Supplementary Information (SI) for Journal of Materials Chemistry C.

Accelerating the prediction of remanent polarization in multicomponent ferroelectrics by Using Variational Autoencoder-Based Data Augmentation

Zixiong Sun[ab], Ruyue Gao[a], Ping Wang[a], Xinying Liu[a], Yujie Bai[a], Jingyu Luo[b],*

Hongyu Yang[c], Wanbiao Hu[d*],*

a School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an 710021, PR China.

b School of Materials Science and Engineering, Shaanxi University of Science and Technology, Xi'an 710021, PR China.

c School of Advanced Materials and Nanotechnology, Xidian University, Xi'an, 710071, China.

d Yunnan Key Laboratory of Electromagnetic Materials and Devices, Yunnan, 650000, PR China.

**Corresponding author: Zixiong Sun, Hongyu Yang, Wanbiao Hu.*

E-mail address: sunzx@sust.edu.cn, huwanbiao@ynu.edu.cn, yanghongyu@xidian.edu.cn.

Supplementary Information

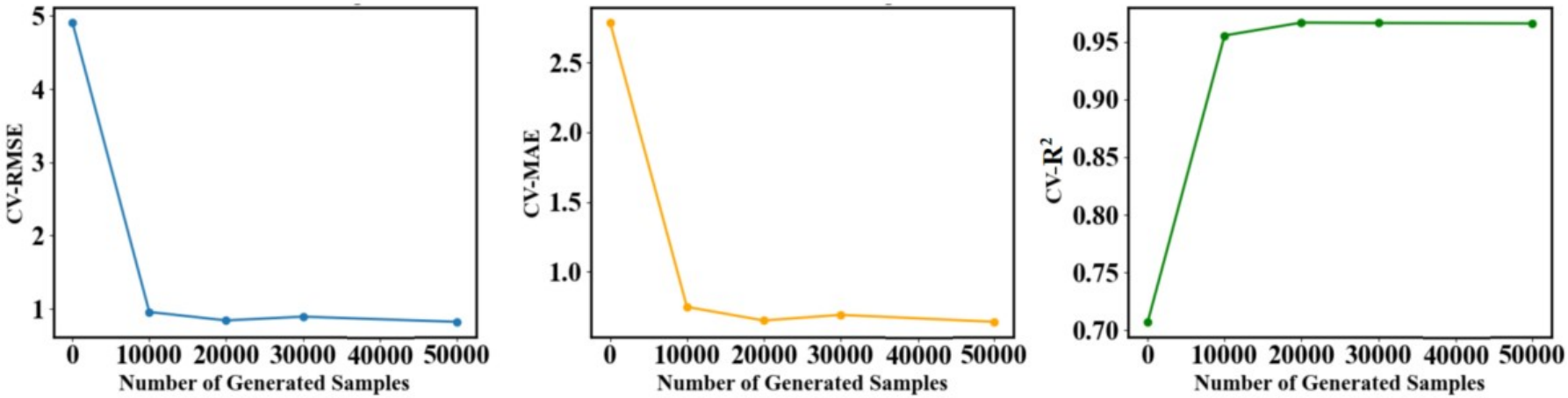

Figure S1.Training curves of the VAE model showing the variation of total loss, reconstruction loss, and KL divergence loss over epochs.

Table S1. Chemical composition of selected samples in the dataset

NN	$NaNbO_3$	NN-BCZ	$(1-x-y)NaNbO_3$-$xBaZrO_3$-$yCaZrO_3$
KNN	$K_{1-x}Na_xNbO_3$	NN-BZT	$(1-x)\ Ba(Zr_yTi_{1-y})O_3$-$xNaNbO_3$
BZ	$BaZrO_3$	NN-CT	$(1-x)NaNbO_3$-$xCaTiO_3$
BT	$BaTiO_3$	NN-BCZT	$(1-x-y)NaNbO_3$-$xBaZrO_3$-$yCaZrO_3$
BCZT	$Ba_{1-x}Ca_x(Ti_{1-y}Zr_y)O_3$	KNN-BT	$(1-x)K_{1-y}Na_yNbO_3$-$xBaTiO_3$
CT	$CaTiO_3$	KNN-BZ	$(1-x)(K_{1-y}Na_y)NbO_3$-$xBaZrO_3$
BCT	$Ba_{1-x}Ca_xTiO_3$	KNN-CZ	$(1-x)(K_{1-y}Na_y)NbO_3$-$xCaZrO_3$
NN-BT	$(1-x)NaNbO_3$-$xBaTiO_3$	KNN-CT	$(1-x)(K_{1-y}Na_y)NbO_3$-$xCaTiO_3$
NN-BZ	$(1-x)NaNbO_3$-$xBaZrO_3$	KNN-BCZT	$(1-x)(K_{1-u}Na_u)NbO_3$-$xBa_{1-y}Ca_y(Ti_{1-z}Zr_z)O_3$
NN-CZ	$(1-x)NaNbO_3$-$xCaZrO_3$	KNN-BCZ	$(1-x-y)(K_{1-x}Na_x)NbO_3$-$xBaZrO_3$-$yCaZrO_3$
KN-BT	$(1-x)BaTiO_3$-$xKNbO_3$	KN-BCT	$(1-x-y)KNbO_3$-$xBaTiO_3$-$yCaTiO_3$
BZT	$BaZr_{1-x}Ti_xO_3$	KN-NN	$(1-x)KNbO_3$–$xNaNbO_3$

Table S2. Specific performance parameters for the 8 regression models

Model	CV_Error	MAE	R^2
RandomForest	1.291657	0.809194	0.943266
DecisionTree	1.431178	0.847509	0.935604
SVR	1.662221	1.068328	0.899552
GradientBoosting	2.010606	1.510060	0.827059
LinearRegression	1.353745	0.868121	0.909955
XGBoost	1.283504	0.812126	0.944982
AdaBoost	1.443968	1.020491	0.895607
Ridge	1.358047	0.868652	0.909734

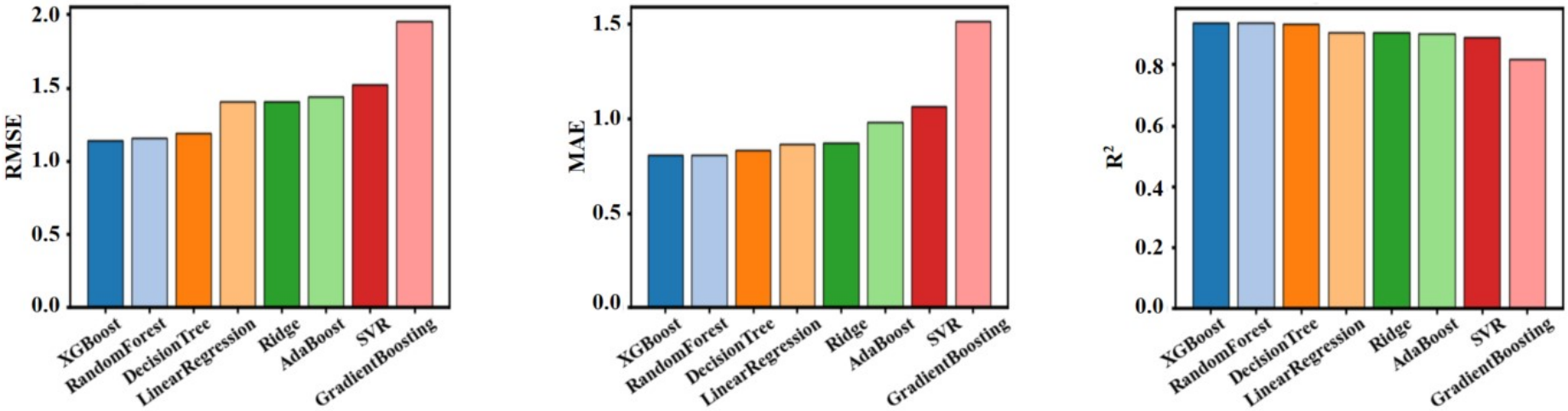

Figure S2. Performance comparison of various models based on R^2, MAE, and CV Error metrics.

Table S3. Key hyperparameters of the eight regression models employed in this study.

Model Name	n_estimators	max_depth	random_state	criterion
RandomForest	8	10	42	squared_error

Model Name	max_depth	random_state	criterion	splitter
DecisionTree	8	42	squared_error	best

Model Name	C	epsilon	kernel	degree
SVR	1.0	0.1	rbf	3

Model Name	n_estimators	max_depth	learning_rate	random_state
GradientBoosting	10	8	0.1	42

Model Name	fit_intercept	normalize	copy_X	n_jobs
LinearRegression	True	deprecated	True	None

Model Name	n_estimators	max_depth	learning_rate	booster
XGBoost	10	8	0.3	gbtree

Model Name	n_estimators	learning_rate	random_state	base_estimator
AdaBoost	10	1.0	42	None

Model Name	alpha	fit_intercept	normalize	solver
Ridge	1.0	True	deprecated	auto

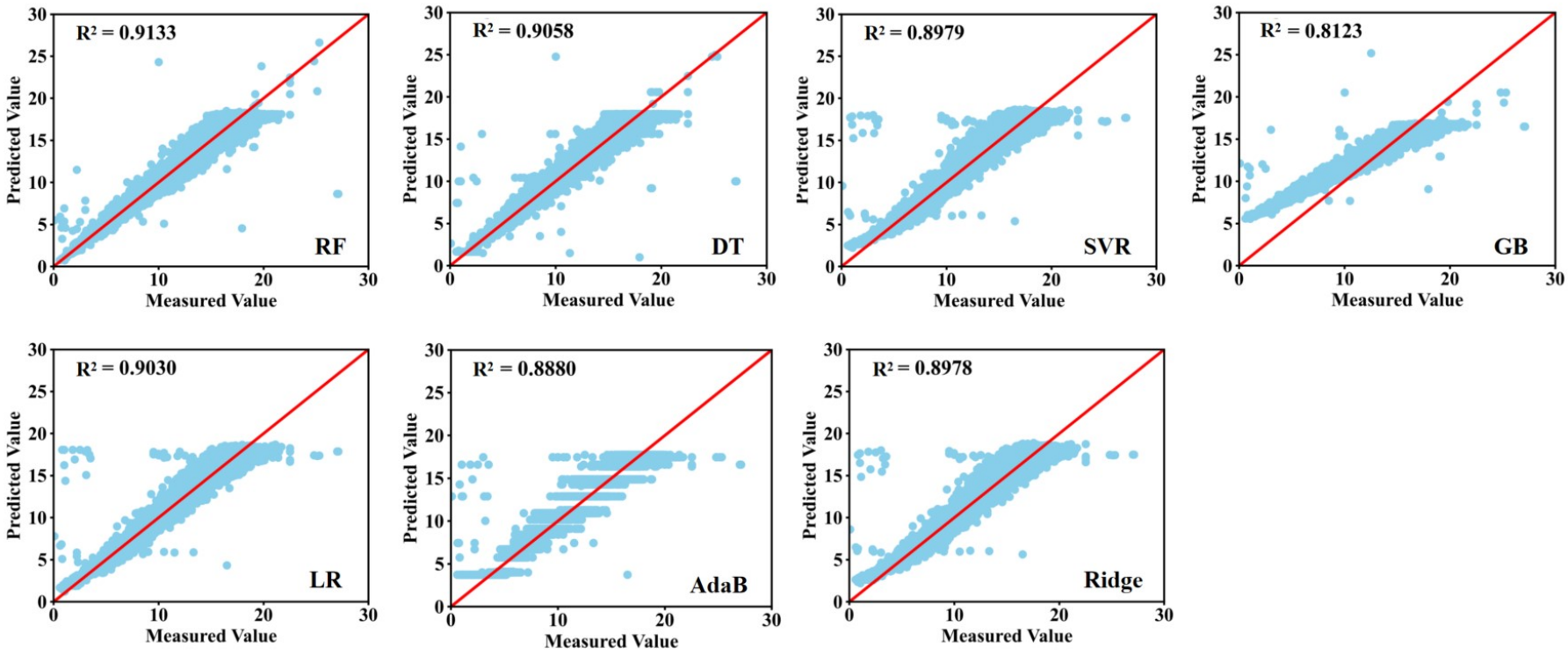

Figure S3. The best 6 features and the performance predictions of the remaining 7 models on the test data.

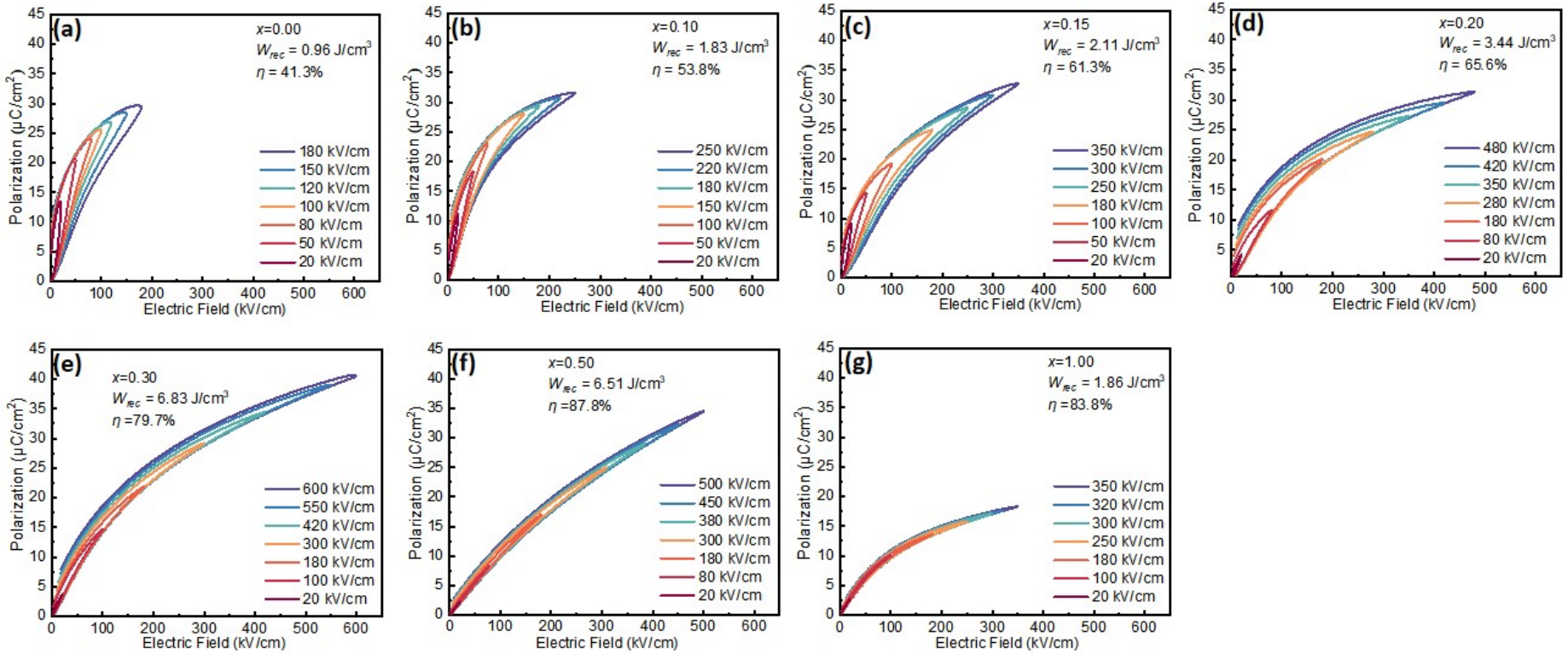

Figure S4. P-E loop of of (1 - x)($K_{0.5}Na_{0.5}$)NbO_3-x$Ba_{0.9}Ca_{0.1}Zr_{0.15}Ti_{0.85}O_3$, (x=0, 0.1, 0.15, 0.2, 0.3, 0.5, 1) ceramics under different electric fields.

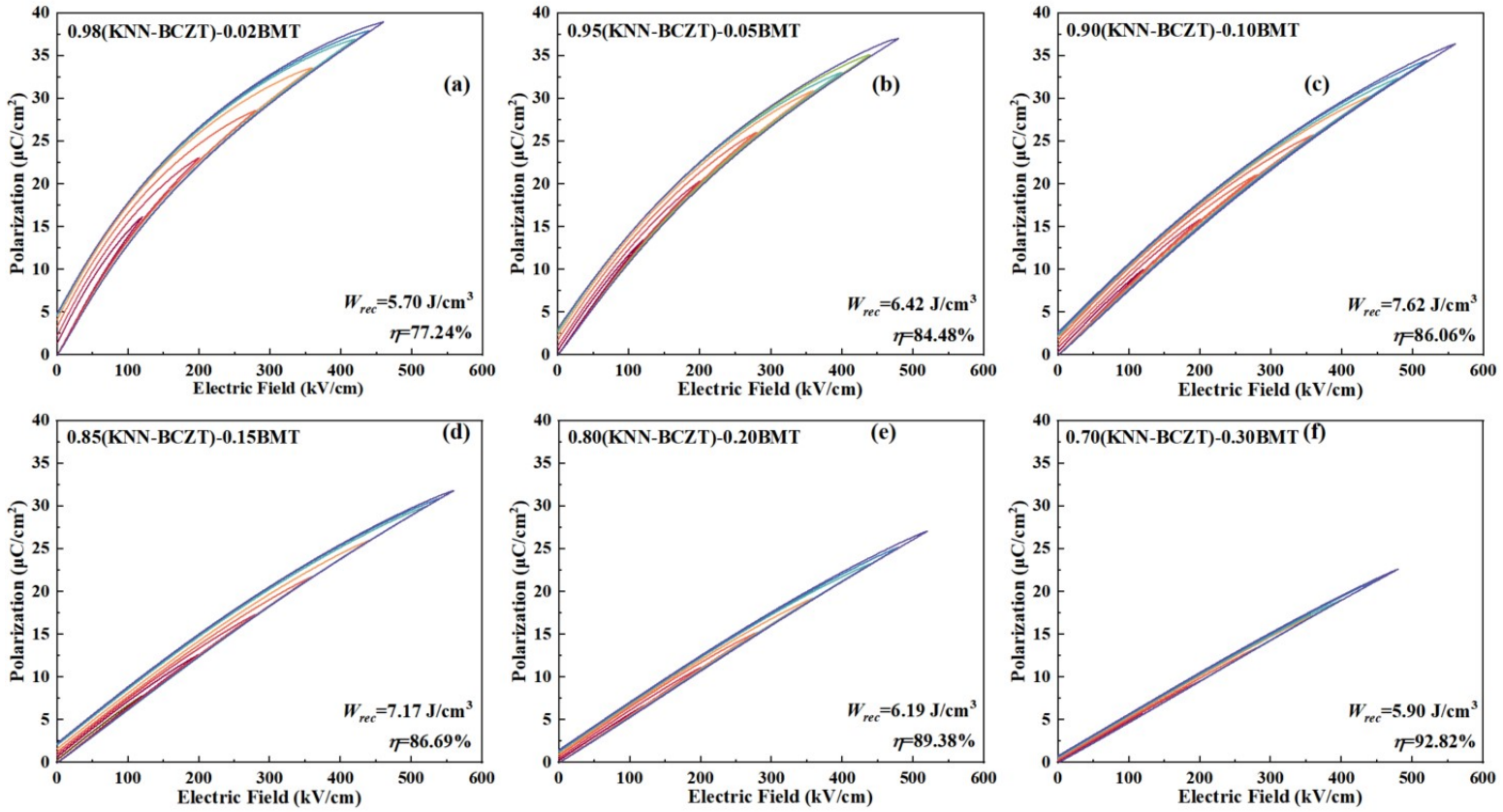

Figure S5. P-E loop of of (1 - x)[$(K_{0.5}Na_{0.5})NbO_3$-x$Ba_{0.9}Ca_{0.1}Zr_{0.15}Ti_{0.85}O_3$]-x$Bi(Mg_{1/3}Ta_{2/3})O_3$, (x=0, 0.1, 0.15, 0.2, 0.3, 0.5, 1) ceramics under different electric fields.