Machine Learning Based Distinguishing between Ferroelectric and Non-Ferroelectric Polarization–Electric Field Hysteresis Loops

Qicheng Huang, Zhen Fan,* Lanqing Hong, Shengliang Cheng, Zhengwei Tan, Guo Tian, Deyang Chen, Zhipeng Hou, Minghui Qin, Min Zeng, Xubing Lu, Guofu Zhou, Xingsen Gao, and Jun-Ming Liu

The polarization–electric field (P–E) hysteresis loop is one of the most important criteria for identifying ferroelectricity. However, a P–E loop with apparent hysteresis window can be generated from non-ferroelectric sources such as leakage current. So far distinguishing between ferroelectric and non-ferroelectric loops is still performed in a manual way, which can be error prone and time consuming, particularly when the loops are not easily distinguishable and the number of loops to be identified is large. Here, two machine learning (ML) approaches are developed, one using the polarization values along the P–E loops as the input dataset (termed as "value-based" approach) and the other using the loop images as the input dataset (termed as "image-based" approach), to identify the P–E loops as ferroelectric or non-ferroelectric. The value- and image-based ML approaches achieve identification accuracies as high as 93.08% and 87.42%, respectively. In addition, it is tested that both approaches complete an identification of about 160 loops in very short time ($\approx$1.0 s). The high accuracy and efficiency therefore demonstrate that the ML approaches significantly outperform the manual way for distinguishing ferroelectric from non-ferroelectric P–E loops, which may greatly facilitate the research on ferroelectrics.

1. Introduction

In recent years, the rapidly advancing machine learning (ML) technique has risen to prominence in the field of materials science, because it can greatly accelerate the discovery of new materials.[1–5] For example, ML has been successfully employed to predict high-performance organic photovoltaic materials,[1] lead-free hybrid organic–inorganic perovskites with high stability and proper bandgaps,[6] piezoelectrics with large electrostrains,[7] ferroelectrics with high curie temperatures,[8] and so on. While these ML applications mainly focused on the prediction of materials with targeted properties, much less attention has been paid to using ML to examine the experimentally measured data of material properties. Indeed, the ML approach can significantly outperform the manual way for the data examination, particularly in cases where high accuracy and efficiency are required. For example, the measured data may contain artifacts and sometimes it is very difficult to distinguish the true signals from artifacts via human eyes. In addition, when an enormous number of experiments are carried out (e.g., in industry), the number of measured data becomes extremely large and thus the manual examination of those data is inefficient and even impossible. All these are no longer a problem for the ML-based data examination. Moreover, the reliability of data after

Q. Huang, Prof. Z. Fan, S. Cheng, Z. Tan, Prof. G. Tian, Prof. D. Chen, Prof. Z. Hou, Prof. M. Qin, Prof. M. Zeng, Prof. X. Lu, Prof. X. Gao
Institute for Advanced Materials, South China Academy of Advanced Optoelectronics
South China Normal University
Guangzhou 510006, China
E-mail: fanzhen@m.scnu.edu.cn
Q. Huang, Prof. Z. Fan, Prof. G. Zhou
Guangdong Provincial Key Laboratory of Optical Information Materials and Technology, South China Academy of Advanced Optoelectronics
South China Normal University
Guangzhou 510006, China

Dr. L. Hong
Department of Industrial Systems Engineering and Management
National University of Singapore
Singapore 117576, Singapore
Prof. G. Zhou
National Center for International Research on Green Optoelectronics
South China Normal University
Guangzhou 510006, China
Prof. J.-M. Liu
Laboratory of Solid State Microstructures and Innovation Center of Advanced Microstructures
Nanjing University
Nanjing 210093, China

**ADVANCED
SCIENCE NEWS**
www.advancedsciencenews.com

**ADVANCED
THEORY AND
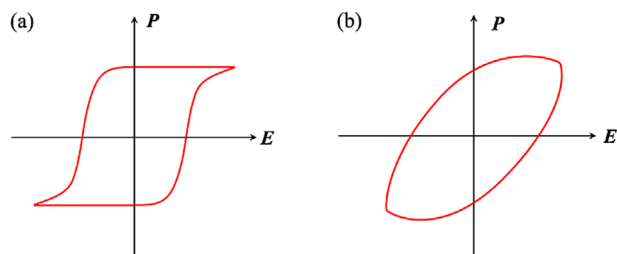SIMULATIONS**
www.advtheorysimul.com

Figure 1. Schematics of typical P–E hysteresis loops for a) an ideal ferroelectric insulator and b) a non-ferroelectric lossy dielectric.

the examination and purification (excluding the artifacts) can be improved, giving rise to a higher accuracy of an ML model built on these data for the prediction of material properties. In this regard, the ML-based data examination can in turn serve the ML-based prediction of material properties.

A prominent example which requires a careful data examination is the polarization–electric field (P–E) hysteresis loop, one of the most important criteria for identifying ferroelectricity[9] (note: ferroelectricity is a property of certain materials having spontaneous polarization that can be reversed by applied electric field). In a typical P–E loop measurement, the current rather than the polarization is directly measured and the integration of current over time gives the polarization.[10] For an ideal ferroelectric insulator, only the polarization switching current is measured, yielding a P–E loop which well captures the hysteretic evolution of polarization as a function of electric field, as shown in **Figure 1**a. However, such loop with an apparent hysteresis window can also be observed for a non-ferroelectric lossy dielectric due to the leakage current, as shown in Figure 1b. While the experts were familiar with the artifact from the leakage current, there were still some researchers who may be new to this field misinterpreting this artifact as a ferroelectric loop,[11–17] even after the publication of the famous article "Ferroelectrics Go Bananas" by Scott.[18] Such situation can be blamed on the fact that so far distinguishing between ferroelectric loops and artifacts still relies on the experience of the researchers, greatly limiting the accuracy of the distinguishment. To change this situation, the ML technique can be used to examine the measured P–E loops and determine whether they are ferroelectric or not, thus avoiding human mistakes. In addition, the ML-based P–E loop examination can be very fast, which is particularly useful for the laboratories and factories where an enormous number of P–E loops are measured.

Here, we develop two ML approaches to identify a measured P–E loop as ferroelectric or non-ferroelectric. The first approach uses arrays of polarization values along the P–E loops as the input dataset, which is termed as the "value-based" approach. Five different ML algorithms are used for the identification, among which the support vector machine (SVM) algorithm achieves the highest accuracy of 93.08%. The second approach constructs the input dataset by using 50×50 grayscale images of the P–E loops, which is termed as the "image-based" approach.[19] An artificial neural network (ANN) algorithm is used for the identification and achieves an accuracy of 87.42%. The obtained high accuracies demonstrate that our value- and image-based ML approaches both work well for the identification of P–E loops. Besides,

the efficiencies of identification are also assessed for our ML approaches.

2. Results and Discussion

The value- and image-based ML approaches are schematically illustrated in **Figure 2**a,b, respectively. The input datasets for these two approaches come from the same set of P–E loops measured for 633 $Pb(Zr_{0.2}Ti_{0.8})O_3$ (PZT) thin-film samples prepared in different conditions. However, the expressions of the P–E loops are different. For the value-based approach, the polarization and electric field values are normalized to [−1,1], and ($E_i^{(j)}$, $P_i^{(j)}$) represents the ith data point in the jth P–E loop ($i = 1, 2,..., 601$; $j = 1, 2,..., 633$; see Figure 2a). Because E_i always adopts the fixed value ($E_1 = 0,..., E_{151} = 1,..., E_{301} = 0,..., E_{451} = −1,..., E_{601} = 0$) while P_i can have different values for different P–E loops, P_i is therefore considered as a feature. Consequently, an array of $P_i^{(j)}$ values with i varying from 1 to 601 represents a sample, that is, the jth sample. For the image-based approach, the P–E loops are plotted as 50×50 grayscale images, and each image represents a sample (see Figure 2b). An image's pixel values are directly used as features and thus the total number of features is $50 \times 50 = 2500$.

A P–E loop is labeled as "1" (ferroelectric) if the polarization switching current peak is present in its corresponding current–electric field (I–E) loop (Figure S1a,b, Supporting Information); otherwise, the P–E loop is labeled as "0" (non-ferroelectric; see Figure S1c,d, Supporting Information). Although in some cases a P–E loop is distorted by the leakage current with only one polarization switching current peak existing in its corresponding I–E loop, such P–E loop is still labeled as "1." Another thing worth mentioning is that a non-ferroelectric loop (the loop labeled as "0") does not mean that the material is non-ferroelectric. Instead, it means that the formation of the loop is due to non-ferroelectric factors, like the leakage current. For example, PZT, a prototype ferroelectric material, can exhibit a non-ferroelectric loop if the sample quality is very poor (too leaky). After labeling, there are 372 ferroelectric loops and 261 non-ferroelectric loops in our dataset.

The dataset containing 633 samples is divided into a training_val set (75%) and a test set (25%) through the stratified random sampling. The training_val set is further divided into a training set (75% × 75% = 56%) and a validation set (75% × 25% = 19%), which are used in the training procedure for training a model and evaluating the hyper-parameters of this model, respectively. To optimize the hyper-parameters, the Bayesian optimization and fivefold cross validation are used in the value-based approach (Figure 2a) while the Bayesian optimization and hold-out validation are used in the image-based approach (Figure 2b). After obtaining the trained model with the optimal hyper-parameters, the prediction performance of the model is assessed using the test set.

Models in the value-based approach are developed based on five different ML algorithms (Figure 2a): SVM,[20] ANN,[21] random forest (RF),[22] gradient boosted trees (GBT),[23] and extra trees (ET).[24] In the image-based approach (Figure 2b), only ANN is used because it is perhaps the most popular algorithm for the image recognition.[25] Besides, the principle component analysis
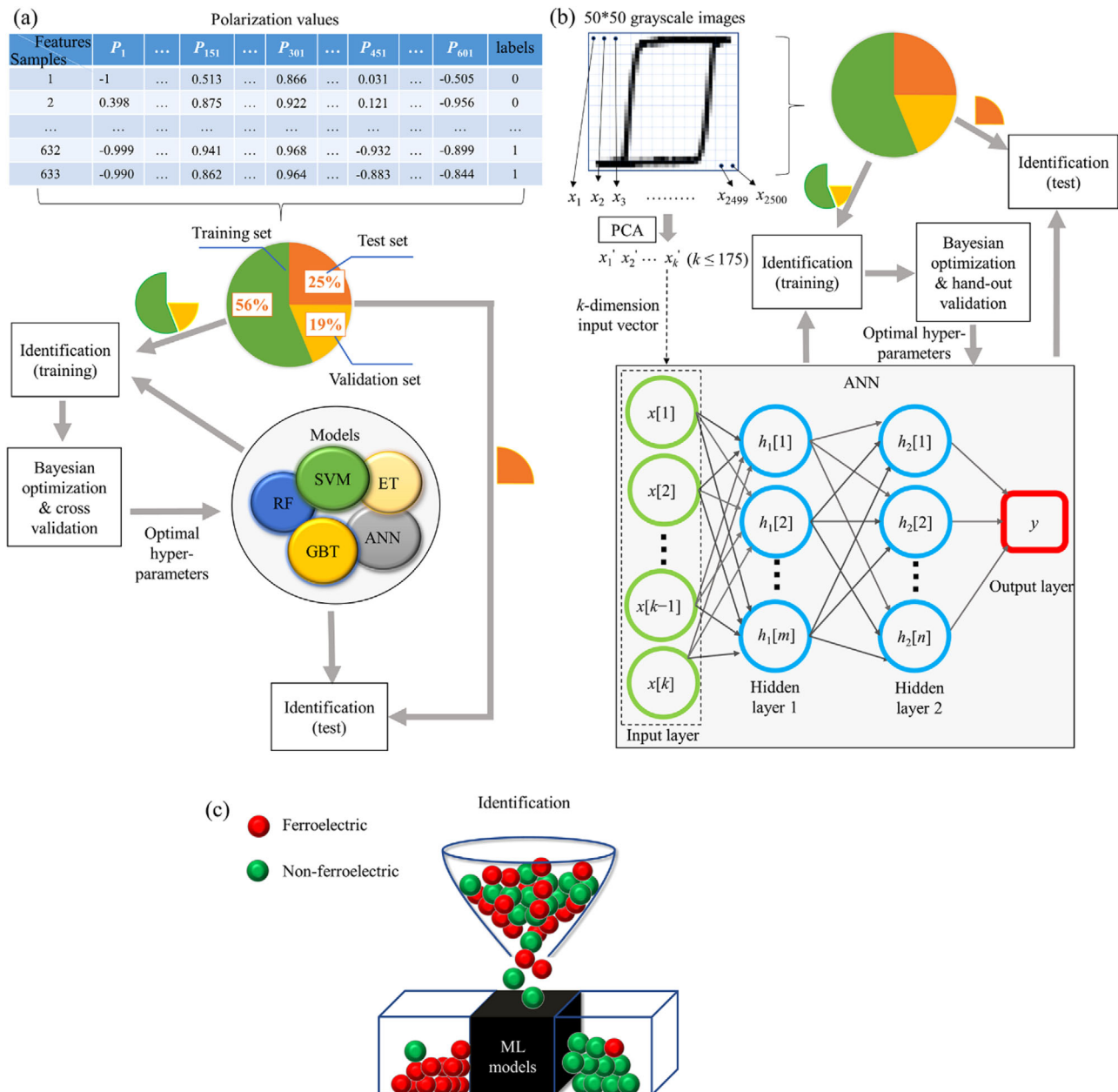
**ADVANCED
SCIENCE NEWS**
www.advancedsciencenews.com

**ADVANCED
THEORY AND
SIMULATIONS**
www.advtheorysimul.com

Figure 2. Workflow charts for the a) value-based and b) image-based ML approaches. In (a), the dataset consists of polarization values and five models including RF, GBT, ET, SVM, and ANN are used; whereas, in (b), the dataset contains the 50 × 50 grayscale images of the P–E loops, and only the ANN model is used. c) Schematic diagram vividly showing the classification of P–E loops as ferroelectric or non-ferroelectric by using an ML model.

(PCA)[26] is also employed to reduce the dimension of the feature space from 2500 to 175 or even lower.

Let us first focus on the performance of the value-based approach. The testing results of the five models used in this approach are presented in **Figure 3**a. The accuracies of all five models are around 90%, among which the highest accuracy of 93.08% is achieved by the SVM model. Figure 3b shows the receiver operating characteristic (ROC) curves[27] of the five models, which all exhibit large area under curve (AUC) values (>0.98). Note that a ROC curve is a plot of the true positive rate (TPR; the ratio of the number of loops which are correctly

identified as "1" to that of truly ferroelectric loops) against the false positive rate (FPR; the ratio of the number of loops which are incorrectly identified as "1" to that of truly non-ferroelectric loops), and an AUC value equaling 1 means that all the truly ferroelectric and non-ferroelectric loops are correctly identified as "1" and "0," respectively. The largest AUC value (0.9924) is achieved by the SVM model, consistent with its highest accuracy. The high accuracies and large AUC values demonstrate that all five models in the value-based approach have outstanding performance in predicting whether a P–E loop is ferroelectric or non-ferroelectric.
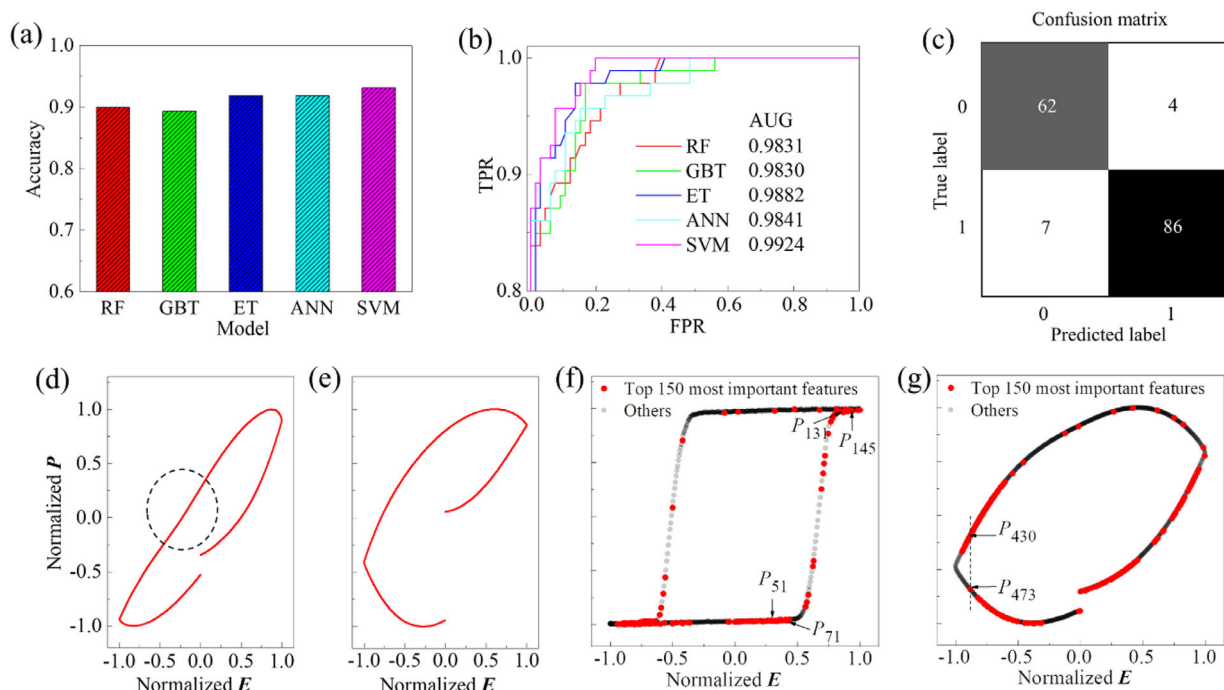
Figure 3. Testing performance of the value-based ML approach. Comparisons of the a) classification accuracy and b) ROC curve among RF, GBT, ET, ANN, and SVM models. In (b), TPR and FPR indicate true positive rate and false positive rate, respectively. c) Confusion matrix for the SVM model. Typical d) ferroelectric and e) non-ferroelectric P–E loops which are misclassified as "0" and "1," respectively. In (d), the S-shaped region is indicated by the dotted circle. Typical f) ferroelectric and g) non-ferroelectric P–E loops with top 150 most important features highlighted in red dots.

A confusion analysis is then performed for the SVM model. Figure 3c shows the confusion matrix[28] containing the counts of the predicted labels versus the true labels. The SVM model correctly identifies most truly ferroelectric (86 out of 93) and non-ferroelectric (62 out of 66) loops as "1" and "0," respectively. However, it also misidentifies small numbers of truly ferroelectric (7 out of 93) and non-ferroelectric (4 out of 66) loops as "0" and "1," respectively. The misidentified P–E loops are specifically analyzed, as exemplified in Figure 3d,e. Figure 3d presents a ferroelectric P–E loop which is misidentified as non-ferroelectric. This loop is distorted due to the leakage current, and only an obscure S-shaped region implying the polarization switching is observed in the negative branch of the loop. In addition, the polarization switching current peak in its corresponding I–E loop is also very weak (Figure S2a,b, Supporting Information). This loop is therefore very confusing and can be easily identified as a non-ferroelectric loop. Figure 3e shows a non-ferroelectric P–E loop which is misidentified as ferroelectric. This loop is also very misleading because it seems to be S-shaped. However, negligible polarization switching current peak is observed in its corresponding I–E loop (Figure S2c,d, Supporting Information), suggesting that this P–E loop is indeed non-ferroelectric. Furthermore, the predicted probability of the loop in Figure 3d for being ferroelectric is 41.68%, while that of the loop in Figure 3e for being non-ferroelectric is 36.31%. These predicted probabilities are not far from 50%, indicating that these misidentifications should not be simply regarded as model failures, but rather as indicators that the loops are very confusing. Note that the probability of 50% means the boundary between the ferroelectric and non-ferroelectric classes. The P–E

loops which are predicted to be near this boundary often do not have distinct ferroelectric or non-ferroelectric features, so these loops are confusing and easy to be misidentified.

The ML models learn how to distinguish between ferroelectric and non-ferroelectric P–E loops through establishing a logic relationship between the features of the P–E loops and their labels. To illustrate this process, the RF model is used as an example. RF consists of many decision trees and it makes the final prediction based on the votes from all decision trees.[22] A decision tree classifies the samples by sorting them down the tree from the root to the leaf nodes.[29] During the training, the decision tree grows by splitting nodes according to certain cutoff values of the features. The optimal cutoff value for a feature is selected if the Gini index is minimized after the node splitting.[30] In other words, the optimal cutoff value makes the two resulting sub-nodes as pure as possible (note: if a sub-node contains only samples with the same label, it is maximally pure and its Gini index has the lowest value of 0). Additionally, the importance of a feature can also be evaluated based on the reduction of the Gini index brought by that feature.[31]

Figure 3f,g shows the top 150 most important features for the RF model, which are distributed mostly in the S-shaped (if exists) and corner regions. Therefore, the S-shaped region and the loop corner seem to contain most important information that is used by the RF model to make the identification. Although the information can be complex, an intuitive and physical interpretation may be given as follows. First, because a ferroelectric loop typically exhibits the S-shaped region while a non-ferroelectric loop does not,[18,32] whether or not the S-shaped region exists may therefore be used as the criterion for the identification. The
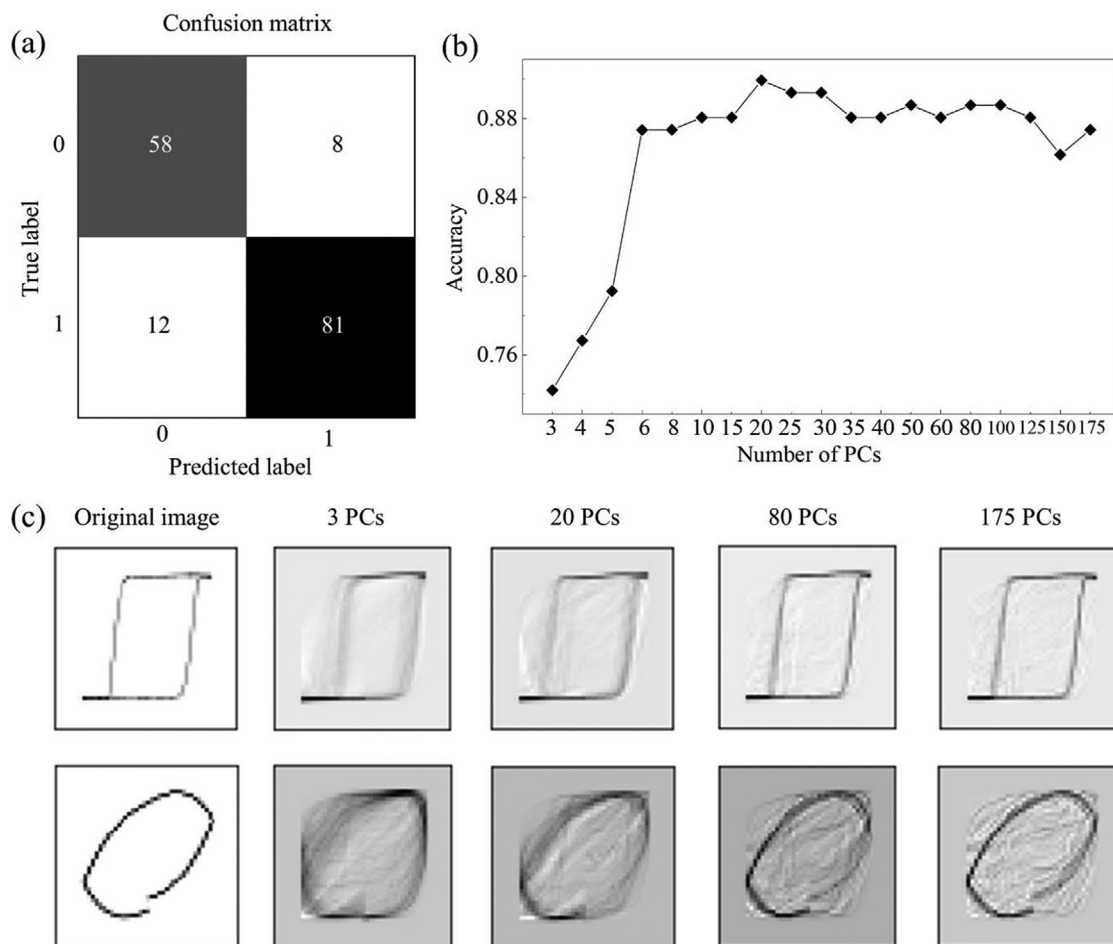
ADVANCED
SCIENCE NEWS
www.advancedsciencenews.com

ADVANCED
THEORY AND
SIMULATIONS
www.advtheorysimul.com

Figure 4. Testing performance of the image-based ML approach. a) Confusion matrix for the ANN model. b) Classification accuracy as a function of the number of PCs. c) Typical images of ferroelectric and non-ferroelectric P–E loops reconstructed using the first 3, 20, 80, and 175 PCs.

RF model can learn from the training data to obtain appropriate cutoff values for the relevant features to well characterize the S-shaped region. Here we use only four features, P_{51}, P_{71}, P_{131}, and P_{145}, as a simple example for illustration. The inequations of $P_{71} - P_{51} < \sigma_1$, $P_{145} - P_{131} < \sigma_2$, and $P_{131} - P_{71} > \sigma_3$, where σ_1 and σ_2 are sufficiently small values while σ_3 is a sufficiently large value, can indicate the existence of an S-shaped region. These inequations are fulfilled by the decision rules of $P_{51} > \epsilon_1$, $P_{71} < \epsilon_1 + \sigma_1$, $P_{131} > \epsilon_2$, and $P_{145} < \epsilon_2 + \sigma_2$, where ϵ_1 and ϵ_2 are appropriate cutoff values and $\epsilon_2 > \epsilon_1 + \sigma_1 + \sigma_3$. Using these decision rules, the RF model can determine whether or not the S-shaped region exists.

Then, the shape of the loop corner may be used as another criterion for the identification because ferroelectric and non-ferroelectric loops typically exhibit sharp and round corners, respectively.[18,33] For the RF model, the shape of the loop corner can be well expressed using appropriate cutoff values for the relevant features. Taking only two features, P_{430} and P_{473}, as an example, a sharp (round) corner typically has a small (large) value of $P_{430} - P_{473}$. Therefore, the RF model may learn and use the decision rules "$P_{430} < \epsilon_3$ and $P_{473} > \epsilon_4$" and "$P_{430} > \epsilon_3$ and $P_{473} < \epsilon_4$" to determine a loop corner as sharp

and round, respectively (note: ϵ_3 and ϵ_4 are appropriate cutoff values).

Based on whether or not the S-shaped region exists and the shape of the loop corner (sharp or round), the loop can be further identified as ferroelectric or non-ferroelectric. Although the actual identification procedures performed by an ML model should be much more complex, the above simple examples allow us to intuitively understand how an ML model identifies a P–E loop. The key to a successful identification is the extraction of important features, for example, those in the S-shaped and corner regions, from the training data, and establish a relationship between these features and the targeted labels.

Next, let us turn our attention to the image-based approach. Because a 50×50 grayscale image contains as many as 2500 features, PCA is thus performed to reduce the number of features.[26] As shown in Figure S3, Supporting Information, about 90% of the total feature variance can be retained as the number of principle components (PCs) reaches 175. The first 175 PCs are therefore used as the new descriptors for the dataset and then the identification is performed with the ANN algorithm. The resulting confusion matrix is presented in **Figure 4**a. 81 out of 93 ferroelectric loops and 58 out of 66 non-ferroelectric loops are

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
THEORY AND
SIMULATIONS**

www.advtheorysimul.com

predicted to have the correct labels of "1" and "0," respectively, resulting in an identification accuracy of 87.42%. This accuracy is slightly lower than that obtained by the ANN model used in the value-based approach (91.82%), which may be because using image pixel values rather than polarization values as features bring more difficulty to establish a relationship between the features and the targeted labels.

Figure 4b shows the dependence of the identification accuracy on the number of PCs. As the number of PCs increases from 3 to 20, the accuracy increases and it reaches a maximum value of 89.94% at the number of PCs = 20. As the number of PCs further increases from 20 to 175, the accuracy slowly decays. These results indicate that the first 20 PCs contain useful information to separate the ferroelectric and non-ferroelectric classes in the dataset, and the use of them for the ANN model can therefore lead to an accurate identification. However, the PCs out of the first 20 ones may contain information about the background noises and the common features shared by the ferroelectric and non-ferroelectric classes,[34] which is useless and even harmful for making an accurate identification. Therefore, the highest accuracy is achieved with the first 20 PCs while further increasing the number of PCs beyond 20 leads to the slight decay of accuracy. To further understand this, typical images of ferroelectric and non-ferroelectric loops are reconstructed using the first 3, 20, 80, and 175 PCs. As shown in Figure 4c, the reconstructed images differ much from their respective original images when only 3 PCs are used. As a result, the ANN model cannot discriminate these two reconstructed images properly, leading to a low identification accuracy of 74.21% (see Figure 4b). When the number of PCs increases to 20, the major characteristics of the ferroelectric (non-ferroelectric) loop, that is, the presence (absence) of the S-shaped region and the sharp (round) loop corner, are clearly visible in the reconstructed images. Therefore, these two reconstructed images can be well distinguished by the ANN model, and this explains why a high identification accuracy of 89.94% is achieved with the first 20 PCs. As the number of PCs further increases from 20 to 175, the loop regions in the reconstructed images change closer to those in the original images but the changes are small. Meanwhile, the background regions in the reconstructed images also change and the changes seem to be chaotic. These two factors have opposite effects on the identification accuracy and eventually lead to the slight decay of accuracy.

The above results have demonstrated that both the value-based and image-based ML approaches can achieve high accuracies for identifying P–E loops as ferroelectric or non-ferroelectric. Besides, these ML approaches also possess very high efficiencies for the identification. For example, to identify 159 loops, the average time needed by the value-based approach is only 0.53 s (0.49, 0.49, 0.13, 1.47, and 0.05 s for the RF, GBT, ET, ANN, and SVM models, respectively), while that needed by the image-based approach is only 2.49 s. Therefore, the ML approaches can greatly accelerate the identification process when the number of samples is very large.

Note that the dataset size in the present study (633 samples) is still limited. Further increasing the dataset size may improve the performance of the ML models, which can be done by conducting more experiments and requesting data from colleagues. Such efforts are currently ongoing. In addition, when antiferroelectric P–E loops are added into the dataset, the problem changes from binary classification to multi-class classification, which can be readily handled by the ML models. Last but not the least, the ML-based identification of P–E loops may pave the way for further ML studies. For example, The P–E loops identified as ferroelectric by our ML approaches may be used to develop new ML models that can extract further information from the loops, such as the remnant polarization values and coercive fields.

3. Conclusions

In summary, we have developed two ML approaches, namely, the value- and image-based approaches, to identify P–E loops as ferroelectric or non-ferroelectric. The input datasets for these two approaches consist of polarization values and 50 × 50 grayscale loop images, respectively. For the value-based approach, five different ML models including RF, GBT, ET, SVM, and ANN are used for the identification, among which the SVM model achieves the highest accuracy of 93.08%. For the image-based approach, the ANN model is used for the identification and it achieves an accuracy as high as 87.24%. The remarkable performance of our ML approaches can be attributed to the ability of the ML algorithms to extract the important information from a P–E loop, for example, whether or not the S-shaped region exists and whether the loop corner is sharp or round, and correlate such information with the label of the P–E loop (ferroelectric or non-ferroelectric). In addition, the average time needed by the ML models to identify about 160 loops is only ≈1.0 s, demonstrating the high efficiency of identification. Given the high accuracy and efficiency, our ML approaches therefore significantly outperform the manual way for distinguishing between ferroelectric and non-ferroelectric P–E loops, which may serve as a useful tool for researchers in the ferroelectric community.[35]

4. Experimental Section

Sample Preparation and Electrical Measurement: Epitaxial and polycrystalline $Pb(Zr_{0.2}Ti_{0.8})O_3$ (PZT) thin films were grown on (001)-oriented $SrRuO_3/SrTiO_3$ (SRO/STO) substrates and $LaNiO_3$ (LNO)/SiO_2/Si substrates, respectively, by using pulsed laser deposition (PLD) with a KrF excimer laser ($\lambda = 248$ nm). For the depositions of different batches of PZT films, different temperatures ranging from 580 to 650 °C and different oxygen pressures ranging from 12 to 18 Pa, were used. After the PZT deposition, patterned Pt top electrodes were deposited on the PZT film through a shadow mask using PLD. The P–E loops of the resultant Pt/PZT (epitaxial)/SRO and Pt/PZT (polycrystalline)/LNO capacitors were measured with a Radiant ferroelectric workstation.

Machine Learning Models: RF was a tree-based ensemble learning algorithm. It was an ensemble of B decision trees $\{T_1(X),...,T_B(X)\}$, where $X = (x_1,...,x_p)$ was a p-dimension vector of features. The ensemble produced B outputs, $\{\hat{y}_1 = T_1(X),..., \hat{y}_B = T_B(X)\}$, where $\hat{y}_b$ ($b = 1, 2,...,B$) was the prediction given by the bth tree. Outputs of all the trees were aggregated to generate the final prediction, $\hat{y}$. Each tree in an RF was trained with a subset of samples drawn from the original training set using the bootstrap sampling method. At each node of a tree, the best split was chosen among a randomly selected subset of the features (rather than all of them) according to certain rules (e.g., minimizing the Gini index).

GBT was an ML algorithm that iteratively constructed an ensemble of weak decision tree learners through gradient boosting. In each iteration, a new tree was created to fit the current residual, that is, the gradient of loss function. This new tree was then added to the ensemble to improve the final prediction.

ET was basically similar to RF in concept, and their major differences are described as follows. ET used all the samples in the training set to train a tree, while RF used the bootstrapped samples. In addition, ET randomly selected a split at each node, while RF selected the best split.

SVM was a supervised learning algorithm, which performed the classification by finding a hyper-plane (i.e., the decision boundary) in the feature space that best separated the data points from two classes. The best hyper-plane was defined as the one that maximized the margin between the two data classes. For nonlinearly separable data, a kernel function was used to map the data points in the low-dimension space to a higher-dimension space, so that the hyper-plane for separation could be easily found.

ANN was an ML algorithm that attempted to mimic how the human brain processed information. An ANN consisted of many neurons arranged in different layers: input, hidden, and output layers. Neurons in a layer were connected with those in the neighboring layers. Each neuron computed a weighted sum of its inputs and then fed the weighted sum to an activation function to produce an intermediate output. The intermediate output then propagated to the next layer and eventually reached the output layer, a process called the forward propagation. An ANN was trained by adjusting the connection weights based on the back propagation algorithm.

PCA was a dimensionality-reduction method that transformed a large set of variables to a smaller set of orthogonal variables called principle components (PCs) while preserving as much information as possible. The PCs were computed in such a way that the first PC accounted for the highest variance of the original variables. The second PC, which was computed under the condition that it was orthogonal to the first PC, accounted for the next highest variance. The other PCs were computed likewise.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Conflict of Interest

The authors declare no conflict of interest.

[1] W. Sun, Y. Zheng, K. Yang, Q. Zhang, A. A. Shah, Z. Wu, Y. Sun, L. Feng, D. Chen, Z. Xiao, S. Lu, Y. Li, K. Sun, *Sci. Adv.* **2019**, *5*, eaay4275.

[2] G. Hautier, C. C. Fischer, A. Jain, T. Mueller, G. Ceder, *Chem. Mater.* **2010**, *22*, 3762.

[3] P. Raccuglia, K. C. Elbert, P. D. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, A. J. Norquist, *Nature* **2016**, *533*, 73.

[4] Z. Li, Q. Xu, Q. Sun, Z. Hou, W. J. Yin, *Adv. Funct. Mater.* **2019**, *29*, 1807280.

[5] D. Xue, P. V. Balachandran, J. Hogden, J. Theiler, D. Xue, T. Lookman, *Nat. Commun.* **2016**, *7*, 11241.

[6] S. Lu, Q. Zhou, Y. Ouyang, Y. Guo, Q. Li, J. Wang, *Nat. Commun.* **2018**, *9*, 3405.

[7] R. Yuan, Z. Liu, P. V. Balachandran, D. Xue, Y. Zhou, X. Ding, J. Sun, D. Xue, T. Lookman, *Adv. Mater.* **2018**, *30*, 1702884.

[8] P. V. Balachandran, B. Kowalski, A. Sehirlioglu, T. Lookman, *Nat. Commun.* **2018**, *9*, 1668.

[9] J. Wu, Z. Fan, D. Xiao, J. Zhu, J. Wang, *Prog. Mater. Sci.* **2016**, *84*, 335.

[10] M. Dawber, K. Rabe, J. Scott, *Rev. Mod. Phys.* **2005**, *77*, 1083.

[11] L. Liu, S. Wang, Z. Yin, W. Liu, X. Xu, C. Zhang, X. Li, J. Yang, *Chin. Phys. B* **2016**, *25*, 097801.

[12] A. Mukherjee, M. Banerjee, S. Basu, N. T. K. Thanh, L. Green, M. Pal, *Phys. B* **2014**, *448*, 199.

[13] A. Kumar, P. Sharma, D. Varshney, *J. Ceram.* **2015**, *2015*, 869071.

[14] S. Gupta, M. Tomar, A. James, V. Gupta, *Ferroelectrics* **2013**, *454*, 41.

[15] M. Basith, A. Billah, M. Jalil, N. Yesmin, M. A. Sakib, E. K. Ashik, S. E. H. Yousuf, S. S. Chowdhury, M. S. Hossain, S. H. Firoz, *J. Alloys Compd.* **2017**, *694*, 792.

[16] M. M. Abolhasani, F. Zarejousheghani, Z. Cheng, M. Naebe, *RSC Adv.* **2015**, *5*, 22471.

[17] K. M. Batoo, J. P. Labis, R. Sharma, M. Singh, *Nanoscale Res. Lett.* **2012**, *7*, 511.

[18] J. Scott, *J. Phys.: Condens. Matter* **2008**, *20*, 021001.

[19] W. Sun, M. Li, Y. Li, Z. Wu, Y. Sun, S. Lu, Z. Xiao, B. Zhao, K. Sun, *Adv. Theory Simul.* **2019**, *2*, 1800116.

[20] J. A. K. Suykens, J. Vandewalle, *Neural Process. Lett.* **1999**, *9*, 293.

[21] S. Dreiseitl, L. Ohno-Machado, *J. Biomed. Inf.* **2002**, *35*, 352.

[22] A. Liaw, M. Wiener, *R news* **2002**, *2*, 18.

[23] J. H. Friedman, *Ann. Stat.* **2001**, *29*, 1189.

[24] P. Geurts, D. Ernst, L. Wehenkel, *Mach. Learn.* **2006**, *63*, 3.

[25] B. Yegnanarayana, *Sadhana* **1994**, *19*, 189.

[26] S. Wold, K. Esbensen, P. Geladi, *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37.

[27] J. A. Hanley, B. J. McNeil, *Radiology* **1982**, *143*, 29.

[28] M. Sokolova, G. Lapalme, *Inf. Process. Manage.* **2009**, *45*, 427.

[29] J. R. Quinlan, *Mach. Learn.* **1986**, *1*, 81.

[30] L. E. Raileanu, K. Stoffel, *Ann. Math. Artif. Intell.* **2004**, *41*, 77.

[31] B. H. Menze, B. M. Kelm, R. Masuch, U. Himmelreich, P. Bachert, W. Petrich, F. A. Hamprecht, *BMC Bioinf.* **2009**, *10*, 213.

[32] K. Du, B. Gao, Y. Wang, X. Xu, J. Kim, R. Hu, F.-T. Huang, S.-W. Cheong, *npj Quantum Mater.* **2018**, *3*, 33.

[33] X.-W. Zhao, S.-N. Dong, G.-Y. Gao, Z.-X. Xu, M. Xu, J.-M. Yan, W.-Y. Zhao, Y.-K. Liu, S.-Y. Yan, J.-X. Zhang, *npj Quantum Mater.* **2018**, *3*, 52.

[34] S. Pyatykh, J. Hesser, L. Zheng, *IEEE Trans. Image Process.* **2012**, *22*, 687.

[35] An easy-to-use tool called "*P–E* loop discriminator" has been developed, and readers can access it via www.peloop.top (website accessed July 2020).