Science and Technology of Advanced Materials: Methods

ISSN: 2766-0400 (Online) Journal homepage: www.tandfonline.com/journals/tstm20

Causal inference in statistics insights into stress-induced ferroelectric states in SrTiO: disentangling piezoelectric and flexoelectric effects from birefringence images

Kazuma Seike, Hirotaka Manaka & Yoko Miura

To cite this article: Kazuma Seike, Hirotaka Manaka & Yoko Miura (16 May 2025): Causal inference in statistics insights into stress-induced ferroelectric states in SrTiO: disentangling piezoelectric and flexoelectric effects from birefringence images, Science and Technology of Advanced Materials: Methods, DOI: 10.1080/27660400.2025.2503698

To link to this article: https://doi.org/10.1080/27660400.2025.2503698

Publisher: Taylor & Francis & The Author(s). Published by National Institute for Materials Science in partnership with Taylor & Francis Group
Journal: *Science and Technology of Advanced Materials: Methods*
DOI: 10.1080/27660400.2025.2503698

Causal inference in statistics insights into stress-induced ferroelectric states in $SrTiO_3$: disentangling piezoelectric and flexoelectric effects from birefringence images

Kazuma Seike[a], Hirotaka Manaka[a], and Yoko Miura[b]
[a]Graduate School of Science and Engineering, Kagoshima University, Korimoto, Kagoshima 890-0065, Japan; [b]National Institute of Technology, Suzuka College, Shiroko-cho, Suzuka, Mie 510-0294, Japan

ARTICLE HISTORY
Compiled March 30, 2025

ABSTRACT
In materials science, experimental conditions are precisely controlled to ensure high reproducibility. This property is well suited for causal inference in statistics, yet its potential remains unrealized. In this study, we integrate causal inference with structural equation modeling (SEM) to analyze birefringence images of the stress-induced ferroelectric $SrTiO_3$. Random forest analysis identified retardance at 14.1 K, $\delta(14.1K)$, as a key predictor of the ferroelectric phase transition temperature, T_F. SEM revealed strong correlations between T_F and δs, although multicollinearity in δs necessitated sparse principal component analysis to transform $\delta(14.1K)$ and $\delta(40.0K)$ into two independent components, $PC1$ and $PC2$, for subsequent causal analysis. Directed acyclic graphs (DAG) based on SEM helped infer causal relationships, revealing $PC1$'s predominant influence on T_F in stress/strain-concentrated regions (cluster E12) and $PC2$'s influence in uniform stress/strain regions (cluster E34). Two-model learner (T-Learner) analysis revealed the factors behind the higher T_F in E12 than in E34. Specifically, the difference in magnitude of the piezoelectric effect, which occurs uniformly throughout the substrate, causes T_F to increase by 0.36 K [0.18 K, 0.54 K], while the flexoelectric effect, which occurs only in E12, causes it to increase by an additional 1.49 K [1.23 K, 1.75 K]. These findings demonstrate the utility of causal inference in disentangling piezoelectric and flexoelectric effects in the stress-induced ferroelectric $SrTiO_3$.

KEYWORDS
Causal inference; SEM; DAG; T-Learner; random forest; flexoelectricity; piezoelectricity; birefringence image; stress-induced ferroelectricity; $SrTiO_3$

Corresponding author Hirotaka Manaka. E-mail: manaka@eee.kagoshima-u.ac.jp

1. Introduction

In materials science, experimental observations reveal intricate relationships among multiple variables. However, such relationships are typically correlative, not causal, and may represent only a subset of the system's complex dynamics, leaving significant gaps in understanding physical mechanisms[1–3]. Furthermore, this challenge is further complicated by spurious correlations, where unrelated variables appear connected due to confounding factors, hindering definitive conclusions based on experimental data alone. Traditional approaches rely on fitting theoretical models to experimental results, but these models are frequently modified based on subjective interpretations of data. Therefore, identifying both correlations and causal relationships within experimental data is crucial to advance our understanding.

Recent advances in causal inference in statistics offer powerful tools to address these challenges[4, 5]. Unlike traditional statistical analysis, causal inference provides a systematic framework for modeling and testing causal relationships beyond simple correlations, allowing researchers to estimate system-wide changes resulting from hypothetical interventions[6]. By integrating domain knowledge and statistical models, these methods disentangle direct and indirect effects, account for confounding factors, and quantify causal effects with greater confidence. Structural equation modeling (SEM), a complementary approach, has been widely utilized to analyze latent variables representing unobserved factors influencing a system[7, 8]. Although SEM excels at revealing correlations, it lacks the capability to directly identify causal relationships. Therefore, SEM can be integrated with causal inference to realize the full potential of these analytical methods, enabling researchers to quantitatively evaluate unobserved regularities[9].

Materials science, characterized by precisely controlled experimental conditions and high reproducibility, is particularly suited for causal inference applications. Leveraging SEM with causal inference holds significant potential for uncovering the hidden regularities governing material properties. This integration reveals causal links between variables, uncovering previously unobservable physical mechanisms that drive complex phenomena. The resulting insights advance materials science by refining theoretical models and inspiring novel experimental approaches. However, this powerful methodology remains underutilized in materials science[10, 11]. In this study, we developed an integrated framework combining causal inference with SEM to analyze birefringence images of stress-induced ferroelectric $SrTiO_3$ under external force[12, 13]. The primary aim was to quantitatively disentangle the contributions of piezoelectric and flexoelectric effects to birefringence during the stress-induced ferroelectric phase transition in quantum paraelectric $SrTiO_3$. Experimentally, distinguishing between these effects is challenging owing to their overlapping spontaneous polarizations[14, 15]. To address this, causal inference with SEM was employed to isolate these overlapping effects, providing insights into the underlying mechanisms.

Unlike piezoelectricity, flexoelectricity generates polarization from strain gradients and occurs in all dielectrics regardless of crystal symmetry [16–18]. This unique property makes flexoelectricity a promising mechanism for applications in stress/strain sensors, nanoscale actuators, and energy harvesting devices in microelectromechanical systems[19, 20]. Historically, flexoelectricity has been more extensively studied in thin films than in bulk materials, primarily due to the ease of generating strain gradients at the nanoscale[21, 22]. Recent advances in surface potential measurement techniques, such as scanning probe microscopy, have broadened the scope of flexoelectricity research[23, 24]. Early theoretical

work suggested that the flexoelectric effect scales with dielectric constant, as materials with higher permittivity exhibit stronger polarization responses to strain gradients[25]. This hypothesis has spurred interest in exploring ferroelectrics and relaxors to enhance flexoelectric responses in bulk materials [26–34].

Bulk $SrTiO_3$ undergoes a structural phase transition from cubic to tetragonal at 105 K, exhibiting a high dielectric constant below this temperature[35–37]. However, quantum fluctuations suppress the emergence of a ferroelectric state, maintaining a quantum paraelectric state. These quantum fluctuations predominantly affect the orientations of local dipole moments formed by small displacements of Ti ions within the oxygen octahedral structure, preventing long-range ferroelectric order[38, 39]. In contrast, fluctuations in the magnitude of dipole moments are relatively small and do not play a dominant role in suppressing ferroelectricity[40]. This suppression mechanism, driven by zero-point energy effects, can be overcome by applying electric fields or mechanical forces, allowing the emergence of a ferroelectric state at low temperatures[12,13,41–51]. Over the past two decades, numerous studies have explored stress/strain-induced ferroelectric states. It is widely believed that spontaneous polarization near surfaces or interfaces arises from both piezoelectric and flexoelectric effects[52–58]. In our previous birefringence imaging measurements on bulk $SrTiO_3$ under an external force of 231 MPa applied along [001], we observed that the ferroelectric phase transition occurs below 30 K[12, 13]. As later discussed in Section 2, the ferroelectric phase transition temperature (T_F) showed a localized increase in stress/strain-concentrated regions, even though the structural phase transition temperature (T_c) remained uniformly distributed[59, 60]. This finding challenges the previous hypothesis that both T_c and T_F are proportional to the magnitude of stress.

In this study, we developed a robust framework integrating causal inference with SEM to systematically evaluate the contributions of piezoelectric and flexoelectric effects to stress-induced ferroelectricity in $SrTiO_3$. By quantifying hypothetical intervention effects, this framework disentangles overlapping polarization mechanisms and provides a comprehensive understanding of stress-induced ferroelectric phase transitions. Furthermore, this data-driven approach offers a scalable methodology for analyzing other complex material systems.

2. Dataset for analysis

Successive birefringence images with a resolution of 384×288 pixels were obtained by lowering the temperature (T) of the $SrTiO_3$ substrate from 300.0 K to 14.1 K, under an external force of 231 MPa applied along [001], as described in Ref. [12]. In this setup, the refractive indices along the orthogonal principal optical axes are denoted as n_1 (slow axis) and n_2 (fast-axis), where $n_1 \geq n_2 \geq 1$. Birefringence (Δn) and retardance (δ) are defined as follows:

$$\Delta n \equiv n_1 - n_2, \tag{1}$$

$$\delta = \Delta n \times t, \tag{2}$$

where t represents the optical path length through the sample[61, 62]. The fast-axis direction (ψ) corresponds to the orientation associated with n_2. In general, δ is proportional to the stress/strain in the material due to the photoelastic effect[63]. Additionally, polarization induced by the piezoelectric and flexoelectric effects generates strain, further increasing δ. ψ reflects the direction of the stress/strain and the polarization[63–67]. Figures 1(a,b) illustrate the spatial distributions of T_F and T_c, respectively, obtained using the Bayesian T-series estimation from the T-dependence of δ at a wavelength of $\lambda = 575$ nm for each pixel[60]. These images, also utilized in this study, focus on a 302×140-pixel region of the substrate in the birefringence dataset. As shown in Figure 1(a), T_F exhibits locally elevated values in stress/strain-concentrated regions, forming a stripe-like pattern along $\langle 111 \rangle$ [59]. This is likely due to slip planes generated along $\langle 111 \rangle$ under an external force at room temperature, rendering ferroelectric domains visible along these planes. In contrast, Figure 1(b) shows a nearly uniform distribution of T_c across the substrate. These varying distributions suggest that the stress/strain concentration has a more profound influence on T_F compared to T_c.

Using the K-shape clustering method described in Ref. [60], the dataset was grouped into four clusters based on the shapes of the δ and ψ T-dependence curves, $\delta(T)$ and $\psi(T)$. Figure 2(a) presents the clustering results, while Figure 2(b) presents the averaged $\delta(T)$ curves for each cluster at 575 nm. Clusters E1 and E2 exhibit higher δ values than E3 and E4, with deviations of approximately 20 nm across all T s. For each cluster, box-and-whisker plots of T_F and T_c and data tables are provided in the Supplementary Materials. The mean T_F values for E1 and E2 are 1.85 K (95 % Bayesian confidence interval: [1.82 K, 1.87 K]) higher than those for E3 and E4. The T_F distributions in E1 and E2 are skewed toward higher temperatures, resulting in higher means than medians. In contrast, the T_F distributions in E3 and E4 are relatively symmetrical, with similar means and medians. These results indicate that T_F increases due to stress/strain concentration in E1 and E2, while uniform stress/strain is primarily observed in E3 and E4. Contrastingly, T_c remains nearly uniform distribution across all clusters, as evident in the box-and-whisker plots, reflecting its uniformity across the substrate. These distinct observations suggest that T_c is affected by stress/strain, while T_F is additionally affected by spontaneous polarization. We propose the following scenario: uniform stress/strain generates polarization throughout the substrate via the piezoelectric effect, while strain gradients induce additional localized polarization via the flexoelectric effect in E1 and E2. As T approaches $\sim T_c$ values, the dielectric constant decreases rapidly, suppressing the flexoelectric effect. According to Ref. [14], the flexoelectric effect manifests within the range $T_F < T < T_c$ and disappears at $\sim T_c$. However, its behavior below T_F remains unclear due to overlapping contributions from piezoelectric and flexoelectric effects, making it challenging to isolate the latter.

Since the ferroelectric phase transition is affected by local stress/strain distributions, birefringence measurements provide an indirect but effective means of assessing their effect on T_F. The distribution of T_F shown in Figure 1(a) was obtained from the T dependence of δ. Since the stress-induced ferroelectric phase transition is sensitive to local mechanical environments, analyzing the spatial distributions of δ and ψ provides an effective means of

examining how these factors influence T_F. Although it is possible to perform T-series causal inference by substituting the time variable with temperature, this approach presents significant computational challenges. The primary issue is not merely the increase in data volume but the need to address statistical dependencies such as unit roots and autocorrelation, which require extensive preprocessing. Previous clustering studies have demonstrated that key tendencies can be captured by analyzing datasets above and below T_F [59]. Therefore, this study focused on birefringence measurements from the spatial distributions of δ and ψ at four T values: 14.1 K ($< T_F$), 40.0 K ($> T_F$), 90.0 K ($< T_c$), and 130.9 K ($> T_c$). Among the three wavelength datasets ($\lambda = 523$, 543, and 575 nm), the 575-nm dataset was selected owing to higher signal-to-noise ratio resulting from the higher incident light intensity[62]. Since there is a strong correlation between the data at the four T points, appropriate preprocessing is required. However, as discussed in Section 4, the data can be processed using conventional statistical methods, so computational cost is not a limiting factor.

Figure 3 illustrates the spatial distributions of $\delta(14.1K)$ and $\delta(40.0K)$ at $\lambda = 575$ nm. In Figure 3(a), the $\delta(14.1K)$ distribution exhibits an unexpected value in the small rectangular region on the right, spanning several pixels. The values intrisitically exceed incident light λ of 575 nm, as described in Ref. [68]. However, due to the limitations of the measurement method, distinguishing between $\delta = \lambda(1+\Delta\theta/(2\pi))$ nm and $\lambda(1-\Delta\theta/(2\pi))$ nm is not possible; hence, $-\Delta\theta$ was assumed for all pixels. The sign could be theoretically determined by cross-referencing datasets for 543 and 523 nm due to the small λ-dependence of δ. However, verifying each pixel is very laborious. As these regions account for less than 1 % of the total pixel count, their impact on the overall analysis is negligible. Therefore, $-\Delta\theta$ was consistently assumed in the all data. The Supplementary Materials provide the spatial distributions for $\delta(90.0\,\mathrm{K})$, $\delta(130.9\,\mathrm{K})$, $\psi(14.1\,\mathrm{K})$, $\psi(40.0\,\mathrm{K})$, $\psi(90.0\,\mathrm{K})$, and $\psi(130.9\,\mathrm{K})$ at $\lambda = 575$ nm. These results demonstrate that the distribution patterns are largely invariant to T. The stripe-like structures in the ψ distributions are attributed to dislocations and occur independently of ferroelectric or structural phase transitions.

3. Data analysis procedure

The data-driven analysis in this study utilized birefringence imaging datasets obtained under 231-MPa external force, as reported in Ref. [12]. The analysis was conducted on a computer equipped with an Intel(R) Core(TM) i9-14900 CPU (up to 5.80 GHz) and 96 GB of memory, using a combination of advanced statistical and machine learning tools in the Python and R environments. In the Python environment (version 3.11.1), random forest regression was employed using the scikit-learn library (version 1.2.2) to investigate the relationship between explanatory variables and T_F. This method was chosen for its robustness in handling complex, nonlinear relationships within the dataset and its ability to rank variable importance. In R (version 4.1.2), SEM was performed to examine the correlations using the lavaan (version 0.6-18) and semTools (version 0.5-6) packages. Lavaan provided a framework for defining and estimating SEM models, while semTools enabled bootstrapped confidence interval calculations and multi-group analyses. To make the explanatory variables orthogonal while preserving explanability, sparse principal component analysis (sparse PCA) was conducted using the "sklearn.decomposition" module in Python. Causal inference was performed using the DoWhy

library (version 0.11.1) in Python[69]. DoWhy provides a robust framework for estimating and validating causal effects by integrating causal graph modeling, effect estimation, and robustness checks. This facilitated counterfactual analysis and the evaluation of potential unobserved confounding factors. To aggregate the estimation results, hierarchical Bayesian estimation was performed using the Hamiltonian Monte Carlo algorithm–-a sophisticated method for Markov chain Monte Carlo (MCMC) simulations. These calculations were performed using the rstan package (version 2.32.6) in the R environment, ensuring efficient and accurate parameter estimation. Statistical reliability of the calculation results from SEM and causal inference was evaluated using p values, which quantify the probability of observing results as extreme as the data under null hypothesis[70]. Small p values (typically less than 0.05) were interpreted as significant evidence against the null hypothesis, while larger values indicated insufficient evidence to reject the null hypothesis.

4. Results and discussion

To investigate the inhomogeneous spatial distribution of T_{F}, datasets of (δ,ψ) at 14.1, 40.0, 90.0, and 130.9 K, and T_{c} were analyzed. In causal inference, careful selection of essential variables simplifies model construction and enhances the explanatory power for identifying causal relationships. The relatively unchanged spatial distributions of δ and ψ across varying T, as shown in Figure 3 and the Supplementary Materials, suggest multicollinearity among these variables. To address this, we used random forests, which are relatively robust to multicollinearity and can handle complex, nonlinear relationships. Unlike gradient boosting, which builds decision trees sequentially and is more prone to overfitting noisy data, random forests train trees independently, increasing robustness to noise[71, 72]. Furthermore, random forests provide an interpretable ranking of the importance of explanatory variables because, unlike gradient boosting, they do not depend on the order in which the trees are constructed. This analysis identified the key variables influencing T_{F}. After reducing the variables for T_{F} from the random forest analysis, SEM was conducted to elucidate correlations between variables and derived a directed acyclic graph (DAG) for further causal interpretation. To validate the robustness of causal inference derived from DAGs, a two-model learner (T-Learner) approach was applied separately for groups with and without treatment[73]. This causal inference technique, combined with machine learning, allowed us to disentangle the flexoelectric and piezoelectric effects. The integration of random forest, SEM, and causal inference provides a more comprehensive understanding of the fundamental mechanisms underlying stress-induced ferroelectricity in $SrTiO_3$.

4.1. Regression analysis using random forests

Random forests, a robust ensemble learning method based on decision trees, are widely recognized for their high prediction accuracy and resilience to overfitting and multicollinearity. In this study, random forests were employed for regression analysis to predict the inhomogeneous distribution of T_{F}. The dataset was divided into 70 % training and 30 % test sets using the "train_test_split" function. A random forest regressor was constructed with 100 decision trees, each trained on bootstrapped samples of the training data. The model employed

root mean squared error (RMSE) reduction as the splitting criterion, minimizing the variance within nodes to ensure effective data partitioning. All features were considered at each split point to capture variable interactions. To maximize model flexibility, the minimum sample size for a split was set to two, with leaf nodes requiring at least one sample. No depth limit was applied, allowing the model to learn detailed patterns in the dataset.

The input dataset consisted of a 302×140-pixel block, excluding cluster numbers. The objective variable was the T_{F} distribution shown in Figure 1(a), while the explanatory variables included nine distribution datasets: $\delta(14.1\,K)$, $\delta(40.0\,K)$, $\delta(90.0\,K)$, $\delta(130.9\,K)$, $\psi(14.1\,K)$, $\psi(40.0\,K)$, $\psi(90.0\,K)$, $\psi(130.9\,K)$, and T_{c}. Each variable was standardized prior to input as a preprocessing step to improve prediction accuracy. Optimizing the random forest model with the nine explanatory variables resulted in an RMSE of 0.223, reflecting its generalization performance when evaluated on an independent test set. Furthermore, when evaluated on the entire dataset, including both training and test data, the RMSE was significantly lower at 0.140, indicating a good fit to the combined dataset. Feature importance rankings shown in Figure 4(a) revealed approximately 60 % contribution of $\delta(14.1\,K)$ to the prediction of T_{F}. The spatial distribution of the residuals ($\equiv$ measurement $-$ prediction), as shown in Figure 4(b), was destandardized to the match the scale of T_{F}. Large absolute values of the residuals were observed in the E1 and E2 regions, where stress/strain is concentrated, despite the exclusion of cluster number information. This suggests that the predictive accuracy for T_{F} is low in E1 and E2 due to dominance of experimental data from E3 and E4, where stress/strain is uniformly distributed and accounts for over 70 % of the total pixels. The results indicate that the regularities governing T_{F} differ between groups E1/E2 and E3/E4. Specifically, the inclusion of flexoelectric effects in predictive models is essential for accurately modeling T_{F} variations in the stress/strain-concentrated regions, such as E1 and E2.

Although random forests excel at capturing complex nonlinear relationships, their feature importance scores can sometimes emphasize variable interactions, potentially obscuring the contributions of individual predictors. To address this, complementary methods such as correlation analysis are valuable. In particular, Spearman's rank correlation coefficient is robust to outliers and provides intuitive understanding by assessing correlations between variables without assuming linearity. Figure 5 presents the Spearman's rank correlation coefficients calculated for each cluster. The analysis revealed a strong positive correlation between T_{F} and $\delta(14.1K)$ across clusters. This finding aligns with the feature importance rankings from the random forest analysis, reinforcing $\delta(14.1K)$ as the dominant predictor of T_{F}. This consistency between the two approaches enhances the reliability of the random forest results. Further examination of correlation coefficients for clusters E1 and E2 revealed strong positive correlations between T_{F} and $\delta(40.0K)$, $\delta(90.0\,K)$, and $\delta(130.9\,K)$. In contrast, these correlations were weak in E3 and E4. Additionally, comparing the correlation coefficients among variables other than T_{F} showed that E1 and E2 exhibited similar trends, as did E3 and E4. These results suggest that the dataset naturally divides into two groups: E1/E2 and E3/E4. The Supplementary Materials include the results from random forest analysis that incorporated neighboring pixel information. Based on these findings, $\delta(14.1\,K)$ remained the highest-importance variable, showing similar trends as the primary analysis.

4.2 Correlation analysis using structural equation modeling (SEM)

The nonlinear analysis using random forests revealed that $\delta(14.1K)$ moderately explains the spatial distribution of T_F. To quantitatively identify which factors contribute to T_F, SEM was employed. SEM integrates factor analysis and linear regression to statistically clarify the relationship between T_F and the explanatory variables.

Figure 6 illustrates the SEM path diagrams, showing the relationships between the explanatory variables (δ, ψ) and the objective variables (T_F, T_c) using the complete 302×140-pixel dataset. In these diagrams, unidirectional arrows with numbers from the explanatory variable to the objective variable represent factor loadings. Table 1 details the estimated factor loadings for each path. The square of each factor loading represents the proportion of the explanatory variable in the objective variable. The variance of each explanatory variable is normalized to 1.000 (see the Supplementary Materials), as indicated by the bidirectional arrows. The sum of squared factor loadings determines the proportion in single objective variable explained by all four explanatory variables. If this value is less than 1, there is some variance that cannot be explained by the four explanatory variables. This unexplained variance is represented by the arrow pointing toward the objective variable, indicating the error variance as an independent component. Table 2 summarizes the results of error variance. As indicated by the values of the bidirectional arrows in Figure 6, the covariance between δs or ψs is considerably large (see the Supplementary Materials). These values are consistent with the Spearman's rank correlation coefficient in Figure 5. Due to notable multicollinearity, some factor loadings exceeded 1, and several p values exceeded 0.05. From the four explanatory variables and one objective variable in this model, the variance–covariance matrix of the observed data contains $5\times(5+1)\div2=15$ unique elements (information content). With four factor loadings and one error variance, the regression analysis model retains ten degrees of freedom. However, the lack of diversity in the observed data arising from multicollinearity reduces the actual degrees of freedom to zero, rendering the χ^2-test invalid. Table 3 presents model fit indices. The RMSE of approximation (RMSEA) is an index used to evaluate the model approximation error, where values below 0.05 suggest a good fit. Comparative fit index (CFI) and Tucker–Lewis index (TLI) are commonly used to assess goodness-of-fit of a model, with values of 0.95 or higher considered good fits. Standardized root mean square residual (SRMR) measures the average of standardized residuals and is acceptable if less than 0.08. Akaike's information criterion (AIC) and Bayesian information criterion (BIC) are indices employed to compare two or more models based on the same criteria. The lower the values, the better the model fit. Based on these indices, the factor loadings and error variances in Tables 1 and 2 appear to follow reliable tendencies, despite the significant multicollinearity. Consequently, the relationship between δ and T_F was explained by the small error variance, whereas other combinations do not show the same reliability. In the following analysis, we will only use the datasets for δ, as the AIC and BIC values for δ are smaller than those for ψ.

Strong multicollinearity complicates causal inference by leaving backdoor pathways open, making it challenging to accurately estimate the average treatment effect. To mitigate this issue, orthogonalizing each explanatory variable via PCA is essential. However, conventional four-variable PCA often results in loss of intuitive understanding due to variable mixing. Therefore, this study employs sparse PCA to maintain explanatory power. Given the near-

identical δ distributions at the four T s, the simplest combination of the datasets, $\delta(14.1\,K)$ and $\delta(40.0\,K)$, was selected for analysis. As shown in Figure 4(a), the importance difference between $\delta(40.0\,K)$ and $\delta(130.9\,K)$ is less than 0.1, favoring the dataset with temperatures close to T_{F}. Sparse PCA uses iterative optimization with lasso regression, introducing an $L1$ norm as a regularization term. Conventionally, the penalty coefficient should be optimized by cross-validation. However, in case of two variables, the outcome is trivial, and the following linear transformation is used:

$$PC1 = \frac{\delta(14.1\,K) + \delta(40.0\,K)}{\sqrt{2}}, \tag{3}$$

$$PC2 = \frac{\delta(14.1\,K) - \delta(40.0\,K)}{\sqrt{2}}, \tag{4}$$

where the principal components $PC1$ and $PC2$ are the orthogonal basis functions. Sparse PCA yielded results identical to the direct linear transformation, except for cases where the penalty for the Lasso regularization is $\alpha > 10$. Although sparse PCA ensures orthogonality and reduces multicollinearity, it does not necessarily preserve the direct physical interpretability of the original variables. In this study, sparse PCA was used primarily for mathematical transformation rather than for organizing physical quantities. Nevertheless, the resulting principal components can be associated with meaningful physical phenomena. $PC1$, which reflects the overall magnitude of δ, is influenced by the concentration of stress/strain, while $PC2$, which is the difference of δ between 14.1 K ($< T_{\mathrm{F}}$) and 40.0 K ($> T_{\mathrm{F}}$), reflects the amount of spontaneous polarization due to the photoelastic effect. Figure 7 shows the spatial distributions of $PC1$ and $PC2$, both of which reveal stripe-like structures.

Figure 8 and Tables 4 and 5 present the SEM estimation results for each cluster, analyzing the contributions of $PC1$ and $PC2$ to T_{F} and T_{c}. Table 6 summarizes the model's goodness-of-fit indices. As $PC1$ and $PC2$ are orthogonal, their covariance is zero, eliminating the need for a bidirectional arrow to represent their relationship. This orthogonality allows independent estimation of factor loadings on T_{F} and T_{c}. Consequently, the factor loadings associated with T_{F} remain unchanged even when the objective variable is restricted to T_{F} alone (see the Supplementary Materials). The goodness-of-fit indices indicate a well-fitting model across clusters, and the AIC and BIC values confirm significant improvements in model performance due to the orthogonalization, even with fewer explanatory variables. The sum of the squared factor loadings and the error variance are equal to 1, indicating that the models fully explain the variance. However, large error variances associated with T_{c} indicate that $PC1$ and $PC2$ do not adequately capture its variance, suggesting the influence of other factors outside the model. Furthermore, the factor loadings from T_{F} to T_{c} are small, with some p values exceeding 0.05, highlighting distinct mechanisms governing T_{F} and T_{c} distributions. Based on these results, this study focuses on T_{F}, which is more reliably explained by $PC1$ and $PC2$. $PC1$ exhibits larger factor loadings in E1 and E2, while PC2 dominates in E3 and E4, although this result only reflects the correlation.

4.3 Causal inference in statistics

For causal inference in statistics using the DoWhy library, a DAG (Figure 9) was constructed based on the SEM diagrams in Figure 8. Prior to analysis, the $PC1$, $PC2$, and T_F datasets were standardized to eliminate scale differences between variables and improve model stability. In this setup, $PC1$ and $PC2$ were the treatment variables, while T_F was the outcome variable. The orthogonality of $PC1$ and $PC2$, achieved via Equations (3) and (4), inherently satisfies the backdoor criterion. This eliminates the need for additional covariates to block potential confounding effects. Consequently, the "identify_effect", "estimate_effect", and "refute_estimates" functions from the DoWhy library were employed to estimate and validate the average treatment effect (ATE). The "backdoor.linear_regression" function was specified for "estimate_effect". In this linear regression model, the variable ATE is expressed as

$$ATE = E(Y_1 - Y_0), \tag{5}$$

where Y_1 and Y_0 are the expected outcomes when the treatment is applied and not applied, respectively. For standardized datasets, the ATE estimates represent the expected change in standard deviation units of T_F for a one-standard-deviation increase in either $PC1$ or $PC2$, assuming no interaction effects between the treatment variables. However, to capture nonlinear relationships, potential interaction terms $(PC1 \times PC2)$ are also discussed below.

Figure 9 shows the DAGs, and Table 7 summarizes the results of these analyses. However, the error variances are not explicitly estimated. As the input data were standardized, if the sum of ATE s equals 1, error variances do not contribute to the residual effect. Conversely, a sum less than 1 suggests contributions from measurement error and/or unobserved confounding factors. In Table 7, the sum of ATE s in E3 is exceeds 1, even within the 95 % confidence interval. This suggests that the effects of $PC1$ and $PC2$ on T_F are not entirely additive. Although ATE provides an average population-level treatment effect, it does not account for the interaction effects variables. In this case, $PC1$ and $PC2$ are assumed to affect T_F independently, while also interacting with each other. This interaction implies that the effect of $PC1$ on T_F may depend on the value of $PC2$, and vice versa. To capture these effects, the conditional average treatment effect (CATE) was introduced. $CATE$ extends ATE by evaluating the treatment effect conditional on specific values of the covariates X. For instance, $CATE(X)$ quantifies the treatment effect on T_F, given particular values of the covariates X. This approach provides a more nuanced understanding of the relationship. The inclusion of $CATE(X)$ is important to accurately model interaction dynamics that cannot be captured by ATE alone. To represent these interactions, the models are modified as follows:

$$T_F \sim PC1 + (PC1 \times PC2), \tag{6}$$

for the $CATE$ of $PC1$, and

$$T_F \sim PC2 + (PC1 \times PC2), \tag{7}$$

for the $CATE$ of $PC2$. These models allow the evaluation of nonlinear relationships, where the

effect of $PC1$ depends on $PC2$, and vice versa. Using $CATE(X)$, ATE for $PC1$ can be redefined as

$$ATE = E(CATE(X)), \tag{8}$$

where

$$CATE(X) = E(Y_1 - Y_0 \mid PC2 = X). \tag{9}$$

$CATE(X)$ can be similarly described for $PC2$. Table 8 summarizes the results of $CATE(X)$ under different conditions for $PC1$ or $PC2$. The $CATE(X)$ of $PC1$ increases as $PC2$ increases, and vice versa. The model described in Equations (6) and (7) includes the ATE estimates of $PC1$ and $PC2$ separately, as well as the interaction term ($PC1 \times PC2$) in both. Consequently, summing the two ATEs results in double-counting the interaction term, overestimating the total effect of $PC1$ and $PC2$ on T_F.

The reliability of the evaluation results is assessed by checking whether the DAG models constructed in Figure 9 are sensitive to spurious causal relationships. The "dummy_outcome_refuter" function from the DoWhy library was employed. This ensures that the model is free from spurious relationships and captures true causal effects. By design, the random dummy outcome has no causal relationship with the treatment. If the estimated effect for the dummy outcome is close to zero, and the p value is high ($p \sim 1$), the null hypothesis cannot be rejected, supporting the conclusion that no spurious causal relationships exist in this model. Table 9 summarizes the results of the dummy outcome tests and the p values. Across all cases, the estimated values are close to zero, and the p values are close to 1, indicating that these models are robust and do not detect any spurious causal relationships. Comparisons of DAG diagrams (Figure 9) and SEM diagrams (Figure 8) reveal near-identical structures for each cluster. Both approaches indicate that $PC1$ contributes more significantly to T_F in E1 and E2, whereas $PC2$ dominates in E3 and E4. To simplify the discussion below, E1 and E2 are merged into a single cluster labeled E12, while E3 and E4 are grouped into E34.

Additionally, to assess the impact of potential unobserved confounding factors, the "random_common_cause" function from the DoWhy library was utilized. Compared with the "dummy_outcome_refuter" function, this method evaluates the sensitivity of the estimated causal effect to the inclusion of an unobserved common cause. If the causal estimates remain stable after the addition of this random variable, the model is considered robust to potential bias from unobserved confounding factors. Prior to the analysis, the $PC1$, $PC2$, and T_F datasets for E12 and E34 were standardized to ensure consistency and comparability. The refuter introduces a random variable, called "random common cause", sampled from a normal distribution $N(0,1^2)$. This variable simulates unobserved confounding factors, allowing for the evaluation of robustness of causal inference. Intensity of the randomness introduced by the refuter is appropriate for simulating realistic potential bias while maintaining the scale consistency of the standardized input dataset.

To address the sample size imbalance between E12 (10,571) and E34 (31,709), the bootstrap method was employed. Specifically, 2,000 samples were drawn with replacement from each group, and the ATE influenced by the random variable was calculated. This process was repeated 10,000 times, generating a distribution of ATE ratios (the ATE with the random

variable divided by the *ATE* without it). To detect statistical differences in the *ATE* ratio distribution, a smaller number of iterations, such as 1,000, can suffice for convergence; however, a larger number of iterations (10,000 iterations) was chosen as a conservative measure to ensure robust and reliable results, even though it would take a few days for computation. This decision reflects a trade-off between computational cost and statistical confidence, with the goal of minimizing the risk of underestimating the effects due to insufficient sampling. Figure 10 presents histograms of the *ATE* ratios for E12 and E34, which converge closely around 1. Furthermore, based on the histograms of p, values are dispersed between 0.8 and 1; thus, the null hypothesis cannot be rejected. This indicates that the inclusion of the "random common cause" has a negligible effect on the *ATE* estimates. The stability of these ratios across iterations underscores the robustness of the causal inference framework, confirming that the influence of unobserved confounding factors is minimal.

Given the robust regression models derived from this straightforward model, the next phase of analysis focuses on understanding why *PC*1 dominates in E12, while *PC*2 dominates in E34. According to the box-and-whisker plot of T_F (see the Supplementary Materials), the T_F values for E12 are distributed toward higher T values compared to E34. Furthermore, in Figure 2(b), the $\delta(T)$ curves for E1 and E2 consistently exhibit values approximately 20 nm higher than those for E3 and E4 across the entire T range. To analyze these differences, the 20-nm increase in δ is defined as the treatment effect, designating E12 as the treated group and E34 as the untreated (control) group. Using the T-Learner framework, the latent scores of the treated E12 group and untreated E34 group are directly compared to gain deeper insights into the variations in treatment effects between these groups[73].

In this study, the T-Learner framework was employed to quantitatively assess the increase in T_F at individual pixels using different regression models. The dataset was split into two groups, E12 and E34, to train separate models for each group for predicting outcomes for the treated (Y_1) and untreated (Y_0) groups based on the observed δ values. The treatment effect for each pixel was estimated as the difference between these predicted outcomes $(Y_1 - Y_0)$. Further, by aggregating these individual estimates, the ATE, ATE on the treated (ATT), and ATE on the untreated (ATU) values were calculated. To preserve the original scale of the T_F dataset and quantitatively evaluate its absolute change, the datasets for $\delta(14.1K)$, $\delta(40.0K)$, and T_F are used without preprocessing. This contrasts with the typical approach where input datasets are standardized to facilitate model training. As noted in the DAG calculations in Figure 9, sparse PCA was used to orthogonalize the predictors and prevent the backdoor pathway from remaining open due to multicollinearity, which is crucial for valid causal inference within the DAG framework. In contrast, the T-Learner framework focuses on estimating the difference in treatment effects predicted by regression models for the treated and untreated groups. Although multicollinearity may affect the stability of individual regression models, their impact on the difference in predicted outcomes is typically minimal. Furthermore, the use of machine learning models that are inherently robust to multicollinearity further mitigates any potential issues, ensuring accurate and reliable estimation of treatment effects.

Bootstrapping was employed to ensure the robustness of the treatment effect estimates. In each iteration, 2,000 samples were drawn with replacement from both E12 and E34, and the estimated T_F, corresponding to the observed δ increase of approximately 20 nm, was computed from the bootstrapped subsampled data. Separate regression models were trained on these

datasets to predict outcomes. This process was repeated 10,000 times to generate treatment effect distributions, enabling the calculation of confidence intervals to assess the reliability of the results. Although the 10,000 bootstrap iterations required several days of computation, their statistical reliability was prioritized. Each dataset was split into 70 % training and 30 % test sets using the "train_test_split" function, separately for E12 and E34, to ensure independent model evaluation. The T-Learner framework integrated five regression models to capture the relationship between the predictors, $PC1$ and $PC2$, and the outcome, T_F. Linear regression (LR), implemented using the "LinearRegression" function from the scikit-learn library, assumed a linear relationship between the predictors and the outcome with default parameters. Support vector machine regression (SV), using the "SVR" function with a linear kernel (kernel='linear'), was employed to capture linear relationships. Random forest regression (RF), implemented with the "RandomForestRegressor" function, utilized 180 estimators to model complex, nonlinear interactions, with unrestricted tree depth and minimum sample requirements set to two for splitting and one for leaf nodes. Gradient boosting regression (GB), using the "GradientBoostingRegressor" function, was configured with 100 estimators, a learning rate of 0.1, and a maximum tree depth of three to balance complexity and overfitting. Finally, a neural network (NN) was implemented using the "MLPRegressor" function, which included two hidden layers of 100 units each to capture highly nonlinear relationships. The NN training was limited to 500 iterations, with early stopping enabled to prevent overfitting when validation performance ceased to improve. Regarding multicollinearity of the input datasets, the LR and SV methods may be affected, whereas the RF and GB methods are more robust. Therefore, considering the characteristics and strengths of each model is essential.

Figure 11 illustrates the effect sizes estimated by the T-Learner across five regression models, presented as box-and-whisker plots. The estimated ATE (ATE^{Est}) in Equation (5) represents the average effect of increasing δ by approximately 20 nm across the entire dataset, covering both E12 and E34. The estimated ATT (ATT^{Est}) represents the average treatment effect within the treated E12 group and is typically expressed as

$$ATT = E(Y_1 - Y_0 \mid Z = 1), \tag{10}$$

where $Z = 1$ indicates that the data belong to the treated group. This conditional expectation quantifies the expected difference in outcomes for individuals in E12 when the treatment is applied versus when it is not. Conversely, the estimated ATU (ATU^{Est}) corresponds to the treatment effect in the untreated E34 group using counterfactual thinking and is generally expressed as

$$ATU = E(Y_1 - Y_0 \mid Z = 0), \tag{11}$$

where $Z = 0$ indicates that the data belong to the untreated group. Thus, ATU^{Est} quantifies the expected difference in outcomes for individuals in E34 if the treatment were hypothetically applied compared to no treatment.

Figure 11 reveals slight differences between the estimated effect sizes depending on the regression models used. The differences between ATT^{Est} and ATU^{Est} can be explained by the linear and nonlinear models. In linear models, $ATT^{Est} > ATU^{Est}$ suggests that E12 is more likely to benefit from increasing δ. For E34, the benefit of increasing δ was evaluated using

counterfactual thinking, which indicates that E34 is less likely to experience a significant increase in T_F compared to E12. This aligns with the general observation that untreated groups, typically composed of individuals less likely to receive treatment, tend to underestimate treatment effects. Furthermore, counterfactual estimates for the untreated group are often sensitive to model assumptions, implying that their flexibility is limited in the linear models: the treatment effect is restricted to a monotonic change in T_F. In contrast, nonlinear models exhibit more flexibility in capturing the mechanisms driving the increase in T_F with increasing δ, as indicated by $ATT^{Est} \simeq ATU^{Est}$.

The relationships between ATE^{Est}, ATT^{Est}, and ATU^{Est} are further analyzed, with a focus on the differences across models. Weighted ATE (ATE^{W}) is calculated as the weighted arithmetic mean of ATT^{Est} and ATU^{Est}:

$$\begin{aligned} ATE^{W} &= \frac{w_{ATT} \times ATT^{Est} + w_{ATU} \times ATU^{Est}}{w_{ATT} + w_{ATU}} \\ &= \frac{w_{ATT} \times E(Y_1 - Y_0 \mid Z = 1)}{w_{ATT} + w_{ATU}} + \frac{w_{ATU} \times E(Y_1 - Y_0 \mid Z = 0)}{w_{ATT} + w_{ATU}} \end{aligned} \quad (12)$$

$$\begin{aligned} &= E(Y_1) - E(Y_0) \\ &= E(Y_1 - Y_0) \\ &= ATE^{Est}, \end{aligned} \quad (13)$$

where the weights w_{ATT} and w_{ATU} correspond to the number of pixels in E12 and E34, respectively, i.e., $w_{ATT} = 10,571$ and $w_{ATU} = 33,709$. To transition from Equation (12) to Equation (13), satisfying the conditions of "unconfoundedness", "positivity assumption", and "stable unit treatment value assumption (SUTVA)" is necessary, as mentioned later[1]. To verify the consistency of ATE^{Est} with ATE^{W}, the difference between them is defined as

$$\Delta ATE = ATE^{Est} - ATE^{W}. \quad (14)$$

From the 10,000 bootstrapped datasets, the 95 % confidence interval for ΔATE is clarified to determine if it included zero. Solely relying on the arithmetic mean of the five regression models risks losing valuable information from each model. To overcome this, hierarchical Bayesian estimation was applied. This approach captures both the variability between models and the uncertainty within each model, offering more robust and interpretable estimates. In this framework, for each model, the treatment-effect estimates, ATE_i^{Est}, ATT_i^{Est}, and ATU_i^{Est} (i = LR, SV, RF, GB, and NN), and their differences, $\Delta ATE_i = ATE_i^{Est} - ATE_i^{W}$, were treated as lower-level variables. The global means of these estimates, ATE_{global}^{Est}, ATT_{global}^{Est}, and ATU_{global}^{Est}, and their differences, $\Delta ATE_{global} = ATE_{global}^{Est} - ATE_{global}^{W}$, were modeled as upper-level variables. Hyperpriors were imposed at the upper level, defined as

$$ATE_{global}^{Est} \sim N(1.832, 0.5^2), \quad (15)$$

$$ATT_{\text{global}}^{\text{Est}} \sim \text{N}(1.049, 0.5^2), \tag{16}$$

$$ATU_{\text{global}}^{\text{Est}} \sim \text{N}(0.515, 0.5^2), \tag{17}$$

$$\Delta ATE_{\text{global}} \sim \text{N}(1.106, 0.5^2), \tag{18}$$

where the initial means of the normal distribution were set to the arithmetic means from the dataset. Compared to the ATE^{Est}, ATT^{Est}, and ATU^{Est} distributions in Figure 11, the standard deviations are set to a sufficiently large values. The relationships between the upper and lower levels is modeled as follows:

$$ATE_i^{\text{Est}} \sim \text{N}(ATE_{global}^{\text{Est}}, \tau^2), \tag{19}$$

$$ATT_i^{\text{Est}} \sim \text{N}(ATT_{\text{global}}^{\text{Est}}, \tau^2), \tag{20}$$

$$ATU_i^{\text{Est}} \sim \text{N}(ATU_{\text{global}}^{\text{Est}}, \tau^2), \tag{21}$$

$$\Delta ATE_i \sim \text{N}(\Delta ATE_{\text{global}}, \tau^2), \tag{22}$$

$$\tau \sim \text{N}(0, 0.5^2), \tag{23}$$

$$\sigma_i \sim \text{N}(0, 0.5^2), \tag{24}$$

where τ and σ_i represent the error terms for the upper and lower levels, respectively, with large standard deviations. Finally, the likelihood for optimization is calculated as

$$ATE_{\text{global}}^{\text{Est}} \sim \text{N}(ATE_i^{\text{Est}}, \sigma_i^2), \tag{25}$$

$$ATT_{\text{global}}^{\text{Est}} \sim \text{N}(ATT_i^{\text{Est}}, \sigma_i^2), \tag{26}$$

$$ATU_{\text{global}}^{\text{Est}} \sim \text{N}(ATU_i^{\text{Est}}, \sigma_i^2), \tag{27}$$

$$\Delta ATE_{\text{global}} \sim \text{N}(\Delta ATE_i, \sigma_i^2), \tag{28}$$

where ATE_i^{Est}, ATT_i^{Est}, and ATU_i^{Est} correspond to the 10,000 data points shown in Figure 11. The Bayesian estimation was performed using MCMC with 16 chains, each running for 11,000 iterations, including a 1,000-iteration warm-up period. Convergence of the MCMC simulations was assessed via trace plots and Gelman–Rubin statistics[74]. The trace plots demonstrated effective mixing and overlap between the chains. The Gelman–Rubin diagnostic value ($\hat{R}$) was 1.00 for all parameters, confirming satisfactory convergence between the chains. Pooling the 160,000 iterations yielded the mean and 95 % Bayesian confidence interval for the parameters, as summarized in Table 10. These results indicate minimal discrepancies between the models for ATE_i^{Est}, but significant differences between the linear and nonlinear models for ATT_i^{Est} and ATU_i^{Est}. From the 160,000 iterations, $\Delta ATE_{\text{global}} = 1.19$ K [0.99 K, 1.40 K] was obtained and $ATE_{\text{global}}^{\text{W}} = 0.66$ K [0.49 K, 0.83 K] was recalculated using the paired values of $ATT_{\text{global}}^{\text{Est}}$ and $ATU_{\text{global}}^{\text{Est}}$.

As explained in the Supplementary Materials, the difference in the means of T_{F} for E12 and E34 was estimated as 1.85 K [1.82 K, 1.87 K] using hierarchical Bayesian estimation. This

value aligns well with $ATE_{\text{global}}^{\text{Est}} = 1.83$ K [1.63 K, 2.04 K]. Explaining the fundamental assumptions for causal inference is also necessary to describe the reason for this agreement as mentioned above[1]. The first "unconfoundedness" assumption posits that all confounding factors have been observed and appropriately adjusted such that treatment assignment is random after conditioning on the observed covariates. This assumption implies the absence of unobserved confounding factors that can simultaneously affect both the treatment and the potential outcomes (Y_1 and Y_0). To evaluate this, sensitivity analysis was conducted using a hypothetical unobserved common cause. The results showed that even under the presence of such a factor, the causal inferences remained stable. This supports the validity of the unconfoundedness assumption. The second "positivity assumption" is defined as

$$0 < P(Z = 1 | X) < 1 \quad \forall X. \tag{29}$$

This assumption ensures that for all the covariates X, sufficient samples exist in both the treated and untreated groups for estimating causal effects. As shown in Table 8, the *CATE* estimates remain stable over various covariates X for the standardized input variables. This result indicates the validity of the "positive assumption". The final 'SUTVA" ensures that the treatment effect for each individual is unaffected by the treatment assignment of others and that the observed outcomes Y directly correspond to the potential outcomes Y_1 and Y_0. In this analysis, the treatment effect–-defined as an increase in δ by approximately 20 nm–-was applied uniformly across all pixels. Experimentally, the values of δ were measured simultaneously and recorded independently for each pixel, ensuring statistical consistency of the treatment definition and minimizing interference between pixels. Therefore, it is reasonable to assume that 'SUTVA" is satisfied. Given the three assumptions, the difference in the means of the observed outcomes between the treated and untreated groups can be expressed as

$$\begin{aligned} & E(Y | Z = 1) - E(Y | Z = 0) = E(Y_1 | Z = 1) - E(Y_0 | Z = 0) \\ & = E(Y_1) - E(Y_0) \\ & = E(Y_1 - Y_0) \\ & = ATE^{\text{Est}}, \end{aligned} \tag{30}$$

where Y is the observed outcome, as previously mentioned. Consequently, the consistency of $ATE_{\text{global}}^{\text{Est}}$ implies that there are no significant unobserved confounding factors and that the T-Learner model is appropriately constructed. Similarly, the value of $ATE_{\text{global}}^{\text{W}}$ should also be equal to $ATE_{\text{global}}^{\text{Est}}$ in Equations (12) and (13); however, $\Delta ATE_{\text{global}} > 0$ was observed. This discrepancy suggests that counterfactual estimates for the untreated group do not adequately capture unobserved factors present in the treated group.

The origin of $\Delta ATE_{\text{global}} > 0$ is explained as follows. The ATE_i^{Est} value includes the full dataset (E12 + E34) and captures both the direct treatment effect–-which increases δ by approximately 20 nm–-and unobserved factors influencing the treatment effect that cannot be explicitly modeled by δ. In contrast, ATT_i^{Est}, estimated only in E12, reflects both the direct

treatment effect and the influence of these unobserved factors. For ATU_i^{Est}, however, the estimation process significantly differs. Counterfactual scenarios for ATU_i^{Est} in E34 solely account for the direct effect of increasing δ by approximately 20 nm, neglecting unobserved factors. Consequently, the linear models, $ATU_{\text{LR}}^{\text{Est}}$ and $ATU_{\text{SV}}^{\text{Est}}$, show a small increase in T_{F}. In contrast, the nonlinear models, $ATU_{\text{RF}}^{\text{Est}}$ and $ATU_{\text{GB}}^{\text{Est}}$, demonstrate greater robustness by redistributing the unobserved treatment effects more equitably between the treated and untreated groups, reducing the disparity between ATT_i^{Est} and ATU_i^{Est}. This underscores the advantage of nonlinear models in addressing biases inherent in linear approaches, making them particularly valuable for causal inference. Therefore, $\Delta ATE_{\text{global}} > 0$ arises from the presence of unobserved treatment effects unique to E12. In contrast, the consistency of ATE_i^{Est} across models further strengthens its reliability as a measure of the overall treatment effect.

Considering this, to reexamine why the observed difference in the means of T_{F} for E12 and for E34 aligns with $ATE_{\text{global}}^{\text{Est}}$, consider the components of the observed outcome Y at E12, which includes both the observed and unobserved treatment effects. Therefore, the difference between the expected outcomes at E12 and at E34 can be expressed as

$$\begin{aligned} &E(Y \mid Z=1) - E(Y \mid Z=0) = E(Y_1^{obs}) + E(Y_1^{unobs}) - E(Y_0^{obs}) \\ &= E(Y_1) - E(Y_0) \\ &= ATE^{\text{Est}} \end{aligned} \tag{31}$$

where the superscript "obs" refers to the treatment effects observed in the causal inference, corresponding to the treatment effect of increasing δ by approximately 20 nm. The superscript "unobs" represents the unobserved treatment effect that is only present in E12. As unobserved effects do not exist for E34, they do not contribute to the outcomes for the untreated group. As expressed in Equation (31), the difference in the means for E12 and E34 measures the total treatment effect, including unobserved treatment effect, and is equal to the value of ATE^{Est}. Additionally, ATT^{Est} in ATE^{W} in Equation (12) accounts for unobserved treatment effect, whereass ATU^{Est} does not. Using $w_{\text{ATT}} = 10,571$ and $w_{\text{ATU}} = 33,709$, the corrected weighted arithmetic mean (ATE^{CW}) can be rewritten as

$$\begin{aligned} ATE_{\text{global}}^{\text{W}} &= \frac{w_{\text{ATT}} \times E(Y_1^{\text{obs}} + Y_1^{\text{unobs}} - Y_0 \mid Z=1)}{w_{ATT} + w_{\text{ATU}}} + \frac{w_{ATU} \times E(Y_1^{\text{obs}} - Y_0 \mid Z=0)}{w_{\text{ATT}} + w_{\text{ATU}}} \\ &= ATE_{\text{global}}^{\text{CW}} + \frac{10,571}{42,280} \times E(Y_1^{\text{unobs}}), \end{aligned} \tag{32}$$

$$\begin{aligned} ATE_{\text{global}}^{\textit{accounts for unobserved treatment effect}} &= ATE_{\text{global}}^{\text{W}} + E(Y_1^{\text{unobs}}) \\ &= ATE_{\text{global}}^{CW} + 1.250E(Y_1^{\text{unobs}}). \end{aligned} \tag{33}$$

Here, $E(Y_1^{\text{unobs}})$ corresponds to $\Delta ATE_{\text{global}}$, representing the contribution of the unobserved treatment effect. This unobserved effect is specific to E12 and does not include any effects

observable in E34. As previously mentioned, linear models underestimate ATU^{Est} under the assumption that the unobserved treatment effect Y_1^{unobs} is localized in E12. In contrast, the nonlinear models provide more flexible estimations incorporating indirect information exchange between E12 and E34, allowing for a small influence of Y_1^{unobs} on ATU^{Est}. This difference is evident in the box-and-whisker plots shown in Figure 11, where the whiskers for ATU^{Est} are slightly longer in the nonlinear models than in the linear models. This suggests that the nonlinear models capture a broader range of possible effects, including small contributions from Y_1^{unobs}. Biases inherent to different models are corrected using the hierarchical Bayesian estimation, resulting in a reliable estimate of $ATU_{\mathrm{global}}^{\mathrm{Est}}$ based on the observed data from E34. This integration balances the underestimation bias of linear models with the flexibility of nonlinear models, ensuring that $ATU_{\mathrm{global}}^{\mathrm{Est}}$ robustly reflects the characteristics of E34. Consequently, the assumption that the unobserved treatment effect does not directly influence E34 is statistically supported. From this discussion, the corrected treatment effect of increasing δ by approximately 20 nm between E12 and E34 increases T_{F} by $ATE_{\mathrm{global}}^{\mathrm{CW}}$ = 0.36 K [0.18 K, 0.54 K]. Additionally, the corrected unobserved treatment effect localized at E12 further increases T_{F} by $1.250E(Y_1^{\mathrm{unobs}}) = 1.49$ K [1.23 K, 1.75 K]. Together, these contributions result in a total increase of T_{F} at E12 by $ATE_{\mathrm{global}}^{\mathrm{Est}}$ = 1.83 K [1.63 K, 2.04 K] compared to E34.

Based on the data-driven analysis presented above, we conclude that the five regression models consistently suggest the presence of an unobserved treatment effect. Our analysis indicates that this effect is likely related to the flexoelectric effect. Specifically, our experimental results show that uniform stress/strain is greater in E12 than in E34, resulting in an enhanced photoelastic effect that produces an approximately 20 nm larger δ in E12. This 20 nm increase in δ is associated with a 0.36 K [0.18 K, 0.54 K] higher T_{F} in E12 compared to E34. In addition, localized stress/strain concentrations in E12 generate strain gradients that appear to be associated with additional polarization changes via the flexoelectric effect, resulting in an additional T_{F} increase of 1.49 K [1.23 K, 1.75 K]. Although a direct causal relationship between the flexoelectric effect and the observed increase has not been definitively established, our statistical analysis provides strong evidence of a potential causal relationship, making the flexoelectric effect the most plausible candidate to explain the observed increase. According to Ref. [60], the K-shape clustering criterion is based on the shapes of the $\delta(T)$ and $\psi(T)$ curves. This suggests that the treatment effect related to the flexoelectric effect may be better explained by ψ, as ψ reflects the direction of stress, strain and polarization, which are highly related to strain gradients[63–67]. Extending the DAG analysis to three or more variables is necessary to validate this hypothesis and explore more advanced treatment methods in the future. Finally, as reported in Ref. [14], the flexoelectric effect occurs in $T_{\mathrm{F}} < T < T_{\mathrm{c}}$. In this study, the distribution of T_{F} and the stress-induced ferroelectric state were analyzed using the $\delta(14.1\,K)$ and $\delta(40.0\,K)$ datasets. However, direct comparisons were not possible. Therefore, analyzing the $\delta(90.0\,K)$, $\delta(130.9\,K)$, $\psi(90.0\,K)$, and $\psi(130.9\,K)$ datasets is necessary. The distribution of T_{F} solely driven by the piezoelectric effect is crucial to explain the uniform distribution of T_{c}.

5. Conclusion

This study demonstrated the utility of integrating causal inference with SEM to analyze birefringence imaging datasets of stress-induced ferroelectric $SrTiO_3$. Through random forest analysis, $\delta(14.1K)$ was identified as the key variable influencing T_F, contributing approximately 60 % to its prediction. Spatial distribution patterns in the residuals revealed distinct mechanisms governing T_F in E12 and E34. SEM analysis indicated that T_F could be explained by δ, while T_c remained inadequately captured due to large error variances, suggesting the influence of unmodeled factors on its spatial distribution. To address the multicollinearity, sparse PCA was applied to $\delta(14.1K)$ and $\delta(40.0K)$, yielding two independent components: $PC1$ (stress/strain magnitude) and $PC2$ (polarization change). Orthogonalization improved the explanatory power and analytical stability, providing clearer insights into the distinct roles of uniform and localized stress/strain on T_F. DAGs were constructed to explore the causal paths from $PC1$ and $PC2$ to T_F. The orthogonalization satisfied the backdoor criterion, ensuring causal effects' unbiased estimation. The DAG analysis revealed that $PC1$ predominantly influenced T_F in E12, while $PC2$ was the dominant factor for E34. This distinction suggests that different mechanisms govern T_F at E12 and E34. Based on T-Learner analysis, the increasing δ by approximately 20 nm between E12 and E34 raised T_F by 0.36 K [0.18 K, 0.54 K], attributed to the piezoelectric effect. Furthermore, the unobserved treatment effect exclusive to E12, which is most plausibly attributed to the flexoelectric effect, further increased T_F by 1.49 K [1.23 K, 1.75 K].

Although causal inference methods cannot directly identify unobserved physical processes, they excel in uncovering hidden data regularities. In this study, we combined causal inference and SEM with domain-specific knowledge to analyze the relationship between δ s and T_F in the stress-induced ferroelectric $SrTiO3$. While traditional methods might attribute the entire 1.83-K increase in T_F to the 20-nm difference in δ, our approach disentangled the contributions of piezoelectricity (0.36 K) and flexoelectricity (1.49 K), revealing the dominant role of the flexoelectric effect. To address the inherent mixing of effects in real-world data, we employed hierarchical Bayesian estimation to correct biases between different models and extract robust estimates of treatment effects. This approach validated our analytical assumptions and demonstrated its capability to handle complex datasets where phenomena are not perfectly separable. By quantifying the contributions of the piezoelectric and flexoelectric effects, our study underscores the potential of integrating advanced statistical methods with phenomenological insights to achieve a deeper understanding of physical phenomenon.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was partially supported by a Grant-in-Aid for Scientific Research (KAKENHI) (Grant Numbers JP21K04897 and JP23K03283) from the Japan Society for the Promotion of Science.

References

[1] Imbens GW, Rubin DB. Causal inference for statistics, social, and biomedical sciences: an introduction. Cambridge University Press; 2015.

[2] Pearl J, Glymour M, Jewell NP. Causal inference in statistics: a primer. Wiley; 2016.

[3] Pearl J, Mackenzie D. The book of why: the new science of cause and effect. Basic Books; 2018.

[4] Sharma A, Kiciman E. Dowhy: an end-to-end library for causal inference. arXiv preprint. 2020;.

[5] Kraev E, Flesch T, Lekunze HT, et al. Out-of-sample scoring and automatic selection of causal estimators. arXiv preprint. 2022;.

[6] Eberhardt F, Scheines R. Interventions and causal inference. Philos Sci. 2007;74(5):981–995.

[7] Bollen KA. Structural equations with latent variables. John Wiley & Sons, Inc.; 1989.

[8] Kline RB. Principles and practice of structural equation modeling, fifth edition (methodology in the social sciences). Guilford Press; 2023.

[9] Pearl J. The causal foundations of structural equation modeling. New York: Guilford Press; 2012.

[10] Ziatdinov M, Nelson CT, Zhang X, et al. Causal analysis of competing atomistic mechanisms in ferroelectric materials from high-resolution scanning transmission electron microscopy data. npj Comput Mater. 2020;6:127.

[11] Ting JYC, Barnard AS. Data-driven causal inference of process-structure relationships in nanocatalysis. Curr Opin Chem Eng. 2022;36:100818.

[12] Manaka H, Uetsubara K, Korogi S, et al. Microscopic observation of ferroelectric and structural phase transitions in $SrTiO_3$ under uniaxial stress using birefringence imaging techniques. J Phys Soc Jpn. 2022;91(8):084704.

[13] Manaka H, Uetsubara K, Miura Y. Stress-induced ferroelectricity in quantum paraelectric $SrTiO_3$ observed by birefringence imaging. JPS Conf Proc. 2023;38:011112.

[14] Zubko P, Catalan G, Buckley A, et al. Strain-gradient-induced polarization in $SrTiO_3$ single crystals. Phys Rev Lett. 2007;99:167601.

[15] Yasui K, Itasaka H, Mimura K, et al. Coexistence of flexo- and ferro-electric effects in an ordered assembly of $BaTiO_3$ nanocubes. Nanomaterials. 2022;12(2):188.

[16] Mashkevich VS, Tolpygo KB. Electrical, optical, and elastic properties of diamond-type crystals. Sov Phys JETP. 1957;3(5):435.

[17] Tolpygo KB. Long wavelength oscillations of diamond-type crystals including long range forces. Sov Phys JETP. 1963;7(4):1297.

[18] Kogan SM. Piezoelectric effect under an inhomogeneous strain and acoustic scattering of carriers in crystals. Sov Phys Solid State. 1964;5(10):2069–2070.

[19] Jiang X, Huang W, Zhang S. Flexoelectric nano-generator: materials, structures and devices. Nano Energy. 2013;2(6):1079–1092.

[20] Bhaskar UK, Banerjee N, Abdollahi A, et al. Flexoelectric mems: towards an electromechanical strain diode. Nanoscale. 2016;8(3):1293–1298.

[21] Catalan G, Lubk A, Vlooswijk AHG, et al. Flexoelectric rotation of polarization in ferroelectric thin films. Nat Mater. 2011;10(12):963–967.

[22] Lee D, Yoon A, Jang SY, et al. Giant flexoelectric effect in ferroelectric epitaxial thin films. Phys Rev Lett. 2011;107(5):57602.

[23] Ahn Y, Son JY. Flexoelectric effect via piezoresponse force microscopy of domain switching in epitaxial $PbTiO_3$ thin films. J Korean Ceram Soc. 2024;61:55–62.

[24] Wang S, Wang X, Tong W, et al. Microstructure designed flexoelectric materials and tip force for multifunctional applications. Nano Energy. 2025;133:110442.

[25] Tagantsev AK. Theory of flexoelectric effects in crystals. Zh Eksp Teor Fiz. 1985;88(6):2108–2122.

[26] Ma W, Cross LE. Observation of the flexoelectric effect in relaxor $Pb(Mg_{1/3}Nb_{2/3})O_3$ ceramics. Appl Phys Lett. 2001;78(19):2920–2921.

[27] Ma W, Cross LE. Large flexoelectric polarization in ceramic lead magnesium niobate. Appl Phys Lett. 2001;79(26):4420–44221.

[28] Ma W, Cross LE. Flexoelectric polarization of barium strontium titanate in the paraelectric state. Appl Phys Lett. 2002;81(18):3440–3442.

[29] Ma W, Cross LE. Strain-gradient-induced electric polarization in lead zirconate titanate ceramics. Appl Phys Lett. 2003;82(19):3293–3295.

[30] Ma W, Cross LE. Flexoelectric effect in ceramic lead zirconate titanate. Appl Phys Lett. 2005;86(7):72905.

[31] Ma W, Cross LE. Flexoelectricity of barium titanate. Appl Phys Lett. 2006;88(23):232902.

[32] Deng F, Deng Q, Shen S. A three-dimensional mixed finite element for flexoelectricity. J Appl Mech. 2018;85(3):031009.

[33] Zhou W, Chen P, Chu B. Flexoelectricity in ferroelectric materials. IET Nanodielectrics. 2019;2(3):83–91.

[34] Tian D, Jeong DY, Fu Z, et al. Flexoelectric effect of ferroelectric materials and its applications. Actuators. 2023;12(3):114.

[35] Cowley RA. Lattice dynamics and phase transitions of strontium titanate. Phys Rev. 1964;134:A981.

[36] Courtens E. Birefringence of $SrTiO_3$ produced by the 105° k structural phase transition. Phys Rev Lett. 1972;29:1380.

[37] Cowley RA. The phase transition of strontium titanate. Phil Trans R Soc A. 1996;354(1720):2799–2814.

[38] Blinc R, Žekš B. Soft modes in ferroelectrics and antiferroelectrics. North-Holland , Wiley; 1974.

[39] Mäller KA, Burkard H. $Srtio_3$: An intrinsic quantum paraelectric below 4 k. Phys Rev B. 1979;19:3593–3602.

[40] Shin D, Latini S, Schäfer C, et al. Quantum paraelectric phase of $srtio_3$ from first principles. Phys Rev B. 2021;104:L060103.

[41] Sawaguchi E, Kikuchi A, Kodera Y. Dielectric constant of strontium titanate at low temperatures. J Phys Soc Jpn. 1962;17(10):1666–1667.

[42] Hegenbarth E. Die feldstärkeabhängigkeit der dielektrizitätskonstanten von $SrTiO_3$-einkristallen im temperaturbereich von 15 bis 80° k. Phys Status Solidi. 1964;6:333.

[43] Fleury PA, Worlock JM. Electric-field-induced raman scattering in $SrTiO_3$ and $KTaO_3$. Phys Rev. 1968;174:613.

[44] Hemberger J, Lunkenheimer P, Viana R, et al. Electric-field-dependent dielectric constant and nonlinear susceptibility in $SrTiO_3$. Phys Rev B. 1968;174:613.

[45] Hemberger J, Nicklas R M Viana, Lunkenheimer P, et al. Quantum paraelectric and induced ferroelectric states in $SrTiO_3$. J Phys Condens Matter. 1996;8:4673.

[46] Manaka H, Nozaki H, Miura Y. Microscopic observation of ferroelectric domains in $SrTiO_3$ using birefringence imaging techniques under high electric fields. J Phys Soc Jpn. 2017;86(11):114702.

[47] Manaka H, Nozaki H, Miura Y. Development of birefringence imaging techniques under high electric fields. J Phys Conf Ser. 2018;969:012119.

[48] Müller KA, Berlinger W, Slonczewski JC. Order parameter and phase transitions of stressed $SrTiO_3$. Phys Rev Lett. 1970;25:734.

[49] Burke WJ, Pressley RJ. Stress induced ferroelectricity in $SrTiO_3$. Solid State Commun. 1971;9(3):191–195.

[50] Uwe H, Sakudo T. Stress-induced ferroelectricity and soft phonon modes in $SrTiO_3$. Phys Rev B. 1976;13(1):271–286.

[51] Fujii Y, Uwe H, Sakudo T. Stress-induced quantum ferroelectricity in $SrTiO_3$. J Phys Soc Jpn. 1987;56(6):1940–1942.

[52] Ohtomo A, Hwang HY. A high-mobility electron gas at the $LaAlO_3/SrTiO_3$ heterointerface. Nature. 2004;427:423–426.

[53] Honig M, Sulpizio JA, Drori J, et al. Local electrostatic imaging of striped domain order in $LaAlO_3/SrTiO_3$. Nat Mater. 2013;12:1112.

[54] Lee PW, Singh VN, Guo GY, et al. Hidden lattice instabilities as origin of the conductive interface between insulating $LaAlO_3$ and $SrTiO_3$. Nat Commun. 2016;7:12773.

[55] Su CP, Singh AK, Wu TC, et al. Impact of strain-field interference on the coexistence of electron and hole gases in $SrTiO_3/LaAlO_3/SrTiO_3$. Phys Rev Mat. 2019;3:075003.

[56] Haeni JH, Irvin P, Chang W, et al. Room-temperature ferroelectricity in strained $SrTiO_3$. Nature. 2004;430:758.

[57] Kim YS, Kim J, Moon SJ, et al. Localized electronic states induced by defects and possible origin of ferroelectricity in strontium titanate thin films. Appl Phys Lett. 2009;94(20):202906.

[58] Iglesias L, Sarantopoulos A, Magén C, et al. Oxygen vacancies in strained $SrTiO_3$ thin films: formation enthalpy and manipulation. Phys Rev B. 2017;95:165138.

[59] Toyoda K, Manaka H, Miura Y. Improvements of birefringence imaging techniques to observe stress-induced ferroelectricity in $SrTiO_3$ based on K-means clustering with circular statistics. Sci Technol Adv Mater Meth. 2023;3:2278322.

[60] Manaka H, Toyoda K, Miura Y. Multivariate temperature-series analysis of stress-induced ferroelectricity in $SrTiO_3$: a machine learning approach with K-shape clustering and hierarchical bayesian estimation. Sci Technol Adv Mater Meth. 2024;4:2342234.

[61] Azzam RM, Bashara NM. Ellipsometry and polarized light. Amsterdam: North Holland; 1977.

[62] Manaka H, Yagi G, Miura Y. Development of birefringence imaging analysis method for observing cubic crystals in various phase transitions. Rev Sci Instrum. 2016;87(7):073704.

[63] Manaka H, Tateishi K, Miura Y. Real-space imaging by magnetic birefringence for $KNiF_3$ under inhomogeneous stress. J Phys Soc Jpn. 2019;88(12):124702.

[64] Manaka H, Sasaki Y, Miura Y. Re-examination of successive structural phase transitions in $(C_3H_7NH_3)_2CuCl_4$ using birefringence imaging and electron paramagnetic resonance spectroscopy. J Phys Soc Jpn. 2017;86(11):114710.

[65] Miura Y, Okumura K, Fukuda T, et al. Observation of ferroelastic domains in layered magnetic compounds using birefringence imaging. J Phys Conf Ser. 2018;969:012153.

[66] Manaka H, Okumura K, Tokunaga K, et al. Observations of successive local-structure and ferroelectric phase transitions in $(C_2H_5NH_3)_2CuCl_4$ using birefringence imaging and electron paramagnetic resonance spectroscopy. J Phys Soc Jpn. 2022;91(11):114701.

[67] Miura Y, Ibushi R, Manaka H. Observation of ferroelectricity in two-dimensional antiferromagnet $(C_2H_5NH_3)_2CuCl_4$ using birefringence imaging techniques. J Phys Conf Ser. 2023;38:011142.

[68] Manaka H, Fukuda T, Miura Y. Birefringence imaging measurements on various structural phase transitions in $(C_nH_{2n+1}NH_3)_2MnCl_4$ with $n=1,2,$ and 3 using multiple wavelengths. J Phys Soc Jpn. 2016;85(12):124701.

[69] Molak A. Causal inference and discovery in Python. Packt Publishing; 2023.

[70] Casella G, Berger R. Statistical inference. Chapman and Hall/CRC; 2024.

[71] Moisen GG, Freeman EA, Blackard JA, et al. Predicting tree species presence and basal area in utah: A comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. Ecol Model. 2006;199(2):176–187.

[72] Natekin A, Knoll A. Gradient boosting machines, a tutorial. Front Neurorobot. 2013;7.

[73] Köunzel SR, Sekhon JS, Bickel PJ, et al. Metalearners for estimating heterogeneous treatment effects using machine learning. Proceedings of the National Academy of Sciences. 2019;116(10):4156–4165.

[74] Gelman A, Carlin J, Stern HS. Bayesian data analysis. Chapman and Hall/CRC; 2013.

Table 1. Calculated factor loadings based on structural equation modeling (SEM) in Figure 6. Relationships between (a) retardance (δ) and ferroelectric phase transition temperature (T_F), (b) fast-axis direction (ψ) and T_F, (c) δ and structural phase transition temperature (T_c), and (d) ψ and T_c.

Condition	Start $\rightarrow$ End	Factor loading	Standard error	p -value
(a)	$\delta(14.1\,K) \rightarrow T_F$	1.043	0.007	0.000
	$\delta(40.0\,K) \rightarrow T_F$	0.015	0.016	0.329
	$\delta(90.0\,K) \rightarrow T_F$	−0.459	0.014	0.000
	$\delta(130.9\,K) \rightarrow T_F$	0.119	0.007	0.000
(b)	$\psi(14.1\,K) \rightarrow T_F$	−0.459	0.030	0.000
	$\psi(40.0\,K) \rightarrow T_F$	0.163	0.023	0.000
	$\psi(90.0\,K) \rightarrow T_F$	0.331	0.014	0.000
	$\psi(130.9\,K) \rightarrow T_F$	−0.128	0.007	0.000
(c)	$\delta(14.1\,K) \rightarrow T_c$	−0.030	0.010	0.004
	$\delta(40.0\,K) \rightarrow T_c$	0.145	0.025	0.000
	$\delta(90.0\,K) \rightarrow T_c$	0.108	0.022	0.000
	$\delta(130.9\,K) \rightarrow T_c$	−0.003	0.011	0.800
(d)	$\psi(14.1\,K) \rightarrow T_c$	0.137	0.031	0.000
	$\psi(40.0\,K) \rightarrow T_c$	0.027	0.023	0.239
	$\psi(90.0\,K) \rightarrow T_c$	−0.095	0.014	0.000
	$\psi(130.9\,K) \rightarrow T_c$	0.037	0.007	0.000

Table 2. Error variance as an independent component of T_F and T_c based on SEM results in Figure 6, considering (a) δ and T_F, (b) ψ and T_F, (c) δ and T_c, and (d) ψ and T_c.

Condition	Variable	Variance	Standard error	p -value
(a)	T_F	0.384	0.003	0.000
(b)	T_F	0.967	0.007	0.000
(c)	T_c	0.951	0.007	0.000
(d)	T_c	0.987	0.007	0.000

Table 3. Model fit indices based on SEM results in Figure 6, analyzing relationships between (a) δ and T_F, (b) ψ and T_F, (c) δ and T_c, and (d) ψ and T_c.

Index	(a)	(b)	(c)	(d)
RMSEA	0.000	0.000	0.000	0.000
CFI	1.000	1.000	1.000	1.000
TLI	1.000	1.000	1.000	1.000
SRMR	0.000	0.000	0.000	0.000
AIC	296,681	358,171	335,013	359,048
BIC	296,811	358,301	335,143	359,178

Table 4. Calculated factor loadings for clusters E1–E4 based on SEM results in Figure 8.

Cluster	Start $\rightarrow$ End	Factor loading	Standard error	p -value
E1	$PC1 \rightarrow T_F$	0.734	0.009	0.000
	$PC1 \rightarrow T_c$	0.309	0.018	0.000
	$PC2 \rightarrow T_F$	0.057	0.009	0.000
	$PC2 \rightarrow T_c$	−0.076	0.012	0.000
	$T_F \rightarrow T_c$	0.146	0.018	0.000
E2	$PC1 \rightarrow T_F$	0.699	0.010	0.000
	$PC1 \rightarrow T_c$	0.255	0.019	0.000
	$PC2 \rightarrow T_F$	0.230	0.010	0.000
	$PC2 \rightarrow T_c$	0.038	0.014	0.008
	$T_F \rightarrow T_c$	0.072	0.020	0.000
E3	$PC1 \rightarrow T_F$	0.399	0.005	0.000
	$PC1 \rightarrow T_c$	0.179	0.010	0.000
	$PC2 \rightarrow T_F$	0.711	0.005	0.000
	$PC2 \rightarrow T_c$	−0.037	0.012	0.003
	$T_F \rightarrow T_c$	−0.034	0.014	0.013

E4	$PC1 \rightarrow T_F$	0.318	0.005	0.000
	$PC1 \rightarrow T_c$	0.121	0.009	0.000
	$PC2 \rightarrow T_F$	0.700	0.005	0.000
	$PC2 \rightarrow T_c$	−0.127	0.012	0.000
	$T_F \rightarrow T_c$	0.015	0.012	0.207

Table 5. Error variance as an independent component of T_F and T_c for clusters E1–E4 based on SEM results in Figure 8.

Cluster	Variable	Variance	Standard error	p -value
E1	T_F	0.458	0.009	0.000
	T_c	0.813	0.015	0.000
E2	T_F	0.458	0.009	0.000
	T_c	0.901	0.018	0.000
E3	T_F	0.335	0.004	0.000
	T_c	0.968	0.011	0.000
E4	T_F	0.409	0.005	0.000
	T_c	0.971	0.011	0.000

Table 6. Model fit indices for clusters E1–E4 based on SEM results in Figure 8.

Index	E1	E2	E3	E4
RMSEA	0.000	0.000	0.000	0.000
CFI	1.000	1.000	1.000	1.000
TLI	1.000	1.000	1.000	1.000
SRMR	0.000	0.000	0.000	0.000
AIC	25,809	24,293	71,291	76,235
BIC	25,855	24,339	71,345	76,288

Table 7. Calculated average treatment effect (ATE) for clusters E1–E4 based on directed acyclic graph (DAG) analysis results shown in Figure 9.

Cluster	Start → End	ATE	95.0% confidence interval	p -value
E1	$PC1 \rightarrow T_F$	0.735	[0.709, 0.756]	0.00
	$PC2 \rightarrow T_F$	0.239	[0.183, 0.301]	0.00
E2	$PC1 \rightarrow T_F$	0.683	[0.666, 0.704]	0.00
	$PC2 \rightarrow T_F$	0.161	[0.132, 0.191]	0.00
E3	$PC1 \rightarrow T_F$	0.379	[0.358, 0.403]	0.00

	$PC2 \rightarrow T_F$	0.722	[0.710, 0.735]	0.00
E4	$PC1 \rightarrow T_F$	0.307	[0.285, 0.336]	0.00
	$PC2 \rightarrow T_F$	0.687	[0.674, 0.702]	0.00

Table 8. Calculated conditional average treatment effect (CATE) for clusters E1–E4 based on DAG analysis results in Figure 9. Due to the analytical conditions, some half-open interval orientations differ; however, this does not affect the discussion.

Cluster	Start → End	*PC*2 interval condition	*CATE* for *PC*1	Start → End	*PC*1 interval condition	*CATE* for *PC*2
E1	$PC1 \rightarrow T_F$	[−15.409, −0.554)	0.629	$PC2 \rightarrow T_F$	[−3.558, −0.797)	0.029
		[−0.554, −0.119)	0.704		[−0.797, −0.416)	0.139
		[−0.119, 0.182)	0.738		[−0.416, 0.151)	0.213
		[0.182, 0.635)	0.769		[0.151, 0.878)	0.322
		[0.635, 2.159)	0.832		[0.878, 3.103)	0.493
E2	$PC1 \rightarrow T_F$	[−8.232, −0.807)	0.632	$PC2 \rightarrow T_F$	[−3.727, −0.752)	−0.002
		[−0.807, −0.204)	0.665		[−0.752, −0.372)	0.081
		[−0.204, 0.266)	0.685		[−0.372, 0.060)	0.140
		[0.266, 0.803)	0.703		[0.060, 0.618)	0.207
		[0.803, 3.739)	0.732		[0.618, 3.633)	0.381
E3	$PC1 \rightarrow T_F$	[−10.128, −0.865)	0.286	$PC2 \rightarrow T_F$	(−4.978, −0.714]	0.533
		[−0.865, −0.244)	0.339		(−0.714, −0.209]	0.655
		[−0.244, 0.115)	0.375		(−0.209, 0.198]	0.723
		[0.115, 0.687)	0.405		(0.198, 0.597]	0.780
		[0.687, 4.451)	0.490		(0.597, 6.232]	0.920
E4	$PC1 \rightarrow T_F$	[−10.655, −0.855)	0.115	$PC2 \rightarrow T_F$	(−7.709, −0.772]	0.540
		[−0.855, −0.182)	0.238		(−0.772, −0.300]	0.629
		[−0.182, 0.229)	0.311		(−0.300, 0.231]	0.682
		[0.229, 0.665)	0.369		(0.231, 0.798]	0.745
		[0.665, 5.583)	0.502		(0.798, 9.279]	0.839

Table 9. Random dummy outcomes for clusters E1–E4 based on DAG analysis results in Figure 9.

Cluster	Start → End	Estimate	*p* -value
E1	$PC1 \rightarrow T_F$	0.002	0.82
	$PC2 \rightarrow T_F$	0.002	0.94

E2	$PC1 \rightarrow T_F$	0.002	0.94
	$PC2 \rightarrow T_F$	0.001	0.80
E3	$PC1 \rightarrow T_F$	0.000	0.98
	$PC2 \rightarrow T_F$	0.000	0.94
E4	$PC1 \rightarrow T_F$	0.001	0.98
	$PC2 \rightarrow T_F$	0.000	0.96

Table 10. Hierarchical Bayesian estimation of mean effect sizes, including ATE^{Est}, ATT^{Est}, ATU^{Est}, and their differences, calculated using various regression models: linear regression (LR), support vector machine (SV), random forest (RF), gradient boosting (GB), and neural network (NN).

Variable	Mean	95 % Bayesian confidence interval
ATE^{Est}_{global}	1.832	[1.626, 2.037]
ATT^{Est}_{global}	0.966	[0.763, 1.172]
ATU^{Est}_{global}	0.525	[0.320, 0.731]
$\Delta ATE^{Est}_{global}$	1.193	[0.987, 1.397]
ATE^{Est}_{LR}	1.845	[1.844, 1.847]
ATE^{Est}_{SV}	1.765	[1.764, 1.766]
ATE^{Est}_{RF}	1.846	[1.845, 1.848]
ATE^{Est}_{GB}	1.846	[1.845, 1.848]
ATE^{Est}_{NN}	1.856	[1.854, 1.858]
ATT^{Est}_{LR}	1.186	[1.185, 1.187]
ATT^{Est}_{SV}	1.365	[1.364, 1.366]
ATT^{Est}_{RF}	0.854	[0.853, 0.856]
ATT^{Est}_{GB}	0.853	[0.851, 0.854]
ATT^{Est}_{NN}	0.553	[0.551, 0.555]
ATU^{Est}_{LR}	0.409	[0.408, 0.410]
ATU^{Est}_{SV}	0.142	[0.141, 0.143]
ATU^{Est}_{RF}	0.778	[0.777, 0.780]
ATU^{Est}_{GB}	0.764	[0.763, 0.766]
ATU^{Est}_{NN}	0.538	[0.536, 0.541]
ΔATE_{LR}	1.242	[1.241, 1.243]
ΔATE_{SV}	1.317	[1.316, 1.318]

ΔATE_{RF}	1.049	[1.048, 1.051]
ΔATE_{GB}	1.060	[1.058, 1.061]
ΔATE_{NN}	1.314	[1.312, 1.316]

Figure 1. Spatial distributions of (a) ferroelectric phase transition temperature (T_F), and (b) structural phase transition temperature (T_c) on $SrTiO_3$(110) under an external force of 231 MPa applied along [001], as reported in Ref. [60].

Figure 2. (a) K-shape clustering and (b) temperature dependence of the averaged retardance for $\lambda = 575$ nm, as reported in Ref. [60]. In (b), the curves for E1 and E2 overlap, as do the curves for E3 and E4.

Figure 3. Spatial distributions of the retardance δ at (a) 14.1 K and (b) 40.0 K for $\lambda = 575$ nm.

Figure 4. (a) Feature importance ranking of explanatory variables for T_F, and (b) spatial distribution of residuals after destandardization to the same scale as T_F, based on random forest analysis.

Figure 5. Spearman's rank correlation coefficients calculated for clusters (a) E1, (b) E2, (c) E3, and (d) E4. Each symbol represents the following: $a = T_F$, $b = T_c$, $c = \delta(130.9\,K)$, $d = \psi(130.9\,K)$, $e = \delta(90.0\,K)$, $f = \psi(90.0\,K)$, $g = \delta(40.0\,K)$, $h = \psi(40.0\,K)$, $i = \delta(14.1\,K)$, and $j = \psi(14.1\,K)$.

Figure 6. SEM diagrams illustrating the correlations between (a) T_F and four δ variables, (b) T_F and four ψ variables, (c) T_c and four δ variables, and (d) T_c and four ψ variables.

Figure 7. Spatial distributions of orthogonalized retardance components: (a) $PC1$ and (b) $PC2$ at $\lambda = 575$ nm, derived using Equations (3) and (4).

Figure 8. SEM diagrams showing the correlations between $PC1$ and $PC2$ with T_F and T_c for

clusters (a) E1, (b) E2, (c) E3, and (d) E4.

Figure 9. Directed Acyclic Graphs (DAGs) illustrating the causal relationships between $PC1$ and $PC2$ with T_F for clusters (a) E1, (b) E2, (c) E3, and (d) E4.

Figure 10. Evaluation of the effect of a random common cause. Histograms of ATE ratios are shown, representing ATE with a random variable vs. ATE without the random variable for (a) $PC1$ at E12, (b) $PC2$ at E12, (c) $PC1$ at E34, and (d) $PC2$ at E34. The insets show the corresponding histograms of p -values for each case.

Figure 11. Box-and-whisker plots of effect sizes for ATE^{Est}, ATT^{Est}, and ATU^{Est} using Two-model learner (T-Learner) with (a) linear regression (LR), (b) support vector machine (SV), (c) random forest (RF), (d) gradient boosting (GB), and (e) neural network (NN).

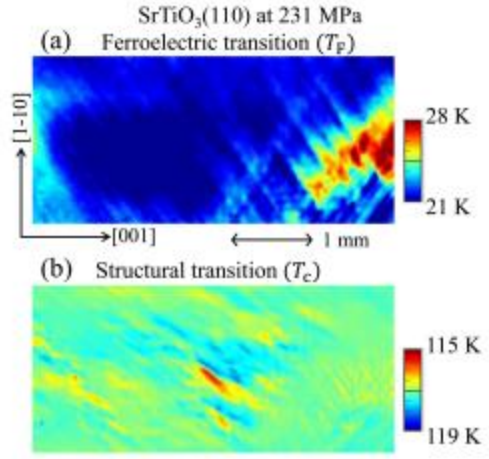


Figure 1. Spatial distributions of (a) ferroelectric phase transition temperature (T_F), and (b) structural phase transition temperature (T_c) on $SrTiO_3$(110) under an external force of 231 MPa applied along [001], as reported in Ref. [60].

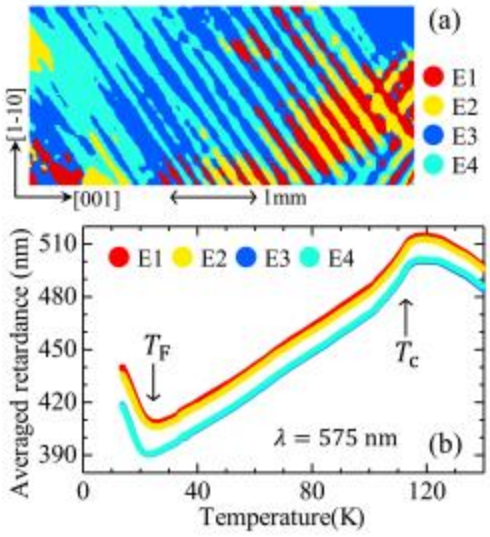


Figure 2. (a) K-shape clustering and (b) temperature dependence of the averaged retardance for $\lambda = 575$ nm, as reported in Ref. [60]. In (b), the curves for E1 and E2 overlap, as do the curves for E3 and E4.

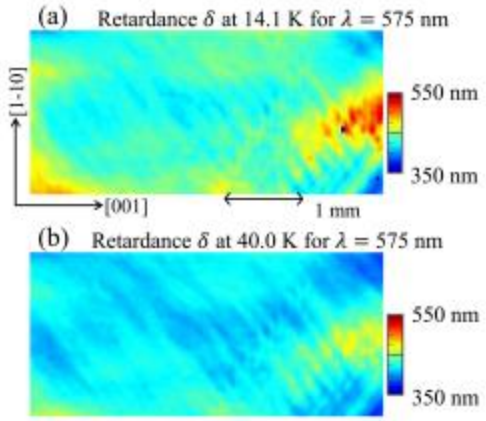


Figure 3. Spatial distributions of the retardance δ at (a) 14.1 K and (b) 40.0 K for $\lambda = 575$ nm.

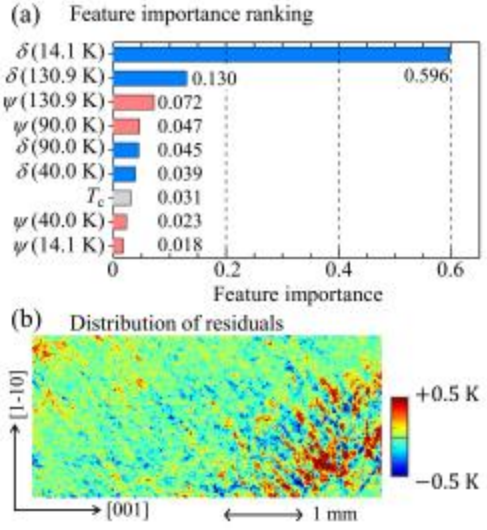


Figure 4. (a) Feature importance ranking of explanatory variables for T_{p}, and (b) spatial distribution of residuals after destandardization to the same scale as T_{p}, based on random forest analysis.

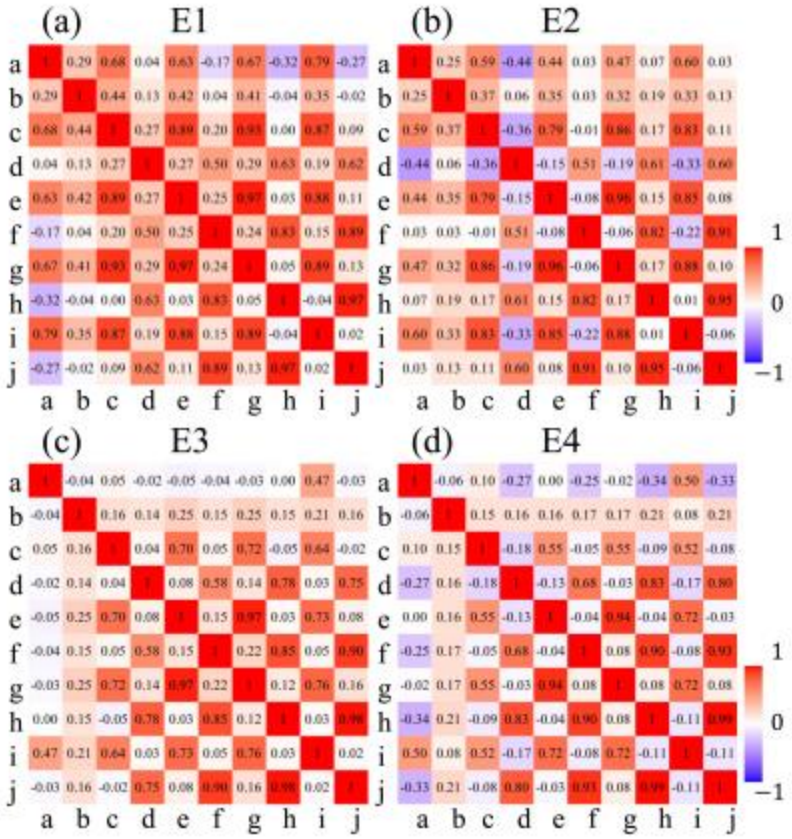


Figure 5. Spearman's rank correlation coefficients calculated for clusters (a) E1, (b) E2, (c) E3, and (d) E4. Each symbol represents the following: a = T_F, b = T_c, c = δ(130.9 K), d = ψ(130.9 K), e = δ(90.0 K), f = ψ(90.0 K), g = δ(40.0 K), h = ψ(40.0 K), i = δ(14.1 K), and j = ψ(14.1 K).

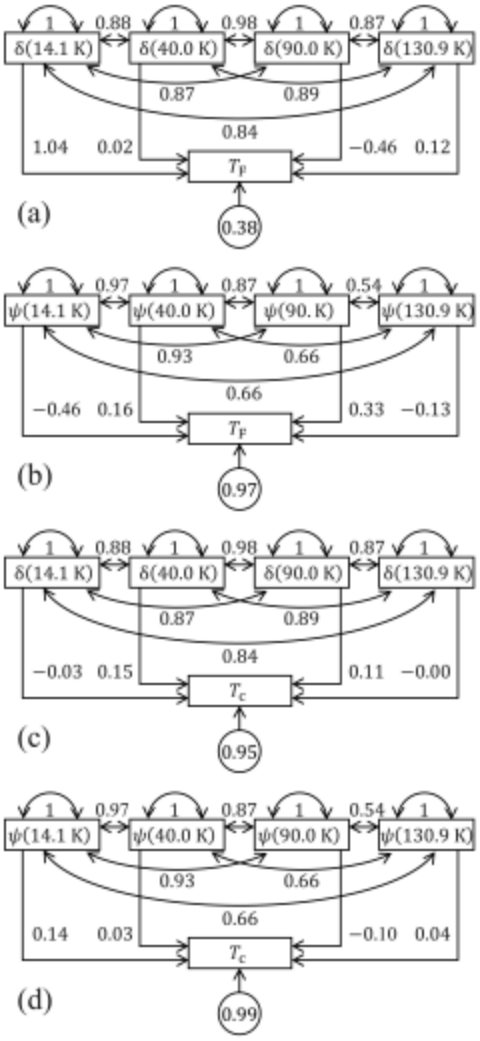


Figure 6. SEM diagrams illustrating the correlations between (a) T_F and four δ variables, (b) T_F and four ψ variables, (c) T_c and four δ variables, and (d) T_c and four ψ variables.

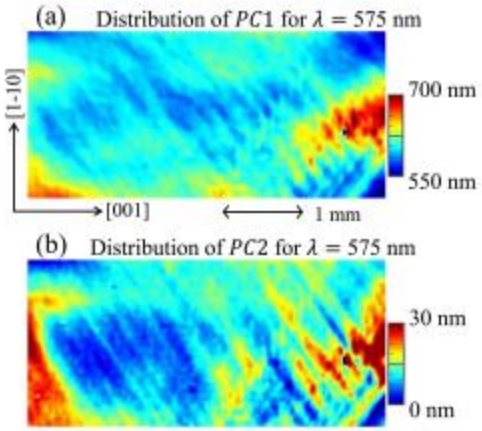


Figure 7. Spatial distributions of orthogonalized retardance components: (a) $PC1$ and (b) $PC2$ at $\lambda = 575$ nm, derived using Equations (3) and (4).

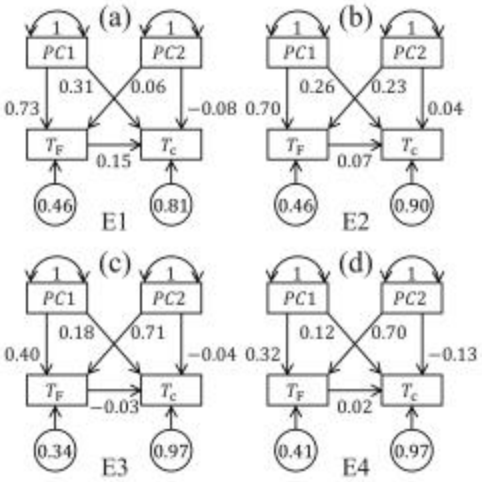


Figure 8. SEM diagrams showing the correlations between $PC1$ and $PC2$ with T_F and T_c for clusters (a) E1, (b) E2, (c) E3, and (d) E4.

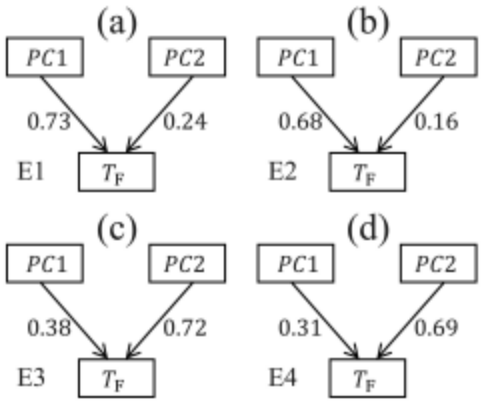


Figure 9. Directed Acyclic Graphs (DAGs) illustrating the causal relationships between $PC1$ and $PC2$ with T_F for clusters (a) E1, (b) E2, (c) E3, and (d) E4.

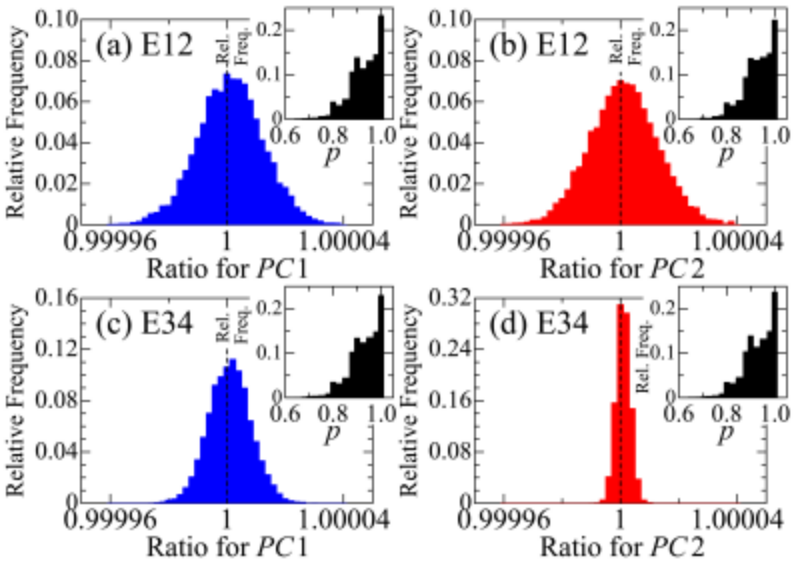


Figure 10. Evaluation of the effect of a random common cause. Histograms of ATE ratios are shown, representing ATE with a random variable vs. ATE without the random variable for (a) $PC1$ at E12, (b) $PC2$ at E12, (c) $PC1$ at E34, and (d) $PC2$ at E34. The insets show the corresponding histograms of p-values for each case.

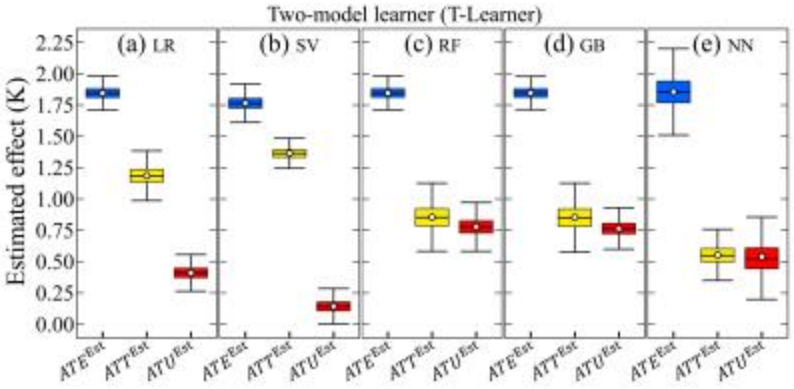
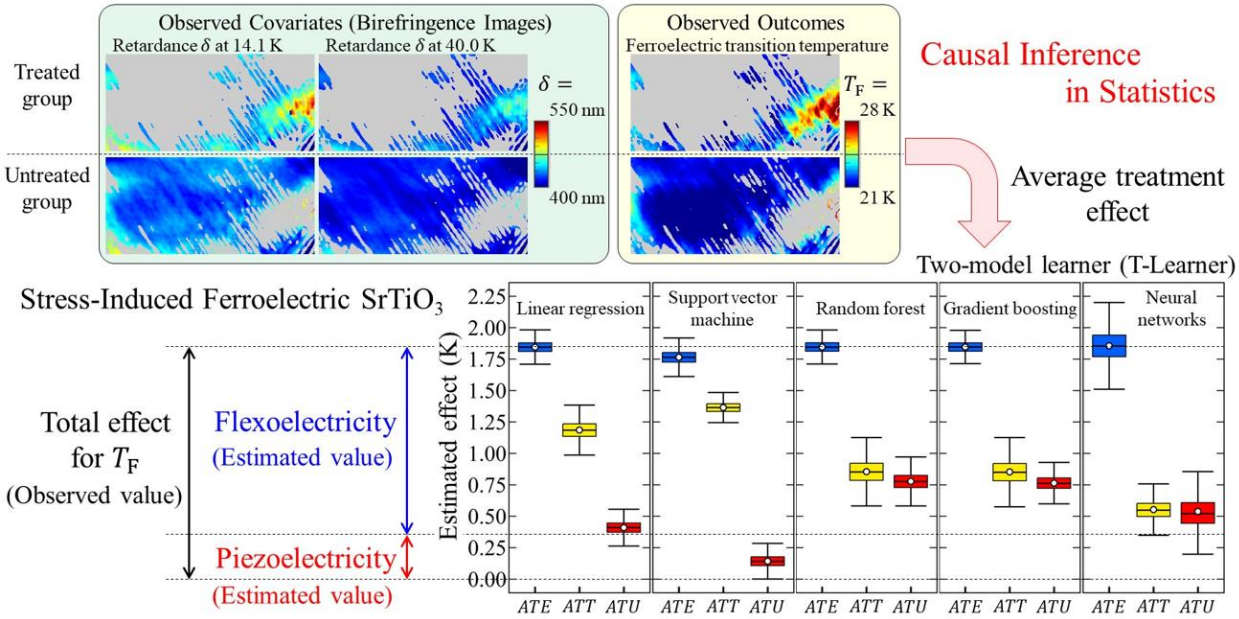


Figure 11. Box-and-whisker plots of effect sizes for ATE^{Est}, ATT^{Est}, and ATU^{Est} using Two-model learner (T-Learner) with (a) linear regression (LR), (b) support vector machine (SV), (c) random forest (RF), (d) gradient boosting (GB), and (e) neural network (NN).

Graphical Abstract



Impact Statement

This study integrates causal inference and SEM to quantitatively analyze the effect of changes in birefringence on the ferroelectric transition temperature of $SrTiO_3$, successfully separating contributions from piezoelectricity and flexoelectricity.