic discrimination in central Asia with emphasis on western China. Bull. Seismol. Soc. Amer. 1997, 87, 551–568.

12. Wang, R.; Schmandt, B.; Kiser, E. Seismic discrimination of controlled explosions and earthquakes near Mount St. Helens using P/S amplitude ratio. J. Geophys. Res.-Solid Earth 2020, 125, e2020JB020338. [CrossRef]

13. O’Rourke, C.; Baker, G.; Sheehan, A. Using P/S amplitude ratio for seismic discrimination at local distance. Bull. Seismol. Soc. Amer. 2016, 106, 2302–2331. [CrossRef]

14. Yildirim, E.; Gülbğa, A.; Horasan, G.; Doğan, D. Discrimination of quarry blasts and earthquakes in the vicinity of Istanbul using soft computing techniques. Comput. Geosci. 2011, 37, 1209–1217. [CrossRef]

15. Lee, K.; Kwon, H.; You. K. Laser-interferometric broadband seismometer for epicenter location estimation. Sensors 2017, 17, 2423. [CrossRef]

16. Cervantes, J.; Lamont, F.; Mazahua, L.; Lopez, A. A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing 2020, 408, 189–215. [CrossRef]

17. Kim, S.; Lee, K.; You, K. Seismic discrimination between earthquakes and explosions using support vector machine. Sensors 2020, 20, 1879. [CrossRef]

18. Amari, S.; Wu, S. Improving support vector machine classifiers by modifying kernel functions. Neural Netw. 1999, 12, 783–789. [CrossRef]

19. Wong, T. A hybrid discretization method for naive Bayes classifiers. Pattern Recognit. 2012, 45, 2321–2325. [CrossRef]

20. Granik, M.; Mesyura, V. Fake news detection using naive bayes classifier. In Proceedings of the 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kyiv, Ukraine, 29 May–2 June 2017; pp. 900–903.

21. Rymarczyk, T.; Kozlowski, E.; Klosowski, G.; Niderla, K. Logistic regression for machine learning in process tomography. Sensors 2019, 19, 3400. [CrossRef]

22. Gao, W.; Goldfarb, D. Block BFGS methods. SIAM J. Optim. 2018, 28 1205–1231. [CrossRef]

23. Minka, T. A Comparison of Numerical Optimizers for Logistic Regression; Technical Report; Department of Statistics, Carnegie Mellon University: Pittsburgh, Pennsylvania, 2003.

24. Byrd, R.; Hansen, S.; Nocedal, J.; Singer, Y. A stochastic quasi-Newton method for large-scale optimization. SIAM J. Optim. 2016, 26, 1008–1031. [CrossRef]

25. Claeskens, G.; Croux, C.; Kerckhoven, J. Variable selection for logistic regression using prediction focused information criterion. Biometrics 2006, 62, 972–979. [CrossRef] [PubMed]

26. Rahman, M.; Davis, D. Addressing the class imbalance problem in medical datasets. Int. J. Mach. Learn. Comput. 2013, 3, 224–228. [CrossRef]

27. Liu, M.; Jervis, M.; Li, W.; Nivlet, P. Seismic facies classification using supervised convolutional neural networks and semisupervised generative adversarial network. Geophysics 2020, 85, 47–58. [CrossRef]

28. Sun, Y.; Wong, A.; Kamel, M. Classification of imbalanced data: A review. Int. J. Pattern Recognit. Artif. Intell. 2009, 23, 687–719. [CrossRef]

29. He, H.; Garcia, E. Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. 2009, 21, 1263−1284.
30. Zhu, Y.; Jia, C.; Li, F.; Song, J. Inspector: A lysine succinylation predictor based on edited nearest-neighbor undersampling and adaptive synthetic oversampling. *Anal. Biochem.* 2020, 593, 113592–113596. [CrossRef]

31. He, H.; Bai, Y.; Garcia, E.; Li, S. ADASYN: adaptive synthetic sampling approach for imbalanced learning. In Proceedings of the 2008 IEEE International Joint Conference on Neural Networks, Hong Kong, 1–8 June 2008; pp. 1322–1328.

32. United States Geological Survey. Available online: https://earthquake.usgs.gov (accessed on 1 December 2021).

33. Sokolova, M.; Japkowicz, N.; Szpakowicz, S. Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation. In Proceedings of the Australasian Joint Conference on Artificial Intelligence, Hobart, Australia, 4–8 December 2006.

34. Chicco, D.; Jurman, G. Optimal classifier for the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genom.* 2020, 21, 1–13. [CrossRef]

35. Liu, X. Classification accuracy and cut point selection. *Stat. Med.* 2012, 31, 2676–2686. [CrossRef]

36. Jha, S.; Pan, Z.; Elahi, E.; Patel, N. A comprehensive search for expert classification methods in disease diagnosis and prediction. *Expert Syst.* 2018, 36, e12343. [CrossRef]