IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 21, 2020, accepted December 30, 2020, date of publication January 5, 2021, date of current version January 13, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3049308

Robust and Efficient Classification for Underground Metal Target Using Dimensionality Reduction and Machine Learning

YADONG WAN[1], TONG LI[1], PENG WANG[1,2], SHIHONG DUAN[1], CHAO ZHANG[3], AND NA LI[4]

[1]School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China
[2]Datang Gohigh Data Networks Technology Company Ltd., Beijing 100191, China
[3]School of Materials Science and Engineering, University of Science and Technology Beijing, Beijing 100083, China
[4]School of Mathematics and Physics, University of Science and Technology Beijing, Beijing 100083, China

Corresponding author: Peng Wang (wangpeng.micl@gmail.com)

This work was supported in part by The National Key Research and Development Program of China under Grant 2017YFB0702300, in part by the National Natural Science Foundation of China under Project 61971031, in part by the Scientific and Technological Innovation Foundation of Shunde Graduate School, USTB under Grant BK19AF007, and in part by the China Postdoctoral Science Foundation under Grant 2019M650778.

ABSTRACT Underground metal target detection technology has been widely applied in industrial production, resource exploration, and engineering construction, etc. However, due to the influence of non-negligible noise and high dimensionality in collected data, achieving efficient and accurate underground target classification remains a grand challenge for further applications of underground metal target detection on portable devices with limited computing capability and energy supply. This study aimed to seek out robust and efficient data-based strategies to classify the underground metal targets of different shapes and materials based on electromagnetic induction detection. We investigated thirty-three classification strategies based on eleven dimensionality reduction methods, namely, the least absolute shrinkage and selection operator (LASSO), genetic algorithm-support vector machine (GA-SVM), Pearson correlation coefficient (PCC), mutual information (MI), maximal relevance minimal redundancy Pearson correlation (mRMRP), maximal relevance minimal redundancy mutual information (mRMRMI), statistical features (SF), principal component analysis (PCA), kernel principal component analysis (KPCA), locally linear embedding (LLE), and stacked denoising autoencoder (SDAE), and three machine learning models, namely, artificial neural network (ANN), linear support vector machine (L-SVM), and Gaussian Naïve Bayes (GNB). Several parameters, including classification accuracy, the number of features after dimensionality, the feature type importance, and the time consumption were considered to evaluate the data-based classification strategies. Among the classification strategies investigated and considering the above evaluation parameters, the artificial neural network (ANN) classifier assisted with the kernel principal component analysis (KPCA) feature extraction method yielded the best performance in the material-based classification (accuracy:0.99) and the shape-based classification (accuracy:0.99). The locally linear embedding (LLE) improved the robustness of machine learning classifiers and efficiency of the artificial neural network in the material-based classification (improvement of average accuracy:0.17, reduction of classification time cost:14%) and shape-based classification (improvement of average accuracy:0.16, reduction of classification time cost:22%). Our comparative investigation provides a robust and efficient data-based strategy for underground metal target classification, which is significant for applications of underground metal target detection on portable devices with limited computing capability and energy supply. The cross-combination strategy of dimensionality reduction methods and machine learning models provides a way to find the optimal machine learning model for underground target detection.

The associate editor coordinating the review of this manuscript and approving it for publication was Yongming Li.

INDEX TERMS Dimensionality reduction, machine learning, underground target classification, underground target detection.

I. INTRODUCTION

Underground metal target detection is the process of estimating the properties of subsurface metal items based on the observed data. It has been widely applied in resource exploration [1], engineering construction [2], military fields [3], and many other fields [4]–[6]. Geophysical exploration systems such as infrared remote sensing (RS) system [7], [8], electromagnetic induction (EMI) system [9]–[12] and ground-penetrating radar (GPR) [2], [13], [14] have proven to be successful in underground target detection. The EMI system, using the induced magnetic fields generated as observed data, has been an efficient way to detect underground metal targets in complex environments [15] owing to its strong penetrability, high precision, nondestructive, and sensitivity to metal materials. As a time-domain EMI method, transient electromagnetic (TEM) induction employs low frequencies ranging from tens of Hz to hundreds of kHz to increase sensitivity to conducting targets [16], [17], which has proven to be efficient in the detection and discriminating metal targets [18]. With the development of underground target detection technology [19], the need for detection has become more specific, and distinguishing the attributes such as the materials and shapes of targets has become an urgent problem, i.e., the underground metal target classification problem.

The existing classification strategies for underground metal targets mainly focus on two aspects: model-based methods [17], [20], [21] and data-based methods [22]–[24]. The model-based method needs to establish a forward model [25] and obtains the properties of the target through inversion. Therefore, the classification performance is greatly dependent on the rationality of the forward model and the reliability of inversion algorithms. Compared with the model-based methods, data-based methods have attracted considerable interest because they do not involve physical mechanisms and can be accessibly applied. The mechanism of data-based methods is the establishment of a mapping relationship between the influential factors (selected or extracted features from observed data) and the properties of underground targets. Therefore, there are two essential factors for data-based methods, namely, the mapping relationship and the influential factors (i.e., the model inputs).

In terms of the mapping relationship, abundant classification models have been applied for data-based classification. These methods include machine learning (ML) models such as support vector machine (SVM), artificial neural networks (ANN), random forest, AdaBoost algorithm, and so on [21], [22], [26]–[29]. Also, pattern matching methods, such as comparing the extracted features of targets with a given dictionary [30] and voting scheme which compares field data polarizabilities against templates in the library [31], are frequently employed for classification.

Regarding model inputs, one of the most important problems in data-based modeling is dealing with the high dimensional input space. The number of features generated by time-domain electromagnetic sensors is potentially very large which is dependent on the time channels, the number of receiver elements, and the spatial sampling density [23], [32]. Notably, there are lots of irrelevant, redundant, and noisy data as the amount of data becomes increasingly greater [33]. On the one hand, large-scale features bring problems such as requiring more storage and greater computational complexity of hardware, which may influence the performance of computers (long modeling time and lack of memory) [34]. On the other hand, large-scale features bring challenging problems in the learning process because of the "curse of dimensionality" and complications in the interpretation of data and results [35]. A model would show a lower prediction performance with the poor data, especially in machine learning, even if the model is very robust and effective. In many cases, the useful information that explains the mechanism generated by the data is in a smaller subset of features [36]. Therefore, extracting and selecting the most discriminative information from observed data is a crucial step in data-based modeling, which is known as dimensionality reduction. As a generally accepted rule, dimensionality reduction can be classified into two categories: feature selection and feature extraction [34], [37]. Feature selection refers to selecting a portion of the original dimensions that are most important for the task, while feature extraction refers to extracting a new representation set from the original dimension space [38]–[40]. For the research field of underground metal target detection, dimensionality reduction techniques have been frequently applied to drive more informative data-based classification models. Carin *et al.* utilized the relevance vector machine (RVM) and Bayesian elastic net to identify features that are relevant for discriminating unexploded ordnance (UXO) from the cluster and improve the generalization ability of the classifier [23]. Four statistic features of A-scans, such as the maximum value of amplitude signal graph, the number of peaks in the signals graph, skewness, and standard deviation values, were extracted for differentiated buried metal targets [24]. Kappler *et al.* applied Bayesian classification on four extracted features from the polarizabilities curve to discriminate UXO and harmless scrap metal [31]. Ammari *et al.* extracted geometric features from the induction data for identifying conductive objects [30]. In [41], dimensionality reducing techniques, such as the discrete spectrum of relaxation frequencies (DSRF) and the singular value decomposition (SVD), were used to improve the efficiency of the dictionary matching algorithm for location and orientation estimation. In those studies, various classification systems with different dimensionality reduction methods have been proposed, where building an appropriate feature set plays an important role in achieving high performance. As an important data preprocessing approach in data-based modeling, dimensionality reducing techniques

directly affect the efficient and discrimination performance of classification models [42]. To further improve the efficiency and the robustness of data-based methods, it is worthwhile to investigate the effect of different dimensionality reduction methods on the underground metal target detection.

In this article, we investigated the classification performance of frequently employed machine learning approaches as well as the discrepancy of different dimensionality reduction methods for data-based underground metal target detection. A simulation platform was established to verify the proposed classification strategies. The purpose of the detection was to estimate the shape type and material type of underground metal targets. The TEM method was utilized to acquire reliable data. Specifically, the general physical model and the approximate spheroid model in the TEM method were introduced. Then we examined the effectiveness of six feature selection methods and five feature extraction methods on three classification models, namely, artificial neural network (ANN), linear support vector machine (L-SVM), and Gaussian Naïve Bayes (GNB). These feature selection methods include the least absolute shrinkage and selection operator (LASSO), genetic algorithm-support vector machine (GA-SVM), Pearson correlation coefficient (PCC) and mutual information (MI), maximal relevance minimal redundancy Pearson correlation (mRMRP), maximal relevance minimal redundancy mutual information (mRMRMI). Besides, Feature extraction methods include statistical features (SF), principal component analysis (PCA), kernel principal component analysis (KPCA), locally linear embedding (LLE), and stacked denoising autoencoder (SDAE). The classification performance of classifiers was mainly estimated by accuracy and confusion matrix on a separate hold-out data set. To evaluate the quality of inputs, the relevance analysis and redundancy analysis were employed for the selected/extracted inputs obtained by different dimensionality methods. Besides, the effect of dimensionality reduction methods on the time consumption of the classifier and the importance of selected feature types in classification tasks were carefully analyzed. The main contributions of this paper can be summarized as follows:

- To seek out robust and efficient data-based strategies to classify the underground metal targets, we investigate thirty-three classification strategies based on eleven dimensionality reduction methods (LASSO, GA-SVM, etc.) and three classification models(ANN, L-SVM, GNB). Besides, the normalized feature type importance (NFTI) coefficient is utilized to analyze the significance of selected feature types and mutual information is proposed to evaluate the redundancy and relevance of inputs.
- The average accuracy of classifiers with different dimensionality reduction methods is examined under different signal-to-noise ratios (SNRs). Compared with original inputs, the locally linear embedding LLE) improved the robustness of machine learning classifiers and efficiency of the artificial neural network in the material-based classification (improvement of average accuracy:0.17, reduction of classification time cost:14%) and shape-based classification (improvement of average accuracy:0.16, reduction of classification time cost:22%). The comparison results indicate that the performance of classification models is sensitive to the noise level and appropriate dimensionality reduction methods (e.g., LLE and KPCA) can markedly enhance the robustness of models to noise.
- The comparative investigation proves that the artificial neural network (ANN) classifier assisted with the kernel principal component analysis (KPCA) feature extraction method yields the best performance in the material-based classification (accuracy:0.99) and the shape-based classification (accuracy:0.99). The classification strategy proposed in this study improves the prediction accuracy of underground metal target detection.

The detailed content of this paper is arranged as follows: Section II provides background on the TEM method. Then we introduce the general physical model and the approximate spheroid model. In Section III and Section IV, a theoretical description of six feature selection methods and five feature extraction methods, three machine learning models, the relevance analysis, the redundancy analysis are addressed. The simulation design and evaluation criteria are introduced in Section V. The simulation results are displayed in Section VI. The discussion is given in Section VII and the conclusion comes in Section VIII.

II. SYSTEM MODEL

The process of underground metal target detection in the TEM system is shown in Fig. 1. A typical TEM detector has a transmitting coil and a receiving coil. To collect the observation data of a target region, the transient current in the transmitting coil generates a pulse magnetic field (e.i., primary field $\boldsymbol{B}_P$) first. Then eddy currents induced in the metal target generate a secondary field $\boldsymbol{B}_S$ which decays over time. Finally, The secondary field $\boldsymbol{B}_S$ induces the measured voltage $\boldsymbol{V}_S$ in the receiving coil. For an enhanced discrimination of metal targets, a series of physics-based models [43]–[46] have been designed to describe the response of targets with intrinsic (e.g., target shape, material, and size) and extrinsic (e.g., target position and orientation) properties. The most frequently used physics-based model is the orthogonal dipole model [21], [43], [44], which is used to simulate the actual TEM response of underground metal targets in this study.

When the distance from the detector to a metal target is larger than the size of the metal target, the secondary field $\boldsymbol{B}_S$ can be approximated by the magnetic field generated by a dipole $\boldsymbol{m}$ [47]. In Fig.1, the secondary field $\boldsymbol{B}_S$ at the position of receiving coil is calculated as:

$$\boldsymbol{B}_S = \frac{\left(3\hat{\boldsymbol{r}}_{td}\hat{\boldsymbol{r}}_{td} - \boldsymbol{I}\right)\cdot \boldsymbol{m}_s}{4\pi r_{td}^3} \tag{1}$$

where $\hat{\boldsymbol{r}}_{td}$ represents the unit vector along $\boldsymbol{r}_{td} = \boldsymbol{r}_d - \boldsymbol{r}_t$, r_{td} is the modulus of $\boldsymbol{r}_{td}$, $\boldsymbol{I}$ denotes the identity matrix. The dipole

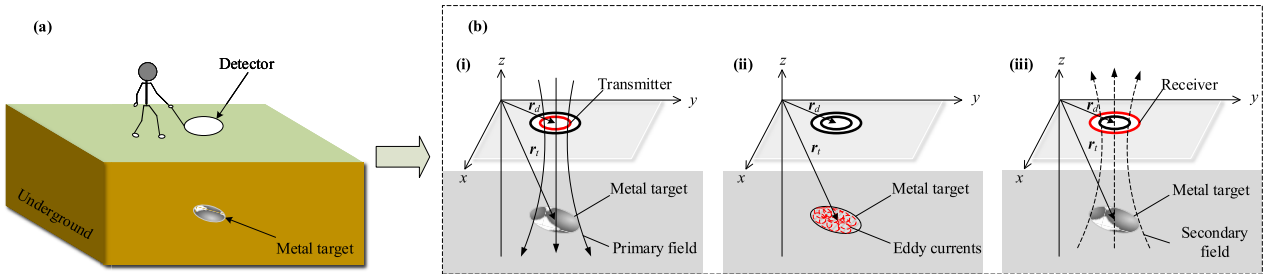

FIGURE 1. Process of underground metal target detection in the TEM system: (a)using a typical TEM detector to detect underground metal targets, and (b) the generation process of EMI response.

moment $\boldsymbol{m}_s$ is given by [48]:

$$\boldsymbol{m}_S = \boldsymbol{m_1} + \boldsymbol{m}_2 + \boldsymbol{m}_3 = \boldsymbol{M}\boldsymbol{B}_P \tag{2}$$

where $\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3$ are the three orthogonal dipoles, $\boldsymbol{B}_P$ represents the primary field at the position of the target, $\boldsymbol{M}$ is the magnetic polarizability tensor (MPT) [49] of the target and can be formulated as

$$\boldsymbol{M} = \boldsymbol{U}\begin{bmatrix} \beta_x & 0 & 0 \\ 0 & \beta_y & 0 \\ 0 & 0 & \beta_z \end{bmatrix}\boldsymbol{U}^T \tag{3}$$

where β_x,β_y,β_z are the principal polarizability elements of MPT, $\boldsymbol{U}$ represents the Euler rotation tensor [50].

In the TEM system, the target response $\boldsymbol{V}_S$ is proportional to the derivative of the secondary field $\boldsymbol{B}_S$ and it is calculated as [11]:

$$\boldsymbol{V}_S = -\frac{\left(3\hat{\boldsymbol{r}}_{td}\hat{\boldsymbol{r}}_{td} - \boldsymbol{I}\right)}{4\pi r_{td}^3}\boldsymbol{U}\begin{bmatrix} \frac{d\beta_x}{dt} & 0 & 0 \\ 0 & \frac{d\beta_y}{dt} & 0 \\ 0 & 0 & \frac{d\beta_z}{dt} \end{bmatrix}\boldsymbol{U}^T\boldsymbol{B}_P \tag{4}$$

The characteristic response matrix $\boldsymbol{L}(t)$ as the negative derivative of $\boldsymbol{M}$.

$$\begin{aligned}\boldsymbol{L}(t) &= \boldsymbol{U}\begin{bmatrix} -\frac{d\beta_x}{dt} & 0 & 0 \\ 0 & -\frac{d\beta_y}{dt} & 0 \\ 0 & 0 & -\frac{d\beta_z}{dt} \end{bmatrix}\boldsymbol{U}^T \\ &= \boldsymbol{U}\begin{bmatrix} l_1(t) & 0 & 0 \\ 0 & l_2(t) & 0 \\ 0 & 0 & l_3(t) \end{bmatrix}\boldsymbol{U}^T \end{aligned} \tag{5}$$

The characteristic response matrix $\boldsymbol{L}(t)$, which is described with the intrinsic properties (target shape, size, and material) of targets, is the key to underground metal target classification. The general model described above is suitable for the response modeling of arbitrary shape targets. However, most metal targets are axisymmetric objects, hence the characteristic response can be further simplified. The characteristic matrix of axisymmetric objects is given by

$$\boldsymbol{L}(t) = \boldsymbol{U}\begin{bmatrix} l_1(t) & 0 & 0 \\ 0 & l_2(t) & 0 \\ 0 & 0 & l_2(t) \end{bmatrix}\boldsymbol{U}^T \tag{6}$$

where $l_1(t)$ is the characteristic response parallel to the symmetry axis of the target, $l_2(t)$ represents the characteristic response perpendicular to the symmetry axis of the target. In the time-domain, the impulse response of the target can be described as an infinite sum of exponentials [48].

$$\begin{aligned} l_1(t) &= m_1(0)\delta(t) + \frac{\partial \sum_k u(t) m_{1k} \exp\left(-\omega_{1k} t\right)}{\partial t} \\ l_2(t) &= m_2(0)\delta(t) + \frac{\partial \sum_k u(t) m_{2k} \exp\left(-\omega_{2k} t\right)}{\partial t} \end{aligned} \tag{7}$$

To reduce the computational complexity of the characteristic response, a simple empirical function defined by a minimum number of parameters is utilized to replicate the features of the characteristic response. The parameterized characteristic response of sphere objects is given by

$$l_s(t) = k\left(1 + \frac{t^{\frac{1}{2}}}{\alpha^{\frac{1}{2}}}\right)^{-\beta} e^{-\frac{t}{\gamma}} \tag{8}$$

where t is the time gate, k, α, β and γ are the fitting parameters and a detailed description of these parameters is given in [51]. Therefore, the characteristic response of sphere objects is described as three different stages of the time decay curve [48]. In the duration of the early time stage, the decrease of response has a characteristic decay of $t^{-1/2}$. At intermediate times, the decay will have a power-law behavior of $t^{-\beta/2}$. And the exponential decay of the response at the late time stage is controlled by the parameter γ. Meanwhile, the axial and transverse responses of ellipsoids can be approximated as the characteristic response of spheres [52].

$$l_n(t) = \frac{2a^2 b}{9R_n^3}\frac{\mu_r + 2}{\mu_r}\left[\frac{1}{1 - A_n} + \frac{\mu_r - 1}{1 + A_n\left(\mu_r - 1\right)}\right] l_s\left(t, R_n\right) \quad n = 1, 2 \tag{9}$$

where R_1 and R_2 respectively correspond to the axial radius and transverse radius of the spheroid, μ_r is the relative permeability, and A_n is the demagnetization factor [52]. The target response $\boldsymbol{V}_S$, which is measured directly with most TEM receiving coil, can also be modeled with equation 9.

III. DIMENSIONALITY REDUCTION

In machine learning, dimensionality refers to the number of features in the data set. In the TEM system, the dimension of observed data is potentially high, depending on the time channels used, the number of receiver elements, and the spatial sampling density [23], [32]. Notably, there are lots of irrelevant, redundant, and noisy data as the amount of data becoming increasingly greater [33]. The advantages of dimensionality reduction can be summarized as follows [34], [53]. First, the data storage space and the computational complexity reduced with the number of dimensions comes down. Secondly, the data quality can be improved by removing redundant, irrelevant, and noisy features. Finally, the dimensionality reduction has been widely used to alleviate the ''curse of dimensionality'' and hence it can improve the generalization ability and efficiency of classification models. The dimensionality reduction strategies can be divided into two categories: feature selection and feature extraction. In this study, six feature selection methods and five feature extraction methods are investigated to improve the performance of classification.

A. FEATURE SELECTION

Feature selection aims to select the most important part of the feature subsets under a specific evaluation criterion and the search strategy. The feature selection methods are often divided into filter methods, wrapper methods and embedded methods [54].

1) WRAPPER METHODS

Wrapper methods utilize a classifier to select the feature subset. To accelerate the feature subset search process, a genetic algorithm-support vector machine (GA-SVM) is used in this study [55]–[57]. The GA-SVM uses genetic algorithms (GAs) to heuristically search the high-dimensional space of the feature subset. Then SVM is utilized to evaluates the fitness of feature subset with three-fold cross-validation. The advantages of GA-SVM can be summarized as follows. First, exhaustive exploration of search space with greater than 100 features is computationally intractable (i.e., 2^{100} possible subsets), while GAs can efficiently search the global optimal subset by evaluating a small number of candidates. Meanwhile, the SVM classifier involves fewer parameters than other machine learning algorithms (e.g., neural networks), has built-in regularization and acceptable classification performance, and is efficient in evaluating the fitness of candidates. Specifically, for the genetic algorithms, individuals in the population were binary strings, with 1 indicating that a feature was included, 0 indicating that it was not included (i.e., one-hot encoding [58]).

2) EMBEDDED METHODS

Embedded methods perform feature selection while training the learning model (jointly train the classifier and select the relevant features). They use a cost function to guide the feature search and are faster than wrapper methods [59]. As one of the most popular embedded methods, LASSO utilized the L_1-norm penalty to regularize the coefficient of linear regression and achieve the purpose of sparse feature selection [53], [60]. LASSO seeks to minimize the objective function as follows:

$$\sum_{i=1}^{n}\left(y_i-\sum_j x_{ij}\omega_j\right)^2+\lambda\sum_{j=1}^{m}|\omega_j| \tag{10}$$

where $\boldsymbol{x}$ is the input sample, n represents the number of samples, m is the number of features, ω_j is the coefficient of the jth feature, y_* is the outcome, λ is the shrinkage parameter. Due to the L_1-penalty, LASSO has the desirable quality of setting the coefficients of redundant features to zero. This provides a natural framework for feature selection in which models are constructed using only features with non-zero coefficients. Specifically, λ controls the degree of coefficient shrinkage and the value of λ is selected using three-folds cross-validation.

3) FILTER METHODS

Filter assess the goodness of features based on evaluation criteria. Since filter-based methods are not biased to any classification algorithm and computationally less expensive, they are popular for handling large feature space problems [61]–[63]. Among the filter methods, mutual information (MI) and Pearson correlation coefficient (PCC) are frequently used to capture linear and non-linear relationships between variables [64]–[66]. Based on whether they consider the dependencies between the features, the filter-based feature selection methods can be divided into bivariate and multivariate methods [54]. The former uses specific evaluations between the feature and the target regardless of other features, while the latter considers not only the correlation between the feature and the target but also the correlations between the feature and other features. In this study, we proposed two bivariate feature selection methods, namely, a ranking method based on Pearson correlation coefficient (PCC) and a ranking method based on mutual information (MI). Furthermore, two multivariate methods, namely, forward feature selection based on maximal relevance minimal redundancy mutual information (mRMRMI) [67] and forward feature selection method based on maximal relevance minimal redundancy Pearson correlation (mRMRP) [68] are proposed. The minimal-redundancy-maximal-relevance (mRMR) criterion, which is given by

$$\max_{x_i\in X-X_{m-1}}\left[f\left(\boldsymbol{x}_i;\boldsymbol{y}\right)-\frac{1}{|S_{m-1}|}\sum_{\boldsymbol{x}_j\in X_{m-1}}f\left(\boldsymbol{x}_i;\boldsymbol{x}_j\right)\right] \tag{11}$$

where $\boldsymbol{y}$ is the label variable and $f(\cdot)$ is the Pearson correlation or mutual information function, $X_{(m-1)}$ is the feature set with $m-1$ selected features, and the task is to selected the mth feature from the set $X-X_{(m-1)}$.

B. FEATURE EXTRACTION

Feature extraction methods extract new features from the original data set such that the features are changed. They can be divided into statistical features calculation, convex techniques (e.g., PCA, KPCA, LLE), and autoencoder (e.g., SDAE).

1) STATISTICAL FEATURES

Statistical features of signals are attractive for providing unique information about different types of the signal [69]. For each of the observed targets, statistical features of response signals are calculated including mean, variance, standard deviation, minimum, maximum, skewness [24], [70]. The total dimension of statistical features is 6.

2) PRINCIPAL COMPONENT ANALYSIS

Convex techniques optimize an objective function that does not contain any local optima, this is, the solution space is convex [71]. principal component analysis (PCA) is a linear technique, which performs dimensionality reduction by projecting the high-dimensional data into a linear subspace by singular value decomposition (SVD) [72]–[74]. To maximize the projection variance, PCA successively selects a set of projection directions with the largest projection variance as the projection matrix $\boldsymbol{M}$, which converts the high-dimensional data into the low-dimensional data representations. Specifically, PCA attempts to find a linear mapping $\boldsymbol{M}$ that maximizes the objective function

$$\begin{aligned} &\max_{M} \operatorname{trace}\left(\boldsymbol{M}^{\mathrm{T}} \operatorname{cov}(\boldsymbol{X}) \boldsymbol{M}\right) \\ &\text{s.t. } \boldsymbol{M}^{\mathrm{T}} \boldsymbol{M}=\boldsymbol{I} \end{aligned} \tag{12}$$

where $cov(\boldsymbol{X})$ is the sample covariance matrix of the data $\boldsymbol{X}$, the linear mapping $\boldsymbol{M}$ can be solved by the following equation

$$\operatorname{cov}(\boldsymbol{X}) \boldsymbol{\omega}_i=\lambda_{\boldsymbol{i}} \boldsymbol{\omega}_i \tag{13}$$

where $\lambda_{\boldsymbol{i}}$ is the ith principal eigenvalues, $\boldsymbol{\omega}_i$ is the ith principal eigenvector. The d principal eigenvectors of $\operatorname{cov}(\boldsymbol{X})$ with the largest eigenvalues are successively selected to form the projection matrix $\boldsymbol{M}$. The d-dimensional data representation $\mathbf{X}'$ is given by

$$\mathbf{X}'=\mathbf{X}\mathbf{M} \tag{14}$$

The dimension of the subspace, i.e., d is selected by three-fold cross-validation.

3) KERNEL PRINCIPAL COMPONENT ANALYSIS

Kernel principal component analysis (KPCA) is the reformulation of PCA in a high-dimensional space that is constructed using a kernel function [75]. Compared with PCA, KPCA computes the principal eigenvectors of the kernel matrix rather than the principal eigenvectors of the covariance matrix. The application of the "kernel trick" endows Kernel PCA with the ability to construct nonlinear mappings. Similarly, Kernel PCA solves the problem:

$$\boldsymbol{K} \boldsymbol{\alpha}_i=\lambda_i \boldsymbol{\alpha}_i \tag{15}$$

where K is a positive semi-definite kernel matrix, α_* is the eigenvector of kernel matrix, λ_* is the eigenvalue. The entries in kernel matrix K are defined by

$$(\boldsymbol{K})_{ij}=\kappa\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) \tag{16}$$

where $\kappa(\cdot)$ is the kernel function. The d principal eigenvectors of the kernel matrix with the largest eigenvalues are successively selected to form the projection matrix $\boldsymbol{\alpha}$. For a given input $\boldsymbol{x}$, the results of the projection is calculated as

$$\boldsymbol{x}'=\left\{\sum_{j=1}^{n} \alpha_1^j \kappa\left(\boldsymbol{x}_j, \boldsymbol{x}\right), \cdots, \sum_{j=1}^{n} \alpha_d^j \kappa\left(\boldsymbol{x}_j, \boldsymbol{x}\right)\right\} \tag{17}$$

where α_i^j denotes the jth values in the vector α_i. Since the kernel matrix is proportional to the square of the number of samples in the training set, the computational complexity of KPCA is larger than PCA. Additionally, the kernel function κ and parameter d play an important role in KPCA and they are selected by three-fold cross-validation.

4) LOCALLY LINEAR EMBEDDING

Locally linear embedding (LLE) attempts to preserve the local properties of the inputs. Specifically, the high-dimensional inputs are approximately reconstructed as a linear combination of their k nearest neighbors [76]. And in the low-dimensional representation of these inputs, LLE attempts to retain the reconstruction weights in the linear combinations (i.e., local properties of the inputs). Hence, finding the reconstruction weights amounts to minimizing the objective function:

$$\begin{aligned} &\min_{\omega_1, \omega_2, \cdots, \omega_n} \sum_{i=1}^{n}\left\|\boldsymbol{x}_i-\sum_{j=1}^{k} \omega_{ij} \boldsymbol{x}_j\right\|_2^2 \\ &\text{s.t. } \sum_{j=1}^{k} \omega_{ij}=1 \end{aligned} \tag{18}$$

where the input $\boldsymbol{x}_i$ which is reconstructed by its neighbors, $\boldsymbol{x}_j$ is the jth neighbors of $\boldsymbol{x}_i$, ω_{ij} is the reconstruction weight. In the low-dimensional space, the reconstruction weights are preserved and the d-dimensional data representation $\boldsymbol{x}'$ can be solved by minimizing the objective function:

$$\min_{\boldsymbol{x}_1', \boldsymbol{x}_2', \cdots, \boldsymbol{x}_n'} \sum_{i=1}^{n}\left\|\boldsymbol{x}_i'-\sum_{j=1}^{k} \omega_{ij} \boldsymbol{x}_j'\right\|_2^2 \tag{19}$$

The parameters of LLE, i.e., k and d are selected by three-folds cross-validation.

5) STACKED DENOISING AUTOENCODER

As an unsupervised feature learning algorithm, the auto-encoder (AE) network learns discriminative and effective features from a large amount of unlabeled data by minimization of the discrepancy of the output values to the input data [77]. The denoising auto-encoder (DAE) is developed from AE but it is more robust since DAE assumes that the input data contain noise and learns features from noisy data [78]. The DAE is trained to reconstruct a clean 'repaired' inputs $\boldsymbol{z}$ from the corrupted input data $\widetilde{\boldsymbol{x}}$. Also, DAEs can be stacked as stacking denoising autoencoders (SDAE) to obtain better feature representation [79]. The training of SDAE is layer-wise and each DAE with one hidden layer is trained independently. The squared loss of DAE is given by

$$L_\alpha(\boldsymbol{x}, z) = \alpha \left(\sum_{j \in \xi(x)} \left(\boldsymbol{x}_j - \mathbf{z}_j\right)^2 \right) + (1-\alpha) \left(\sum_{j \notin \xi(x)} \left(\boldsymbol{x}_j - \mathbf{z}_j\right)^2 \right) \quad (20)$$

where $\boldsymbol{x}_*$ is the input data, $\xi(\boldsymbol{x})$ denotes the index set of the components of $\boldsymbol{x}$ that were corrupted (i.e., $\widetilde{\boldsymbol{x}}$), α represents the weight of the reconstruction error on $\widetilde{\boldsymbol{x}}$. After training, the decoding layers of SDAE are removed and the encoding layers that produce features are retained. In this study, the SDAE consists of two DAEs with one hidden layer. The number of units in each hidden layer is 200 and 100, respectively. The weight α is considered as hyperparameters and is set as 0.1.

IV. MACHINE LEARNING MODELS

Each of the eleven feature sets obtained by different dimensionality reduction methods is utilized as an input of machine learning models. The machine learning models can be divided into generative classifiers and discriminative classifiers [80]. The generative classifiers learn the joint probability ($P(\boldsymbol{x}, y)$) of the input $\boldsymbol{x}$ and label y at first. Then, the posterior probability $P(y \mid \boldsymbol{x})$ is calculated by Bayes rules to make predictions and the most likely label y is picked. The discriminative classifiers directly model the posterior probability $P(y \mid \boldsymbol{x})$ from the training set. In this paper, three widely used classifiers are selected, including gaussian naïve Bayes (GNB), support vector machines and artificial neural network, and we briefly outline the attributes of them.

A. GAUSSIAN NAÏVE BAYES

The gaussian naïve Bayes (GNB) is a typical generative classifier that assumes conditional independence between every pair of features given the value of the class variable [81]. With a set of continuous features, GNB assumed the likelihood of the features to be a Gaussian distribution which can be written as

$$P\left(x_i \mid y\right) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{\left(x_i - \mu_y\right)^2}{2\sigma_y^2}\right) \quad (21)$$

where σ_y and μ_y are the standard deviation and mean of the ith feature x_i and they are estimated using maximum likelihood. Then, the posterior probability $P(y \mid \boldsymbol{x})$ is calculated as

$$P\left(y \mid x_1, x_2, \cdots, x_m\right) = \frac{P(y)\prod_{i=1}^{m} P\left(x_i \mid y\right)}{P\left(x_1, x_2, \cdots, x_n\right)} \quad (22)$$

where $P(y)$ is the prior probability for label y, $P(x_1, x_2, \cdots, x_n)$ is a constant for a given input $\boldsymbol{x}$. In this paper, the prior probability of each label variable is $\frac{1}{3}$ in the material-based classification task and is $\frac{1}{2}$ in the shape-based classification task.

B. SUPPORT VECTOR MACHINE

The support vector machine (SVM) [57], [82] is a discriminative classifier that attempts to maximize the soft margin while incurring a penalty when a sample is misclassified. The optimization problem of SVM is given by

$$\begin{aligned} &\min_{\boldsymbol{\omega}, b} \frac{1}{2}\|\boldsymbol{\omega}\|^2 + C\sum_{i=1}^{n}\xi_i \\ &\text{s.t. } \xi_i \geq 0, y_i\left(\boldsymbol{\omega}^{\boldsymbol{T}}\phi\left(\boldsymbol{x}_i\right) + b\right) \geq 1 - \xi_i \end{aligned} \quad (23)$$

where ω is a vector of coefficients, b is the intercept term, ξ_i controls the allowable margins on either side of the hyperplane, and ϕ is the identity function. The value of ω can be optimized by standard techniques of convex quadratic programming problems. Hence, the discriminant function is written as

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} \alpha_i y_i K\left(\boldsymbol{x}, \boldsymbol{x}_i\right) + b \quad (24)$$

where $K\left(\boldsymbol{x}, \boldsymbol{x}_i\right)$ is a kernel function and α_i is the dual coefficients. In this paper, the kernel function is taken as the linear kernel, C is taken as 1.

C. ARTIFICIAL NEURAL NETWORK

The artificial neural network (ANN) is a discriminative classifier. The ANN is a mathematical model that mimics the structure and function of the biological neural network [83]. The multi-layer perceptron (MLP) with one hidden layer can solve the nonlinearly separable problems and have been widely used in underground target detection [21], [24], [28]. In most applications, the MLP network contains an input layer, hidden layers, and an output layer. The training process in the neurons is given by

$$\mathrm{y}_{\mathrm{p}}^{(k)} = f\left(\sum_{i=1}^{N^{(k-1)}} \omega_{ip}^{(k-1)} \cdot y_i^{(k-1)} - \beta_p^{(k)}\right) \quad (25)$$

where $N^{(k-1)}$ represents the number of neurons in the $(k-1)$th layer, $\omega_{ip}^{(k-1)}$ is the connection weight between the ith neuron in the $(k-1)$th layer and pth neurons in the kth layer, $\mathrm{y}_{\mathrm{p}}^{(k)}$ is the output of the pth neuron in the kth layer, $\beta_p^{(k)}$ is the threshold of the pth neuron in the

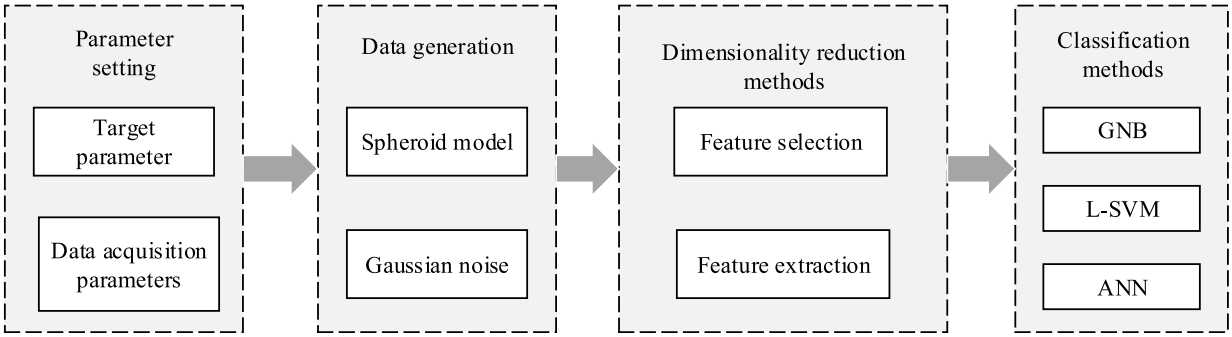

FIGURE 2. Process of the simulation platform.

kth layer, $f(\cdot)$ is the activation function. In this paper, the MLP with one hidden layer (50 neurons) is used for classification problems. The activation function of the hidden layer is the tanh function, which is given by

$$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{26}$$

And the limited-memory BFGS (L-BFGS) [84] have been used to optimize the weights and thresholds in the layers. The acronym for each dimensionality reduction method and classification models was listed in Table 1.

TABLE 1. Summary of the used dimensionality reduction and classification models with the acronyms and full names.

Acronym	Full name
GNB	Gaussian Naïve Bayes
L-SVM	Linear support vector machine
ANN	Artificial neural network
MI	Mutual information
PCC	Pearson correlation coefficient
mRMRMI	Maximal relevance minimal redundancy mutual information
mRMRP	Maximal relevance minimal redundancy Pearson correlation
LASSO	Least absolute shrinkage and selection operator
GA-SVM	Genetic algorithm-support vector machine
SF	statistical feature
PCA	Principal component analysis
KPCA	Kernel principal component analysis
LLE	Locally linear embedding
SDAE	Stacked denoising autoencoder

V. SIMULATION DESIGN

A. DATA PREPROCESSING

We build a simulation platform, which is based on the ellipsoid model proposed in Section II, to verify the improvement of classification performance by applying dimensionality reduction methods. The simulation parameters are set as shown in Table 2 and Table 3.

The shape types of metal targets are divided into two categories according to the aspect ratio $\rho = \frac{R_2}{R_1}$: $\rho > 1$ (i.e., the oblate ellipsoid), $0 < \rho < 1$(i.e., the prolate ellipsoid). The material types of metal targets are divided into three categories, including steel, nickel(Ni), and aluminum(Al). The process of the simulation platform is shown in Fig. 2.

TABLE 2. Data acquisition parameters of simulation platform.

Parameter	Description	Optional values	Unit
SNR	Signal to noise ratio	5,10,15,20,25,30,noiseless	dB
T	Time range of recorded responses	$[10^{-8}, 10]$	s
m	Data length of recorded responses	400	/

TABLE 3. Target parameters of spheroid mode.

Parameter	Description	Range	Unit
R_1	Axial radius of metal targets	$[0.01, 1.50]$	m
R_2	Transverse radius of metal targets	$[0.01, 1.50]$	m
μ_r	The relative permeability of metal targets	$\{696.30, 99.47, 1.00\}$	/

During underground metal target detection, there are some signal sources from the presence of geologic noise. One of the most common ways of describing noise is to assume that it is random and can be denoted by Gaussian statistics [48]. The unbiased noise sources are quantified with SNR, which is given by

$$SNR = 10 \log \frac{P_s}{P_n} \tag{27}$$

where P_s represents the signal energy, P_n denotes the noise energy. In this paper, the SNR is set to $30dB$ as default. Also, min-max normalization is used to scale the range of features, which could improve the performance of the algorithms based on the gradient optimization (e.g., neural network). The normalized feature value is given by

$$x_{norm} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{28}$$

where x_{norm} represents the normalized feature value of x, and x_{max} and x_{min} are estimated from the given data set. The simulation platform is implemented in Python and run on a PC feature Intel core i7-6700 CPU(3.4 GHz) and 16 GB of RAM.

B. FRAMEWORKS OF CLASSIFICATION PERFORMANCE EVALUATION

To avoid double-dipping, the performance of classification strategies are obtained by using two test modes: hold-out and k-fold cross-validation frameworks [85], [86].

1) HOLD-OUT APPROACH

In hold-out frameworks, the data set is split into the training set and the testing set by a certain percentage ratio. In this paper, for the 1026 ellipsoid samples with various shapes (513 oblate ellipsoid samples, 513 prolate ellipsoid samples) and materials (342 steel, 342 Ni, 342 Al), we randomly assign 718 samples to the training set and 308 samples to the testing set according to a ratio of 7: 3. The training set is exclusively used for training dimensionality reduction methods and classification models and the held-out testing set is exclusively used for model evaluation.

2) K-FOLD CROSS-VALIDATION

In k-fold cross-validation, the data set is split into k equal parts. Then, $k-1$ parts are used for training and the rest part is used for validating alternately. In this paper, the three-fold cross-validation is used in the training phase of dimensionality reduction methods.

C. RELEVANCE AND REDUNDANCY ANALYSES

The candidate inputs usually include features that are either irrelevant to the classification problem or redundant. The irrelevant features are uninformative variables, which add noise and complexity to the model, while the redundant features increasing the dimensionality of the model identification problem without providing any additional predictive benefit. To quantify the redundancy and correlation of inputs, the Pearson correlation coefficient and mutual information are utilized to evaluate the linear and nonlinear relation between features and label variables in the data set, respectively. The relevance of the feature set X is defined as

$$Relevance(X, \boldsymbol{y}) = \frac{1}{|X|} \sum_{\boldsymbol{x}^i \in X} f\left(\boldsymbol{x}^i; \boldsymbol{y}\right) \tag{29}$$

where y is the label variable, x_i is the ith feature, and $f(\cdot)$ is the Pearson correlation or mutual information function. And the redundancy of X is defined as

$$Redundancy(X) = \frac{1}{|X|} \sum_{\boldsymbol{x}^i \in X} \frac{1}{|X_{m-1}|} \sum_{\boldsymbol{x}^j \in X_{m-1}} f\left(\boldsymbol{x}^i; \boldsymbol{x}^j\right) \tag{30}$$

where $X_{(m-1)}$ represents the feature set with $m-1$ features. The operator combines the above two metrics is defined as

$$\phi = Relevance - Redundancy \tag{31}$$

In this paper, ϕ, *Relevance* and *Redundancy* metrics are utilized to quantify the quality of the given data set.

D. CLASSIFICATION ERROR EVALUATION CRITERIA

In this paper, accuracy and confusion matrix are utilized to quantify the performance of classification models. The accuracy is defined as the proportion of the instances correctly classified in the data set. True positive (TP, the correctly predicted positive instances number), false positive (FP, the incorrectly predicted positive instances number), false negative (FN, the incorrectly predicted negative instances number), and true negative (TN, the correctly predicted positive instances number) are utilized to calculate the accuracy, which is given by

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{32}$$

The accuracy reflects the overall performance of classification models. Besides, the confusion matrix [87] is used to further explore the performance of classification models on specific metal targets. The confusion matrix of the binary classification problem is shown in Table 4.

TABLE 4. Confusion matrix for binary classification.

True class	Predicted class	
	Positive	Negative
Positive	TP	FN
Negative	FP	TN

E. EVALUATION OF SELECTED FEATURE NUMBER IN FILTER-BASED FEATURE SELECTION

The selected feature number is a key parameter, which plays an important role in the performance of the filter-based feature selection approaches (i.e., PCC, MI, mRMRMI, and mRMRP). To reduce the performance bias caused by different classifiers, the average accuracy of three classifiers is utilized to evaluate the selected feature number for filter-based feature selection methods. For each filter-based feature selection approach, a range of feature number from 5 to 400 with an interval of 5 is selected. The three classifiers are utilized to evaluate the prediction accuracy with three-fold cross-validation.

F. EVALUATION OF FEATURE TYPE IMPORTANCE IN CLASSIFICATION

Feature selection methods are normally distinguished from feature extraction methods in that they preserve the original set of features. Therefore, the importance of response signals in different time stages can be quantified by analyzing the distribution of selected features. In this paper, the normalized feature type importance (NFTI) coefficient [88] is used to describe the selected feature types for six feature selection methods. Specifically, we count the number of selected features corresponding to their time stage. Then, the numbers are normalized by the selected feature number for all feature types. Finally, we get the NFTI coefficient for three feature types in each feature selection method.

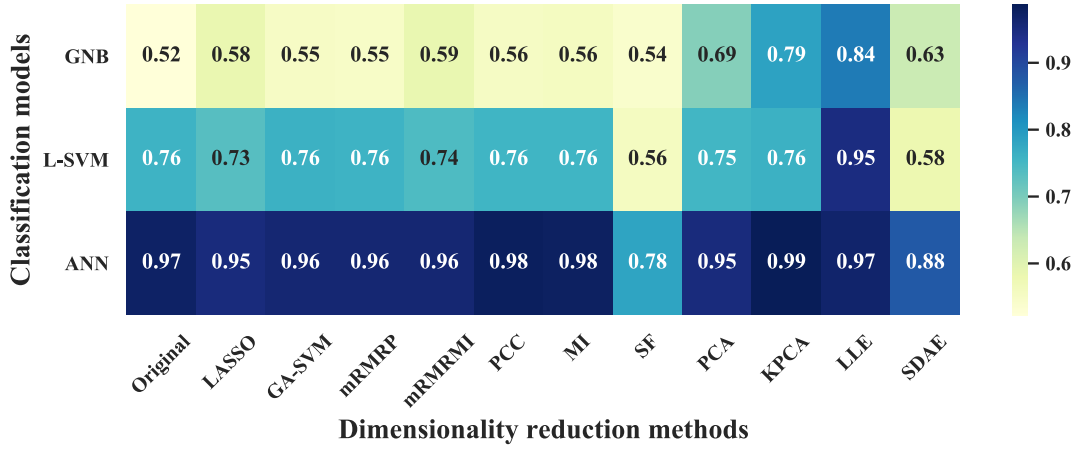

FIGURE 3. Accuracy heatmap of classification models (in columns) and dimensionality reduction methods (in rows) in material-based classification.

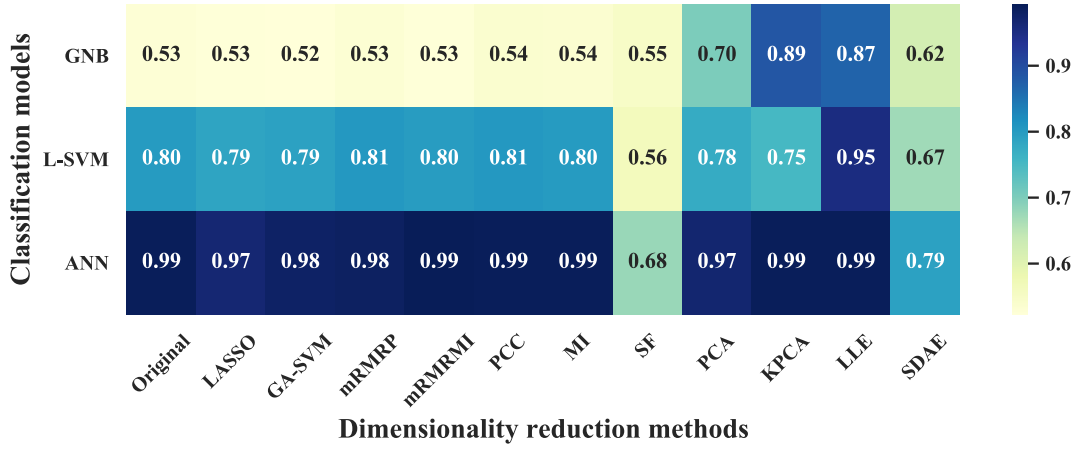

FIGURE 4. Accuracy heatmap of classification models (in columns) and dimensionality reduction methods (in rows) in shape-based classification.

VI. RESULTS

A. COMPARISON OF DIMENSIONALITY REDUCTION METHODS AND CLASSIFICATION MODELS

To compare different dimensionality reduction methods for underground metal target classification problems, we extracted and selected informative features from the response signals with 11 dimensionality reduction methods. Then, we evaluated the performance of 33 combinations of dimensionality reduction methods and classification models on the held-out test set.

Fig. 3 and Fig. 4 depict the accuracy results, which highlight four interesting characteristics. Firstly, regardless of the dimensionality reduction method, classifier ANN is always superior to the other two classifiers for material-based classification and shape-based classification. Secondly, the accuracy of shape-based classification is slightly better than material-based classification, Thirdly, among the eleven dimensionality reduction methods, feature extraction method LLE exhibited a significant accuracy improvement for classifier GNB (material-based classification: 0.32, shape-based classification: 0.34) and classifier L-SVM (material-based classification: 0.19, shape-based classification: 0.15). Finally, compared with the original inputs, most of the data sets after dimension reduction a better classification performance with majority classifiers. Concretely, for the material-based classification problem, feature extraction method KPCA and classifier ANN achieved the highest prediction accuracy (accuracy: 0.99), followed by feature selection method PCC + classifier ANN(accuracy: 0.98) and feature selection method MI+classifier ANN (accuracy: 0.98). Meanwhile, for the shape-based classification problem, feature extraction method KPCA+classifier ANN, feature extraction method LLE+ classifier ANN, feature selection method mRMRMI +classifier ANN, feature selection method PCC +classifier ANN, feature selection method MI +classifier ANN achieved the highest prediction accuracy (accuracy:0.99). However,

feature extraction methods SF and SDAE showed low accuracy with majority classifiers, and classifier GNB showed lower accuracy.

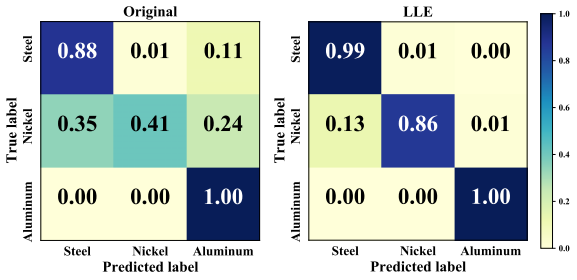

FIGURE 5. The normalized confusion matrices in material-based classification.

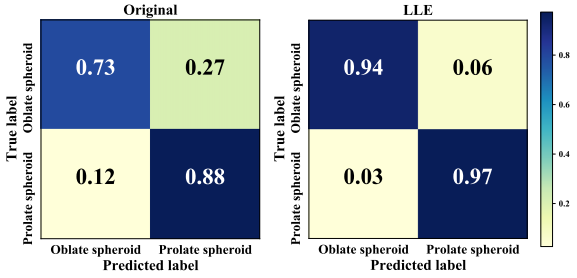

FIGURE 6. The normalized confusion matrices in shape-based classification.

To further explore the performance of classification models on specific metal targets and the corresponding influence of dimensionality reduction methods, Fig. 5 and Fig. 6 shows the normalized confusion matrices (dimensionality reduction method: LLE, classifier: L-SVM) in material-based classification and shape-based classification respectively. For the material-based classification, the classification performance of aluminum targets is best for all of the aluminum targets are classified correctly. But the classification performance of nickel targets is relatively poor since 0.59 of nickel targets are misclassified. After dimension reduction with LLE, the precision of nickel targets is increased to 0.86 and the precision of steel targets is increased to 0.99. Meanwhile, for the shape-based classification, the classification performance of prolate spheroid targets is better. After dimension reduction with LLE, the precision of oblate spheroid targets is increased to 0.94, and the precision of prolate spheroid targets is increased to 0.97.

B. IMPACTS OF DIMENSIONALITY REDUCTION METHODS ON FEATURE DENOISING

To examine the impact of the eleven dimensionality reduction methods on feature denoising, the performance of dimensionality reduction methods under different SNR are shown as follows. Fig. 7 and Fig. 8 depict the results of material-based classification and shape-based classification respectively.

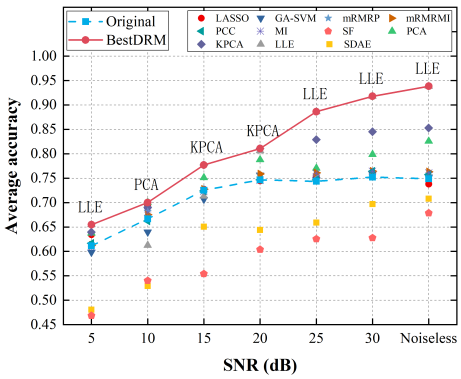

FIGURE 7. The average accuracy (in columns) for different dimensionality reduction methods under different SNR (in rows) in material-based classification.

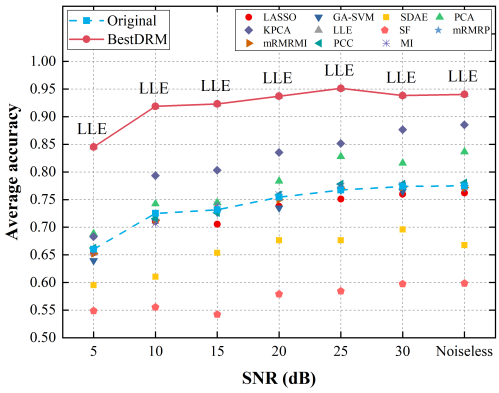

FIGURE 8. The average accuracy (in columns) for different dimensionality reduction methods under different SNR (in rows) in shape-based classification.

To reduce the performance bias caused by different classifiers, the average accuracy of three classifiers was utilized to evaluate the performance of dimensionality reduction methods. The comparison results highlight two important characteristics. First, the average accuracy increases with the SNR in Fig. 7 and Fig. 8. It indicates that the classification models have limited anti-noise ability, this is, the noise in the data set would greatly affect the classification performance. Besides, the average accuracy of shape-based classification is higher than material-based classification under a specific SNR, and classifiers for the shape-based classification are more robust. Secondly, majority dimensionality reduction methods would not significantly reduce the average accuracy of classifiers in the material-based classification and shape-based classification. Specifically, compared with the original data set, each subset obtained by feature selection methods, including mRMRP, mRMRMI, PCC, MI, GA-SVM, and LASSO, has equivalent average accuracy. Feature extraction methods, such as LLE, KPCA, and PCA, greatly improve the average accuracy in both material-based classification and shape-based classification. Particularly, LLE is the best dimensionality reduction methods (BestDRM) in material-based classification when the SNR

is greater or equal to 25dB. Meanwhile, LLE outperformed the other dimensionality reduction methods on the average accuracy under different SNR in shape-based classification. However, feature extraction methods, such as SF and SDAE would reduce the average accuracy of classifiers in material-based and shape-based classification.

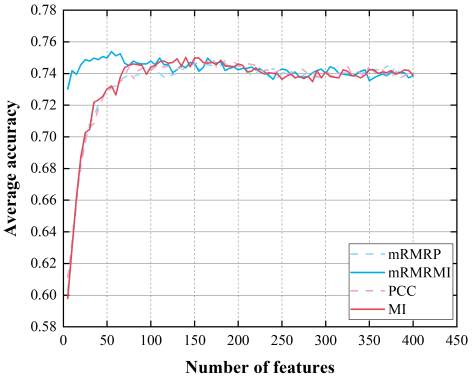

FIGURE 9. The average accuracy (in columns) for different filter-based feature selection methods with different selected feature numbers (in rows) in material-based classification.

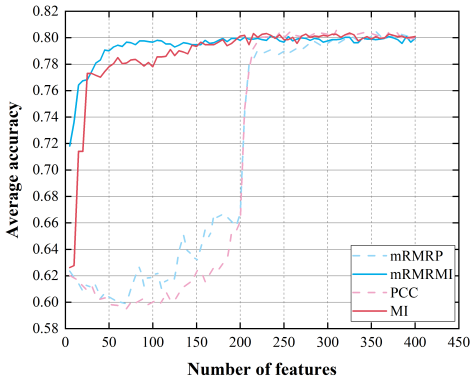

FIGURE 10. The average accuracy (in columns) for different filter-based feature selection methods with different selected feature numbers (in rows) in shape-based classification.

C. IDENTIFYING SELECTED FEATURE TYPES FOR CLASSIFICATION

The selected feature number plays an important role in the performance of the filter-based feature selection approaches, such as mRMRP, mRMRMI, PCC, and MI. Hence, we adjusted the parameters for each method to obtain a range of selected subsets. In this paper, the range of selected feature number is from 5 to 400 with an interval of 5 as shown in Fig. 9 and Fig. 10. And each of the selected features is then trained by three classifiers respectively with repeated three-fold cross-validation. For the material-based classification, mRMRMI is the fastest feature selection to achieve the highest average accuracy (accuracy: 0.75, feature number: 55). Besides, we found that the average accuracy increased with the selected feature number when the feature

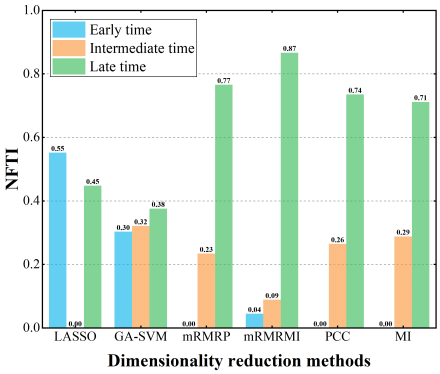

FIGURE 11. The NFTI coefficients (in columns) of feature sets obtained by different methods (in rows) in material-based classification.

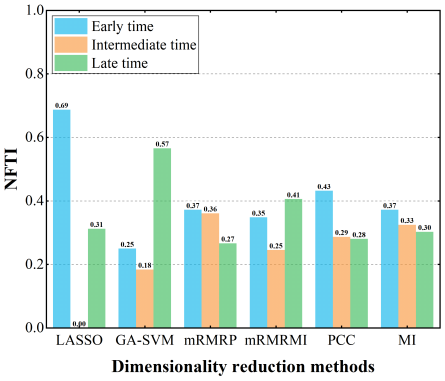

FIGURE 12. The NFTI coefficients (in columns) of feature sets obtained by different methods (in rows) in shape-based classification.

number is smaller than 150. Then, the average accuracy is relatively stable and slightly decreases when the selected feature number is larger than 150. For the shape-based classification, mRMRMI and MI achieve a stable average accuracy (average accuracy: 0.78) with a smaller selected feature number (feature number: 100) than the majority of feature selection methods. Then, the optimal parameters of filter-based feature selection methods are the feature number with the highest average accuracy. The decay response signals in the TEM system can be divided into three stages, including the early time stage, the intermediate time stage, and the late time stage. Fig. 11 and Fig. 12 depict the NFTI results of material-based classification and shape-based classification, respectively. For the material-based classification, the late time response signals have the highest proportion in the selected subsets for most feature selection methods, e.g., mRMRP, mRMRMI, PCC, and MI. Meanwhile, the early time response is hardly selected by those feature selection methods. For shape-based classification, the selected features of most feature selection methods (e.g., GA-SVM, mRMRP, mRMRMI, PCC, and MI) contained nearly all kinds of feature types, and the proportions of these three types are quite close. However, the LASSO, which has lower accuracy with the majority classifiers than other feature selection methods, rarely selects the intermediate time response signals.

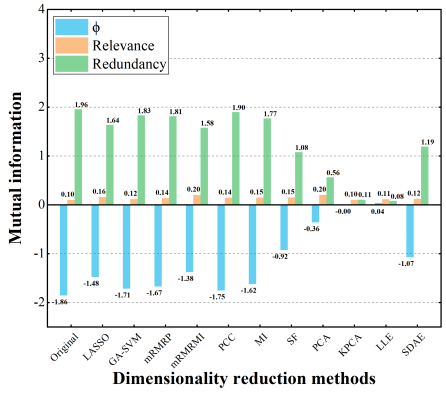

FIGURE 13. The mutual information coefficients (in columns) of feature sets obtained by different methods (in rows) in material-based classification.

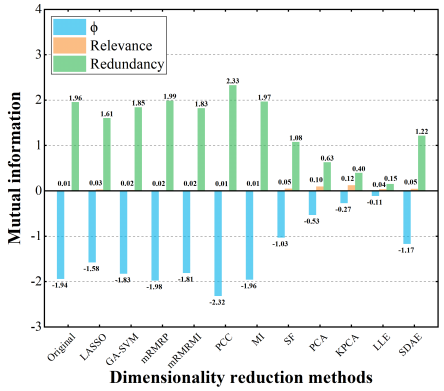

FIGURE 14. The mutual information coefficients (in columns) of feature sets obtained by different methods (in rows) in shape-based classification.

D. IMPACTS OF DIMENSIONALITY REDUCTION METHODS ON REDUNDANCY AND RELEVANCE OF INPUTS

The impacts of dimensionality reduction methods on redundancy and relevance of input are then evaluated. The *Relevance*, *Redundancy*, and ϕ metrics are utilized to quantify the quality of different data sets as shown in Fig. 13 and Fig. 14. For the material-based classification, all of the eleven dimensionality reduction methods reduce the redundancy and increase the relevance, thus, those data sets have larger values of ϕ than the original data set. Moreover, compared with feature selection methods, feature extraction methods, including SF, PCA, KPCA, LLE, and SDAE, are much more efficient in reducing the redundancy. Specifically, KPCA and LLE reduce 94.37% and 95.92% redundancy respectively. Similarly, for the shape-based classification, the redundancy of the data sets obtained by mRMRP and PCC are 1.99 and 2.33 while the redundancy of the data sets obtained by KPCA and LLE are 0.40 and 0.15. Meanwhile, KPCA and LLE have the highest relevance.

E. IMPACTS OF DIMENSIONALITY REDUCTION METHODS ON THE EFFICIENCY OF CLASSIFICATION MODELS

According to the simulation results, most of the dimensionality reduction methods reduce the redundancy of the

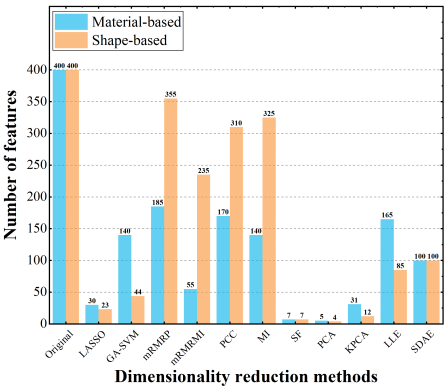

FIGURE 15. The feature number (in columns) corresponding to different dimensionality reduction methods (in rows).

data set, we further analyzed the detailed time consumption of classifier ANN which achieved the highest accuracy for material-based classification and shape-based classification. The dimensionality of inputs is an important factor affecting the complexity of classification models and the RAM storage of hardware, hence we counted the feature number corresponding to different dimensionality reduction methods as shown in Fig. 15. For the material-based classification, the dimensions of eleven data sets after dimensionality reduction are approximately reduced by 50% compared to the original data. Specifically, feature selection methods, e.g., LASSO and mRMRMI reduce the dimension of original data by 92.5% and 86.3% respectively while feature extraction methods, e.g., FS, PCA, and KPCA reduce the dimension of inputs by 98.3%, 98.8%, 92.3% respectively. Meanwhile, for the shape-based classification, feature selection methods, e.g., LASSO and GA-SVM reduce the dimension of original data by 92.5% and 89.0% while feature extraction methods FS, PCA, and KPCA reduce the dimension of inputs by 98.3%, 99.0%, 97.0%, respectively.

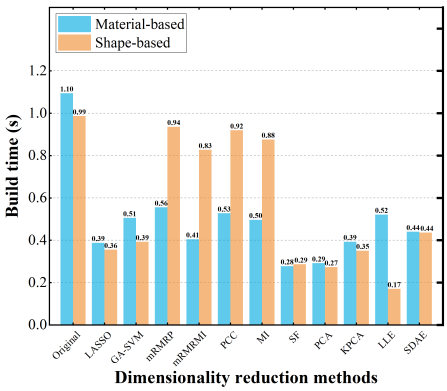

FIGURE 16. The build time consumption (in columns) of ANN with different dimensionality reduction methods (in rows).

Then we compared the impacts of dimensionality reduction methods on the efficiency of classifier ANN. Fig. 16 and Fig. 17 depict the build time and classification time

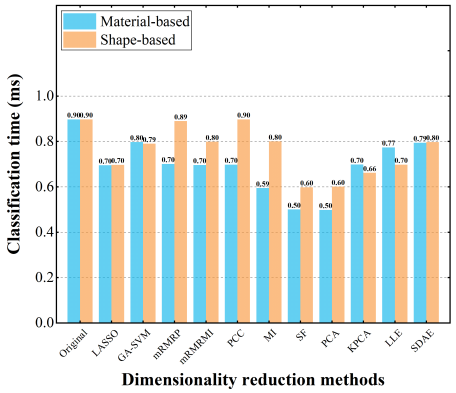

FIGURE 17. The classification time consumption (in columns) of ANN with different dimensionality reduction methods (in rows).

results respectively. As shown in Fig. 16 and Fig. 17, most dimensionality reduction methods reduced the build time and classification time of ANN, and the build time consumption far outweighed the classification time. Feature selection method LASSO obtained the shortest build time (material-based classification: 0.39s, shape-based classification: 0.36s), followed by mRMRMI (material-based classification: 0.41s, shape-based classification: 0.83s) and GA-SVM (material-based classification: 0.51s, shape-based classification: 0.39s). Feature extraction method SF obtained the shorted build time (0.28s) for material-based classification, followed by PCA (0.29s) and KPCA (0.39s). Meanwhile, feature extraction method LLE obtained the shortest build time (0.17s) for shape-based classification, followed by PCA (0.27s) and SF (0.29s).

VII. DISCUSSION

In the TEM system, the properties of underground metal targets are estimated based on high dimensional observation data generated by time-domain electromagnetic sensors. In this paper, different dimensionality reduction methods and classification models were investigated to improve the classification performance for data-based underground metal target detection. Also, other controllable variables, e.g., the number of selected features, feature type were discussed for guiding the data acquisition parameters setting of time-domain electromagnetic sensors.

Eleven dimensionality reduction methods and three frequently employed classifiers were investigated for underground metal target detection. The results showed that the KPCA feature extraction method combined with the ANN classifier achieve the highest prediction accuracy in both the material-based classification (accuracy: 0.99) and the shape-based classification (accuracy: 0.99) problems, which is higher than the classification performance in the previous study [21] (the accuracy of material-based classification: 0.79, the accuracy of shape-based classification: 0.89). Remarkably, the prediction accuracy of the ANN classifier was always superior to the other two classifiers. Noted that the ANN has been proven to be an efficient classification model previously [21], [24]. Among the eleven dimensionality reduction methods, the feature extraction method LLE exhibited superior prediction accuracy with majority classifiers. Statistic features are frequently utilized in previous studies [24], [70], and it is proved to be an efficient feature extraction method. However, it is not outstanding compared with the other dimensionality reduction methods in terms of accuracy. Although the feature selection method GA-SVM could avoid the pitfall of local optima [57], the performance of GA is depending on the parameters (e.g., population size, crossover, and mutation operators) [89]. The GA-SVM feature selection method does not perform better than filter methods (e.g., mRMRP, mRMRMI, PCC, MI). The confusion matrix results indicated that dimensionality reduction methods (e.g., LLE) can significantly improve the performance of classifiers (e.g., L-SVM, GNB).

To further investigate the effect of noise on the accuracy, we compared the average accuracy of dimensionality reduction methods under different SNRs. The results showed that the performance of classification models was sensitive to the noise level. Notably, the robustness of classification can be improved by adopting appropriate dimensionality reduction methods (e.g., LLE, KPCA, and PCA). Specifically, the feature extraction method LLE improved the robustness of machine learning classifiers in the material-based classification (SNR:30dB, improvement of average accuracy:0.17) and shape-based classification (SNR:30dB, improvement of average accuracy:0.16). Although feature selection methods(e.g., GA-SVM, mRMRMI, and LASSO) reduced the dimension of inputs, the improvement of prediction accuracy was still limited. The comparison results of selected feature number of four filter-based feature selection methods illustrated that nonlinear criteria (i.e., mRMRMI and MI) achieved the stable average accuracy with smaller selected feature number compared with linear criteria (i.e, mRMRP and PCC) which indicated that the relationship between features and label variables (shape type and material type) are nonlinear. Moreover, the average accuracy of material-based classification slightly decreased when the selected feature number exceeded 150, indicating that feature selection was an efficient strategy to improve the classification performance. Through the feature type analysis, we found that the late time response signals were most important for material-based classification although other features were necessary for shape-based classification as well. These results provided crucial guidance for the data acquisition parameters setting of time-domain electromagnetic sensors.

We proposed a data quality quantify scheme based on mutual information. The redundancy and relevance analysis results indicate that the original inputs are highly redundant and partial feature extraction methods (e.g., KPCA, LLE, PCA) exhibited outstanding ability on reducing the redundancy and increasing the relevance of data sets. Moreover, the analysis results demonstrated the poor performance of GNB since it assumed conditional independence between

every pair of features. This strong assumption of GNB led to the loss of classification accuracy, which was consistent with corresponding results in the previous study [90].

We further compared the impacts of each dimensionality reduction method on the efficiency of the ANN classifier. The results showed that dimensionality reduction methods approximately reduced 50% dimensions of inputs and reduced the build and classification time consumption of ANN classifier, e.g., compared with original inputs, feature extraction method LLE reduced the time consumption of ANN in material-based classification(reduction of build time:52%, reduction of classification time:14%) and shape-based classification (reduction of build time:82%, reduction of classification time:22%). Utilizing such a method, our machine learning strategy became more effective for the devices with limited computing capability and energy supply in the TEM system.

VIII. CONCLUSION

In this article, we investigated thirty-three classification strategies based on eleven dimensionality reduction methods and three classification models to seek out a robust and efficient data-based strategy to classify the underground metal targets of different shapes and materials based on electromagnetic induction (EMI) detection. A simulation platform was established to verify the proposed classification strategies. Among all the classification strategies explored, the ANN classifier assisted with the KPCA feature extraction method yielded the best performance in the material-based classification (accuracy:0.99) and the shape-based classification (accuracy:0.99), which is higher than the model-based methods in the previous study (the accuracy of material-based classification: 0.79, the accuracy of shape-based classification: 0.89). The results revealed that the ANN classifier was superior to the other two classifiers for the classification of the underground metal target with majority dimensionality reduction methods. A scheme based on mutual information was proposed to evaluate the redundancy and relevance of inputs, which proved the necessity of dimensionality reduction. The comparative investigation indicated that the effect of noise is non-negligible, i.e., the average accuracy of machine learning classifiers decreased rapidly with the decrease of SNR. Dimensionality reduction methods could improve the robustness of machine learning classifiers in the material-based classification (e.g., feature extraction method: LLE, improvement of average accuracy:0.17) and shape-based classification (e.g., feature extraction method: LLE, improvement of average accuracy:0.16). The efficiency analysis also indicated that dimensionality was effective in reducing the time consumption of ANN classifier, e.g., feature extraction method LLE reduced more than 50% dimensions of original inputs and reduced the build and classification time consumption in material-based classification (reduction of build time:52%, reduction of classification time:14%) and shape-based classification (reduction of build time:82%, reduction of classification time:22%). Our comparative investigation provides a robust and efficient data-based strategy (feature extraction method KPCA and classifier ANN) for underground metal target classification, which is significant for applications of underground metal target detection on portable devices with limited computing capability and energy supply. The cross-combination strategy of dimensionality reduction methods and machine learning models provides a way to find the optimal machine learning model for underground target detection. In the future, we will concentrate on optimizing the classification models to further improve the robustness of classification strategies under low SNR for underground metal target detection.

REFERENCES

[1] H. Zhang, K. Yang, Z. Yang, P. Zhang, Y. Lu, and P. Yan, "Hyperspectral mineral mapping technology applied to geology based on hymap data," in *Hyperspectral Remote Sensing Applications and Environmental Monitoring and Safety Testing Technology*, vol. 10156. Bellingham, WA, USA: SPIE, 2016, Art. no. 101560Y.

[2] J. Li, T. Guo, H. Leung, H. Xu, L. Liu, B. Wang, and Y. Liu, "Locating underground pipe using wideband chaotic ground penetrating radar," *Sensors*, vol. 19, no. 13, p. 2913, Jul. 2019.

[3] F. Shubitidze, B. E. Barrowes, J. B. Sigman, K. O'Neill, and I. Shamatava, "UXO classification procedures applied to advanced EMI sensors and models," in *Proc. 21st Int. Seminar/Workshop Direct Inverse Problems Electromagn. Acoustic Wave Theory (DIPED)*, Sep. 2016, pp. 173–177.

[4] T. N. Ibrahim Alrumaih, "The construction of a robotic vehicle metal detector as a tool for searching archaeology sites," in *Proc. 1st Int. Conf. Comput. Appl. Inf. Secur. (ICCAIS)*, Apr. 2018, pp. 1–6.

[5] P. Mlambo, H. Dera, E. Chiweshe, and E. Jonathan, "Inductive metal detectors and the design of prospecting robots: A possibility," in *Proc. EAI Int. Conf. Res., Innov. Develop. Afr.*, 2018, p. 10.

[6] Y. Wan, Z. Wang, P. Wang, C. Zhang, S. Duan, and N. Li, "A comparative study of inversion optimization algorithms for underground metal target detection," *IEEE Access*, vol. 8, pp. 126401–126413, 2020.

[7] Y. Wu, Z. Li, X. Qu, and T. Zhang, "Combined infrared simulation and pix2pix model for underground target detection," *Proc. SPIE*, vol. 11429, Feb. 2020, Art. no. 114290M.

[8] T. Zhang, M. Wenxuan, C. Lu, H. Longwei, W. Yuehuan, N. Sang, W. Yang, and H. Zhu, "Method for infrared imaging detection and positioning of underground tubular facility in plane terrain," U.S. Patent 10,365,399, Jul. 30, 2019.

[9] H. Harkat, A. E. Ruano, M. G. Ruano, and S. D. Bennani, "GPR target detection using a neural network classifier designed by a multi-objective genetic algorithm," *Appl. Soft Comput.*, vol. 79, pp. 310–325, Jun. 2019.

[10] W. Li, H. Li, K. Lu, H. Cui, and X. Li, "Multi-scale target detection for underground spaces with transient electromagnetics based on differential pulse scanning," *J. Environ. Eng. Geophys.*, vol. 24, no. 4, pp. 593–607, Dec. 2019.

[11] S. Chen, S. Zhang, J. Zhu, and X. Luan, "Accurate measurement of characteristic response for unexploded ordnance with transient electromagnetic system," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1728–1736, Apr. 2020.

[12] L.-P. Song, D. W. Oldenburg, L. R. Pasion, S. D. Billings, and L. Beran, "Temporal orthogonal projection inversion for EMI sensing of UXO," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 1061–1072, Feb. 2015.

[13] C.-C. Chen, M. B. Higgins, K. O'Neill, and R. Detsch, "Ultrawide-bandwidth fully-polarimetric ground penetrating radar classification of subsurface unexploded ordnance," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 6, pp. 1221–1230, Jun. 2001.

[14] X. Núñez-Nieto, M. Solla, P. Gómez-Pérez, and H. Lorenzo, "Signal-to-noise ratio dependence on ground penetrating radar antenna frequency in the field of landmine and UXO detection," *Measurement*, vol. 73, pp. 24–32, Sep. 2015.

[15] S. D. Billings, L. R. Pasion, L. Beran, N. Lhomme, L.-P. Song, D. W. Oldenburg, K. Kingdon, D. Sinex, and J. Jacobson, "Unexploded ordnance discrimination using magnetic and electromagnetic sensors: Case study from a former military site," *Geophysics*, vol. 75, no. 3, pp. B103–B114, May 2010.

[16] F. Shubitidze, K. O'Neill, I. Shamatava, K. Sun, and K. D. Paulsen, "Fast and accurate calculation of physically complete EMI response by a heterogeneous metallic object," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 8, pp. 1736–1750, Aug. 2005.

[17] D. Ambrus, D. Vasic, and V. Bilas, "Robust estimation of metal target shape using time-domain electromagnetic induction data," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 4, pp. 795–807, Apr. 2016.

[18] S. Chen, S. Zhang, H. Jiang, and J. Zhu, "Location and characterization of unexploded ordnance-like targets with a portable transient electromagnetic system," *IEEE Access*, vol. 8, pp. 150174–150185, 2020.

[19] P. Zhang, X. Guo, N. Muhammat, and X. Wang, "Research on probing and predicting the diameter of an underground pipeline by GPR during an operation period," *Tunnelling Underground Space Technol.*, vol. 58, pp. 99–108, Sep. 2016.

[20] I. Shamatava, G. Schultz, and F. Shubitidze, "Accessing UXO classification technologies at a challenging live-UXO site," in *Proc. 23rd Int. Seminar/Workshop Direct Inverse Problems Electromagn. Acoustic Wave Theory (DIPED)*, Sep. 2018, pp. 24–27.

[21] S. Duan, Y. Li, Y. Wan, P. Wang, Z. Wang, and N. Li, "Sensitivity analysis and classification algorithms comparison for underground target detection," *IEEE Access*, vol. 7, pp. 116227–116246, 2019.

[22] I. Mitiche, M. David Jenkins, P. Boreham, A. Nesbitt, B. G. Stewart, and G. Morison, "Deep residual neural network for EMI event classification using bispectrum representations," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2018, pp. 186–190.

[23] L. Carin, L. Kennedy, X. Zhu, and N. Dasgupta, "2009 Estcp Uxo discrimination study, San Luis Obispo, Ca," Signal Innovations Group Inc., Durham, NC, USA, Tech. Rep. MR-200501, 2010.

[24] S. N. A. M. Kanafiah, A. A. Firdaus, N. F. Jefri, N. N. Karim, N. S. Khalid, I. I. Ismail, M. J. M. Ridzuan, M. A. Ismail, and M. R. Ahmad, "Fundamental shape discrimination of underground metal object through one-axis ground penetrating radar (GPR) scan," *J. Telecommun., Electron. Comput. Eng.*, vol. 10, nos. 1–13, pp. 43–47, 2018.

[25] L. Pasion, L. Beran, K. Kingdon, and S. Billings, "Inversion and classification using the point dipole model: Practical experiences from munitions response demonstrations," in *Proc. SEG Tech. Program Expanded Abstr.* Tulsa, OK, USA: Society of Exploration Geophysicists, 2011, pp. 3758–3762.

[26] J. Makkonen, L. A. Marsh, J. Vihonen, A. Visa, A. Järvi, and A. J. Peyton, "Classification of metallic targets using a single frequency component of the magnetic polarisability tensor," *J. Phys., Conf. Ser.*, vol. 450, Jun. 2013, Art. no. 012038.

[27] J. Sigman, Y. Wang, K. O'Neill, B. Barrowes, and F. Shubitidze, "An expert-free technique for live site Uxo target classification," in *Proc. Symp. Appl. Geophys. Eng. Environ. Problems*, Mar. 2014, pp. 456–463.

[28] M. P. Bray and C. A. Link, "Learning machine identification of ferromagnetic UXO using magnetometry," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 2, pp. 835–844, Feb. 2015.

[29] S. Lameri, F. Lombardi, P. Bestagini, M. Lualdi, and S. Tubaro, "Landmine detection from GPR data using convolutional neural networks," in *Proc. 25th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2017, pp. 508–512.

[30] H. Ammari, J. Chen, Z. Chen, D. Volkov, and H. Wang, "Detection and classification from electromagnetic induction data," *J. Comput. Phys.*, vol. 301, pp. 201–217, Nov. 2015.

[31] K. N. Kappler and E. Gasperikova, "A hybrid method for UXO vs. non-UXO discriminationhybrid method for uxo discrimination," *J. Environ. Eng. Geophys.*, vol. 16, no. 4, pp. 177–189, 2011.

[32] E. Lavely, R. Grimm, and P. Weichman, "Detection and discrimination of landmines and UXO," in *Proc. Sens. Manag. Environ. IEEE Int. Geosci. Remote Sensing. Symp. (IGARSS)*, vol. 1, Jul. 1998, pp. 514–516.

[33] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artif. Intell.*, vol. 97, nos. 1–2, pp. 245–271, Dec. 1997.

[34] X. Xu, T. Liang, J. Zhu, D. Zheng, and T. Sun, "Review of classical dimensionality reduction and sample selection methods for large-scale data processing," *Neurocomputing*, vol. 328, pp. 5–15, Feb. 2019.

[35] M. Laib and M. Kanevski, "A new algorithm for redundancy minimisation in geo-environmental data," *Comput. Geosci.*, vol. 133, Dec. 2019, Art. no. 104328.

[36] J. A. Lee and M. Verleysen, *Nonlinear Dimensionality Reduction*. Cham, Switzerland: Springer, 2007.

[37] W. Pindah, S. Nordin, A. Seman, and M. S. Mohamed Said, "Review of dimensionality reduction techniques using clustering algorithm in reconstruction of gene regulatory networks," in *Proc. Int. Conf. Comput., Commun., Control Technol. (I4CT)*, Apr. 2015, pp. 172–176.

[38] S. Pölsterl, S. Conjeti, N. Navab, and A. Katouzian, "Survival analysis for high-dimensional, heterogeneous medical data: Exploring feature extraction as an alternative to feature selection," *Artif. Intell. Med.*, vol. 72, pp. 1–11, Sep. 2016.

[39] M. Aljanabi, M. A. Ismail, and V. Mezhuyev, "Improved TLBO-JAYA algorithm for subset feature selection and parameter optimisation in intrusion detection system," *Complexity*, vol. 2020, pp. 1–18, May 2020.

[40] A. H. Ali, "Fuzzy generalized Hebbian algorithm for large-scale intrusion detection system," *Int. J. Integr. Eng.*, vol. 12, no. 1, pp. 81–90, 2020.

[41] K. Krueger, W. R. Scott, and J. H. McClellan, "Location and orientation estimation of buried targets using electromagnetic induction sensors," in *Proc. 17th Detection Sens. Mines, Explosive Objects, Obscured Targets*, May 2012, p. 83570.

[42] H. R. Ibraheem, Z. F. Hussain, S. M. Ali, M. Aljanabi, M. A. Mohammed, and T. Sutikno, "A new model for large dataset dimensionality reduction based on teaching learning-based optimization and logistic regression," *Telkomnika*, vol. 18, no. 3, p. 1688, Jun. 2020.

[43] J. P. Fernandez, B. E. Barrowes, T. M. Grzegorczyk, N. Lhomme, K. O'Neill, and F. Shubitidze, "A man-portable vector sensor for identification of unexploded ordnance," *IEEE Sensors J.*, vol. 11, no. 10, pp. 2542–2555, Oct. 2011.

[44] T. M. Grzegorczyk, B. E. Barrowes, F. Shubitidze, J. P. Fernandez, and K. O'Neill, "Simultaneous identification of multiple unexploded ordnance using electromagnetic induction sensors," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 7, pp. 2507–2517, Jul. 2011.

[45] F. Shubitidze, K. O'Neill, B. E. Barrowes, I. Shamatava, J. P. Fernández, K. Sun, and K. D. Paulsen, "Application of the normalized surface magnetic charge model to UXO discrimination in cases with overlapping signals," *J. Appl. Geophys.*, vol. 61, nos. 3–4, pp. 292–303, Mar. 2007.

[46] F. Shubitidze, J. P. Fernandez, B. E. Barrowes, I. Shamatava, A. Bijamov, K. O'Neill, and D. Karkashadze, "The orthonormalized volume magnetic source model for discrimination of unexploded ordnance," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4658–4670, Aug. 2014.

[47] Y. Tao, W. Yin, W. Zhang, Y. Zhao, C. Ktistis, and A. J. Peyton, "A very-low-frequency electromagnetic inductive sensor system for workpiece recognition using the magnetic polarizability tensor," *IEEE Sensors J.*, vol. 17, no. 9, pp. 2703–2712, May 2017.

[48] L. R. Pasion, "Inversion of time domain electromagnetic data for the detection of unexploded ordnance," Ph.D. dissertation, Dept. Graduate Student, Univ. British Columbia, Vancouver, BC, Canada, 2007.

[49] P. D. Ledger and W. R. B. Lionheart, "An explicit formula for the magnetic polarizability tensor for object characterization," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3520–3533, Jun. 2018.

[50] F. Shubitidze, "A complex approach to UXO discrimination: Combining advanced EMI forward models and statistical signal processing," Sky Res. Inc., Ashland, OR, USA, Tech. Rep. MR-1572, 2012.

[51] J. T. Smith, H. F. Morrison, and A. Becker, "Parametric forms and the inductive response of a permeable conducting sphere," *J. Environ. Eng. Geophys.*, vol. 9, no. 4, pp. 213–216, Dec. 2004.

[52] J. T. Smith and H. F. Morrison, "Approximating spheroid inductive responses using spheres," *Geophysics*, vol. 71, no. 2, pp. G21–G25, Mar. 2006.

[53] B. S. C. Wade, S. H. Joshi, B. A. Gutman, and P. M. Thompson, "Machine learning on high dimensional shape data from subcortical brain surfaces: A comparison of feature selection and classification methods," *Pattern Recognit.*, vol. 63, pp. 731–739, Mar. 2017.

[54] A. Jović, K. Brkić, and N. Bogunović, "A review of feature selection methods with applications," in *Proc. 38th Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, May 2015, pp. 1200–1205.

[55] L. D. Whitley, J. R. Beveridge, C. Guerra-Salcedo, and C. R. Graves, "Messy genetic algorithms for subset feature selection," in *Proc. ICGA*, 1997, pp. 568–575.

[56] J. Yang and V. Honavar, "Feature subset selection using a genetic algorithm," in *Feature Extraction, Construction and Selection*. Cham, Switzerland: Springer, 1998, pp. 117–136.

[57] D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut, "Comparison of linear, nonlinear, and feature selection methods for eeg signal classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 141–144, Jun. 2003.
[58] H. Alkharusi, "Categorical variables in regression analysis: A comparison of dummy and effect coding," *Int. J. Edu.*, vol. 4, no. 2, p. 202, Jun. 2012.
[59] L. T. Vinh, S. Lee, Y.-T. Park, and B. J. d'Auriol, "A novel feature selection method based on normalized mutual information," *Int. J. Speech Technol.*, vol. 37, no. 1, pp. 100–120, Jul. 2012.
[60] Y. Liao, Y. Weng, and R. Rajagopal, "Urban distribution grid topology reconstruction via lasso," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Jul. 2016, pp. 1–5.
[61] J. Quilty, J. Adamowski, B. Khalil, and M. Rathinasamy, "Bootstrap rank-ordered conditional mutual information (broCMI): A nonlinear input variable selection method for water resources modeling," *Water Resour. Res.*, vol. 52, no. 3, pp. 2299–2326, Mar. 2016.
[62] A. Senawi, H.-L. Wei, and S. A. Billings, "A new maximum relevance-minimum multicollinearity (MRmMC) method for feature selection and ranking," *Pattern Recognit.*, vol. 67, pp. 47–61, Jul. 2017.
[63] K. Ren, W. Fang, J. Qu, X. Zhang, and X. Shi, "Comparison of eight filter-based feature selection methods for monthly streamflow forecasting–three case studies on CAMELS data sets," *J. Hydrol.*, vol. 586, Jul. 2020, Art. no. 124897.
[64] H. Zou, K. Tuncali, and S. Silverman, "Correlation and simple linear regression," *Radiology*, vol. 227, no. 3, pp. 617–628, 2003.
[65] R. J. May, H. R. Maier, G. C. Dandy, and T. M. K. G. Fernando, "Non-linear variable selection for artificial neural networks using partial mutual information," *Environ. Model. Softw.*, vol. 23, nos. 10–11, pp. 1312–1326, Oct. 2008.
[66] L. Chen, L. Ye, V. Singh, J. Zhou, and S. Guo, "Determination of input for artificial neural networks for flood forecasting using the copula entropy method," *J. Hydrologic Eng.*, vol. 19, no. 11, Nov. 2014, Art. no. 04014021.
[67] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.
[68] A. Tsanas, M. A. Little, and P. E. McSharry, "A methodology for the analysis of medical data," in *Handbook of Systems and Complexity in Health*. Cham, Switzerland: Springer, 2013, pp. 113–125.
[69] C. Altın and O. Er, "Comparison of different time and frequency domain feature extraction methods on elbow gesture's EMG," *Eur. J. Interdiscipl. Stud.*, vol. 2, no. 3, pp. 35–44, 2016.
[70] S. Shihab and W. Al-Nuaimy, "A comparison of segmentation techniques for target extraction in ground-penetrating radar data," *Near Surf. Geophys.*, vol. 2, no. 1, pp. 49–57, Feb. 2004.
[71] L. Van Der Maaten, E. Postma, and J. Van den Herik, "Dimensionality reduction: A comparative," *J. Mach. Learn. Res.*, vol. 10, nos. 66–71, p. 13, 2009.
[72] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educ. Psychol.*, vol. 24, no. 6, p. 417, 1933.
[73] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *London, Edinburgh, Dublin Phil. Mag. J. Sci.*, vol. 2, no. 11, pp. 559–572, Nov. 1901.
[74] G. H. Dunteman, *Principal Component Analysis*, no. 69. Newbury Park, CA, USA: Sage, 1989.
[75] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998.
[76] S. T. Roweis, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
[77] G. E. Hinton, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.
[78] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Mach. Learn. (ICML)*, 2008, pp. 1096–1103.
[79] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, and L. Bottou, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371–3408, 2010.
[80] B. Hammer, D. Nebel, M. Riedel, and T. Villmann, "Generative versus discriminative prototype based classification," in *Advances in Self-Organizing Maps, Learning Vector Quantization, Clustering and Data Visualization*. Cham, Switzerland: Springer, 2014, pp. 123–132.
[81] H. Zhang, "Exploring conditions for the optimality of Naïve Bayes," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 19, no. 2, pp. 183–198, Mar. 2005.
[82] K. Crammer and Y. Singer, "On the algorithmic implementation of multiclass kernel-based vector machines," *J. Mach. Learn. Res.*, vol. 2, pp. 265–292, Mar. 2001.
[83] M. van Gerven and S. Bohte, "Artificial neural networks as models of neural information processing," *Frontiers Comput. Neurosci.*, vol. 11, p. 114, Dec. 2017.
[84] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Math. Program.*, vol. 45, nos. 1–3, pp. 503–528, Aug. 1989.
[85] M. N. Halgamuge, "Machine learning for bioelectromagnetics: Prediction model using data of weak radiofrequency radiation effect on plants," *Mach. Learn.*, vol. 8, no. 11, pp. 223–235, 2017.
[86] N. Kriegeskorte, W. K. Simmons, P. S. Bellgowan, and C. I. Baker, "Circular analysis in systems neuroscience: The dangers of double dipping," *Nature Neurosci.*, vol. 12, no. 5, p. 535, 2009.
[87] K. M. Ting, "Confusion matrix," in *Encyclopedia of Machine Learning and Data Mining*, vol. 260. Boston, MA, USA: Springer, 2017.
[88] P. Sun, D. Wang, V. C. Mok, and L. Shi, "Comparison of feature selection methods and machine learning classifiers for radiomics analysis in glioma grading," *IEEE Access*, vol. 7, pp. 102010–102020, 2019.
[89] M. Alsajri, M. A. Ismail, and S. Abdul-Baqi, "A review on the recent application of Jaya optimization algorithm," in *Proc. 1st Annu. Int. Conf. Inf. Sci. (AiCIS)*, Nov. 2018, pp. 129–132.
[90] S. A.-B. Salman, A.-H. A. Salih, A. H. Ali, M. K. Khaleel, and M. A. Mohammed, "A new model for iris classification based on Naïve Bayes grid parameters optimization," *Int. J. Sci., Basic Appl. Res.*, vol. 40, no. 2, pp. 150–155, 2018.

YADONG WAN received the B.E. and Ph.D. degrees from the University of Science and Technology Beijing in 2003 and 2010, respectively. He was a Visiting Scholar with the Worcester Polytechnic Institute in area of wireless location from 2012 to 2013. His current research interests include underground target detection, industrial wireless sensor network, signal processing, and data science of materials genome initiative.

TONG LI received the B.E. degree from the University of Science and Technology Beijing, China, in 2019, where he is currently pursuing the master's degree with the School of Computer and Communication Engineering. His current research interest is underground target classification.

PENG WANG received the B.E. and Ph.D. degrees from the University of Science and Technology Beijing in 2012 and 2018, respectively. His current research interests include underground target detection, low-frequency wireless communication, wireless sensor network, adaptive signal processing, and near-field electromagnetic localization.

SHIHONG DUAN received the B.E. and Ph.D. degrees in pattern recognition and computer science from the University of Science and Technology Beijing (USTB), China, in 1998 and 2012, respectively. She was an Assistant Professor with the School of Computer and Communication Engineering, USTB, from 1998 to 2013, where she has been an Associate Professor since 2013. From 2014 to 2015, she was a Visiting Scholar with the Center for Wireless Information Network Studies, Department of Electrical and Computer Engineering, Worcester Polytechnic Institute. Her research interests include wireless channel study, human activity recognition, mobile robotics, and underground target detection.

CHAO ZHANG received the master's degree in physics and the Ph.D. degree in material science and engineering from the University of Science and Technology Beijing in 2003 and 2018, respectively. He has been involved in the study of magnetic sensors since 2006. His current research interest is weak magnetic field detection system.

NA LI received the B.E. and M.Sc. degrees from the University of Science and Technology Beijing, China, in 2001 and 2006, respectively. She joined the University of Science and Technology Beijing in 2006. Her current research interests include multivariate statistical analysis, statistical learning theory, and machine learning.