Machine-learning detection of the Berezinskii-Kosterlitz-Thouless transition and the second-order phase transition in XXZ models

Yusuke Miyajima and Masahito Mochizuki

Machine-Learning Detection of the Berezinskii-Kosterlitz-Thouless Transition and the Second-Order Phase Transition in the XXZ models

Yusuke Miyajima[1] and Masahito Mochizuki[1]

[1]*Department of Applied Physics, Waseda University, Okubo, Shinjuku-ku, Tokyo 169-8555, Japan*

We propose two machine-learning methods based on neural networks, which we respectively call the phase-classification method and the temperature-identification method, for detecting different types of phase transitions in the XXZ models without prior knowledge of their critical temperatures. The XXZ models have exchange couplings which are anisotropic in the spin space where the strength is represented by a parameter $\Delta(> 0)$. The models exhibit the second-order phase transition when $\Delta > 1$, whereas the Berezinskii-Kosterlitz-Thouless (BKT) phase transition when $\Delta < 1$. In the phase-classification method, the neural network is trained using spin or vortex configurations of well-known classical spin models other than the XXZ models, e.g., the Ising models and the XY models, to classify those of the XXZ models to corresponding phases. We demonstrate that the trained neural network successfully detects the phase transitions for both $\Delta > 1$ and $\Delta < 1$, and the evaluated critical temperatures coincide well with those evaluated by conventional numerical calculations. In the temperature-identification method, on the other hand, the neural network is trained so as to identify temperatures at which the input spin or vortex configurations are generated by the Monte Carlo thermalization. The critical temperatures are evaluated by analyzing the optimized weight matrix, which coincide with the result of numerical calculation for the second-order phase transition in the Ising-like XXZ model with $\Delta = 1.05$ but cannot be determined uniquely for the BKT transition in the XY-like XXZ model with $\Delta = 0.95$.

PACS numbers:

I. INTRODUCTION

Machine learning techniques are capable of solving classification and regression problems by extracting hidden features in data [1, 2]. Recently, the techniques have begun to be exploited for scientific research [3]. It is widely recognized that the architecture of machine learning that performs classification, prediction, and presumption by learning certain features from enormous data has compatibility with the procedure of scientific research particularly in physics, which approaches universal concepts of nature by extracting essences from accumulated data and knowledge.

Indeed, the machine learning techniques have recently been applied to the condensed-matter physics intensively [4]. One important subject in research of this kind is description of quantum many-body states using neural networks [5–7], which is based on the universal approximation theorem of neural networks and the variational ansatz. Recently, the ground-state phase diagram of a frustrated J_1-J_2 quantum Heisenberg model has been revealed by calculating the energies for optimized neural networks describing quantum spin states [7]. This phase diagram has not been obtained by any numerical methods so far. In this sense, this is an important example that a machine learning elucidated a new physics.

Another important subject of the machine-learning research in condensed-matter physics is detection of phases and phase transitions in physical models [8]. Recently, phase diagrams of several quantum models with randomness have been investigated using machine learning techniques by classifying spatial-map images of electron wavefunctions at various parameters to corresponding phases. The machine learning detections of phases and phase transitions are basically based on image recognition and pattern classification. Several kinds of classical spin models, in particular, the Ising models in two dimensions, have been investigated as research targets for benchmark tests [9–20]. The second-order phase transition accompanied by spontaneous symmetry breaking in the square-lattice Ising model exhibits apparent difference in the spatial spin configurations between the paramagnetic phase and the (anti)ferromagnetic phase because of a trivial change in the order parameter. Therefore, the machine learning technique can detect the phase transition in this model rather easily.

According to the Mermin-Wagner's theorem [21], spin models with continuous symmetry do not show any spontaneous symmetry breaking at finite temperatures in one and two dimensions. Indeed, the XY model in two dimensions has the U(1) symmetry and thus does not exhibit any phase transitions with symmetry breaking. However, this model is known to exhibit the BKT transition which is a topologically characterized phase transition between a high-temperature phase with individual vortices and antivortices and a low-temperature phase with vortex-antivortex bound pairs [22, 23]. It is known that the BKT transition is difficult to detect by numerical techniques based on the statistical mechanics such as the Monte Carlo techniques. This is because anomalies, i.e., jump, peak, and kink, in the thermodynamic quantities are not necessarily related with the BKT transition point directly. For example, a peak of the specific heat does not show up at the transition point. The magnetic susceptibility exhibits an anomaly at the transition point but is suffered from significant finite-size effects with a

logarithmic correction term, which makes a size-scaling extrapolation to the thermodynamic limit difficult.

Recently, several attempts have been made to detect the BKT transition in the XY model and the q-state clock model with machine learning techniques [24–35]. However, because the BKT transition is a phase transition without any spontaneous symmetry breakings, the machine learning techniques based basically on the pattern recognition are not necessarily powerful for its detection. Most of the attempts are based on supervised learning methods, and they require feature engineering in advance to make preprocessed input data, that is, the data of quantity which captures the features of the phases and the phase transitions such as vortex configurations, histogram of the spin orientations, and the spin correlation functions. Moreover, they need prior knowledge of fundamental properties of the model, e.g., number of phases, approximate transition temperatures, and order parameters. These are critical problems because the first one means that we must know features that characterize the phases and the phase transitions hosted by the models in advance, and the second one means that the method cannot be applied to unknown models.

Here it should also be mentioned that there have been a lot of experimental efforts to search for the BKT behavior in real magnetic materials so far [36–40], but they are not necessarily successful. A main obstacle for the observation of the BKT behavior in real materials is the small but almost unavoidable interlayer coupling combined with the smallness of the easy-plane anisotropy. From the theoretical point of view, the lack of powerful theoretical framework to discuss possible BKT behavior in complicated spin models might be another obstacle. The BKT phases and the BKT transitions in theoretical spin models have been studied usually by using the Monte Carlo techniques. In the Monte Carlo analysis, the quantity called helicity modulus is usually employed to detect the BKT transition. The expression of this quantity does not have a general form but has a form specific to each spin model, which must be derived by hand for each case. Its derivation becomes difficult for complicated theoretical spin models that describe real magnets with further exchange coupling, magnetic anisotropies, and higher-order interactions. Therefore, a powerful and versatile machine-learning method applicable to general spin models is highly demanded for further development of the research on the BKT behavior in real magnets.

In this paper, we propose two machine-learning methods based on neural networks with simple architectures as versatile tools to detect the BKT phase and the BKT transition in classical spin models. The two methods are referred to as the phase-classification method and the temperature-identification method, respectively. We employ the XXZ models for examination of these methods and try to detect two types of phase transitions, i.e., the second-order transition and the BKT transition, in the XXZ models without prior knowledge of their critical temperatures. The XXZ models have exchange cou-

plings which are anisotropic in the spin space with an anisotropy parameter $\Delta(> 0)$ and exhibits the second-order transition when $\Delta > 1$ (Ising-like case) and the BKT transition when $\Delta < 1$ (XY-like case). In the phase-classification method, the spin or vortex configurations of the XXZ models generated by the Monte Carlo thermalization are classified into corresponding phases by utilizing a neural network trained with those of well-known classical spin models, e.g., the Ising model and the XY model as inspired by previous studies in Refs. [41–43]. We demonstrate that the trained neural network successfully performs the classification and determines the phase-transition point for both $\Delta > 1$ and $\Delta < 1$. The evaluated critical temperatures coincide well with those evaluated by numerical calculations. In the temperature-identification method, on the other hand, the neural network is trained so as to output correct temperatures at which the input spin or vortex configurations are generated by the Monte Carlo thermalization, and the critical temperatures are evaluated by analyzing the optimized weight matrix after the training as inspired by studies in Refs. [9, 11]. We show that the critical temperature evaluated by this analysis coincides with the result of Monte Carlo calculation for the second-order phase transition for $\Delta = 1.05$, whereas the critical temperature cannot be evaluated uniquely for the BKT transition for $\Delta = 0.95$.

Both of our proposed methods have advantages over the Monte Carlo methods that have been traditionally used for studying the phase-transition phenomena in spin models in terms of computational cost and generality. Moreover, our methods also have advantages over the previously proposed machine learning methods, that is, our methods require no or less prior knowledge of the models and no or least featrure engineering with preprocessing of input data in contrast to the previous methods. These advantages will be discussed in the last section in detail. We expect that our work will pave the way to exploring novel phases and phase-transition phenomena in theoretical spin models and will contribute to the further development of this research field.

II. MODEL

We study ferromagnetic $(J > 0)$ XXZ models on square lattices. The Hamiltonian is given by,

$$\mathcal{H} = -J \sum_{\langle i,j \rangle} \left(S_i^x S_j^x + S_i^y S_j^y + \Delta S_i^z S_j^z \right), \qquad (1)$$

where $\boldsymbol{S}_i = (S_i^x, S_i^y, S_i^z)$ is a continuous classical spin vector on the ith site, norm of which is set to be unity $(|\boldsymbol{S}_i| = 1)$. Here the parameter $\Delta(> 0)$ describes the anisotropy of the exchange coupling in the spin space, and the summation is taken over the nearest-neighbor site pairs $\langle i, j \rangle$. The XXZ models have been studied intensively because many materials and systems such as ultrathin ferromagnetic films and superconducting films
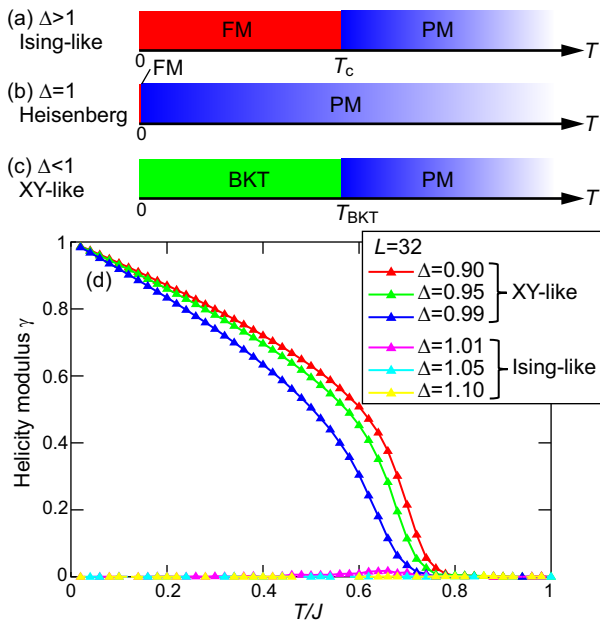
FIG. 1: Schematic temperature phase diagrams of the XXZ models on a square lattice for (a) $\Delta > 1$ (Ising-like case), (b) $\Delta = 1$ (Heisenberg case), and (c) $\Delta < 1$ (XY-like case). Here FM and PM denote the ferromagnetic and paramagnetic phases, respectively. (d) Temperature profiles of the helicity modulus calculated using the Monte Carlo technique for various values of Δ. This quantity becomes finite in the XY-like case with $\Delta < 1$ below a certain temperature, whereas it vanishes in the Ising-like case with $\Delta > 1$. The calculations are performed for a system size of L^2 with $L = 32$.

have turned out to be well described by these models [44–51].

Schematic temperature phase diagrams of the XXZ models are shown in Figs. 1(a)-(c). We find that the models exhibit distinct behaviors depending on the anisotropy Δ. When $\Delta > 1$, the model exhibits a single second-order phase transition as in the Ising model, whereas the model exhibits a single BKT transition as in the XY model when $\Delta < 1$. At $\Delta = 1$, the model is equivalent to the classical Heisenberg model in two dimensions and thus does not exhibit any phase transition at finite temperatures.

To see the Δ-dependent behaviors of the XXZ models, we calculate a quantity called helicity modulus [52–54] by measuring variations of the free energy when uniform and infinitesimal twists are introduced in the spin alignment in a certain direction. Figure 1(d) shows temperature profiles of the helicity modulus γ for various values of the anisotropy parameter Δ calculated by the Monte Carlo calculations. The calculations are performed for a square lattice of 32×32 sites, and the twist along the x-axis is considered. We find that γ is suppressed to be zero at any temperatures for the Ising-like case with $\Delta > 1$, whereas it increases with decreasing temperature below a certain temperature for the XY-like case with $\Delta < 1$ which manifests the emergence of BKT phase.

III. METHOD

We examine two machine-learning methods based on neural networks for detecting the phase transitions in the XXZ models, which are referred to as the phase-classification method and the temperature-identification method, respectively. The neural networks used in these methods are implemented with the machine-learning library KERAS [55]. We perform supervised learning for these neural networks using a set of training input and corresponding answer labels.

A. Phase-classification method

We first implement a neural network shown in Fig. 2(a). The binary classification of two phases is considered here. Inputs to this neural network are spin or vortex configurations of well-known classical spin models other than the XXZ models, i.e., the Ising models and the XY models in the present study, prepared by the Monte Carlo thermalization. The outputs are vectors with two components, each of which represents a probability that the input belongs to the phase. This neural network is referred to as the phase-classification type hereafter.

For detection of the second-order phase transition in the Ising-like XXZ model with $\Delta = 1.05$, we simply use this neural network, which comprises one input layer, one or two hidden layers, and one output layer. The input layer has $3L^2$ nodes corresponding to all the spin components on the $L \times L$ lattice, whereas the output layer has two nodes corresponding to the ferromagnetic phase and the paramagnetic phase. The two outputs are probabilities $P_{\mathrm{FM}}(T)$ and $P_{\mathrm{PM}}(T)$ that the input belongs to the ferromagnetic phase and the paramagnetic phase, respectively. The number of the hidden nodes in each hidden layer is determined so as to encode the $3L^2$-component inputs into the two-component outputs with the same compression ratio at each layer. Adjacent layers are connected by weight matrices which weight the importance of data transmitted from the former layer to the next layer. The Rectified Linear Unit (ReLU) function is employed as an activation function for the hidden layers, whereas the softmax function for the output layer. The sum of probabilities $P_{\mathrm{FM}}(T) + P_{\mathrm{PM}}(T)$ equals to unity because the softmax function normalizes the components of output.

For detection of the BKT transition in the XY-like XXZ model with $\Delta = 0.95$, the input data are processed by using the convolutional neural network [2] before feeding them into the phase-classification-type neural network. The convolutional neural network is adopted to capture two-dimensional features of spatial configurations of vortices and antivortices to judge whether they are bound or unbound. The convolutional neural network comprises one input layer, three hidden layers, and one output layer as shown in Fig. 3, where the first hidden layer is the convolution layer. The ReLU function is em-

(a) Phase-classification-type neural network
for XXZ model with Δ>1 (Ising-like case)

Training data: Spin configs. of Ising model above/below T_c
Input data: Spin configs. of XXZ model at various T

input layer
($3L^2$ nodes)   hidden layer x 2 or 3   output layer
(2 nodes)

(b) Temperature-identification-type neural network
for XXZ model with Δ>1 (Ising-like case)

Training data: Spin configs. of XXZ model at various T

input layer
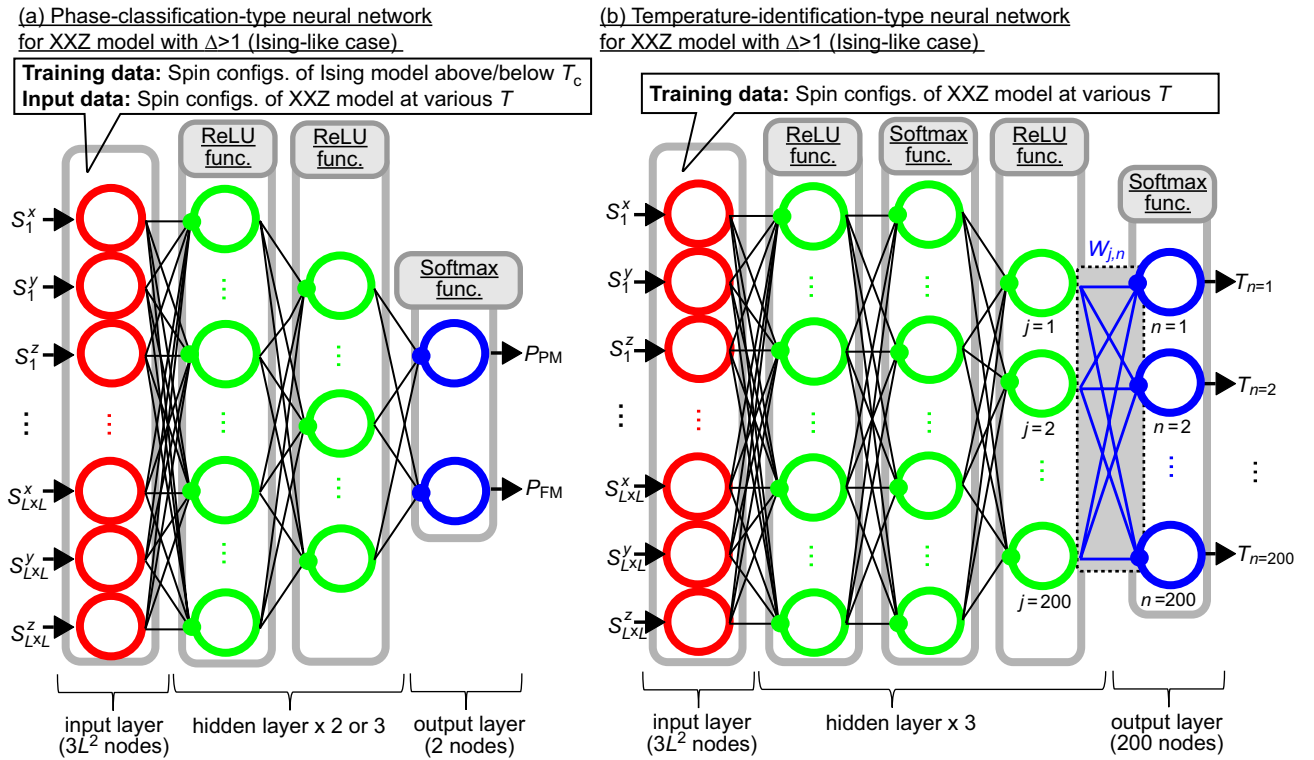($3L^2$ nodes)   hidden layer x 3   output layer
(200 nodes)

FIG. 2: Neural networks used for detection of the second-order phase transition in the Ising-like XXZ model with $\Delta > 1$. (a) Phase-classification type neural network. Spin configurations of $\boldsymbol{S}_i = (0, 0, S_{iz})$ with $S_{iz} = +1$ or -1 for the Ising model are used for training data, while those of $\boldsymbol{S}_i = (S_{ix}, S_{iy}, S_{iz})$ for the XXZ model are used for the input data. This neural network output probabilities P_{PM} and P_{FM} that the input spin configuration is paramagnetic and ferromagnetic, respectively, where their sum $P_{\mathrm{PM}} + P_{\mathrm{FM}}$ is normalized to be unity. The weight matrices connecting the layers in this neural network are trained so as to output a higher probability for the phase to which the spin configuration indeed belongs. (b) Temperature-identification-type neural network. Spin configurations of $\boldsymbol{S}_i = (S_{ix}, S_{iy}, S_{iz})$ for the XXZ model are used for training data. The weight matrices in this neural network are trained so as to output a correct temperature at which the input spin configuration is obtained. All the spin configurations used as training and input data are prepared by the Monte Carlo thermalization. The activation functions adopted in respective hidden and output layers are also presented.

ployed for the three hidden layers, whereas the softmax function for the output layer. In the convolution layer, the input data with L^2 components are transformed into a data with $(L+3)^2$ components by adding zeros as additional components in the so-called zero-padding procedure. The convolutional procedure is performed by applying a filter with 4^2 weight components to the $(L+3)^2$ data to obtain L^2 data. Then the ReLU function is applied to each component of the processed L^2 data for activation, and the two-dimensional data are flattened to one-dimensional L^2 data.

The convoluted L^2 data are fed to the phase-classification-type neural network for detection of the BKT transition. The two nodes in the output layer correspond to the BKT phase and the paramagnetic phase, which output the probabilities $P_{\mathrm{BKT}}(T)$ and $P_{\mathrm{PM}}(T)$ that the input vortex configuration belongs to the BKT phase and the paramagnetic phase, respectively. The sum of probabilities $P_{\mathrm{BKT}}(T) + P_{\mathrm{PM}}(T)$ again equals to unity. We determine the number of nodes in the hidden layers in the same manner as the Ising-like case.

We prepare sets of the spin (vortex) configurations of the Ising (XY) model and corresponding phase labels as training data. The neural network is trained so as to correctly guess phases to which the input spin (vortex) configurations belong. We use a cost function to quantify the difference between the outputs of the neural network and the answer labels. As a cost function, the cross-entropy error function is employed, which is given by,

$$E(\boldsymbol{t}^i, \boldsymbol{y}^i) = \sum_i \sum_k t_k^i \log y_k^i, \qquad (2)$$

where i is the label of input data, and k is the index of output nodes. y_k^i represents the kth component of the output vector from the neural network, and t_k^i represents the kth component of the answer label for the ith input training data. In the training procedure, we minimize the loss function by tuning the components of weight matrices using the optimization algorithm called Adam [56].

After the training procedure, we feed 100 spin or vortex configurations of the XXZ model generated at each of 200 temperature points to the neural network as the input
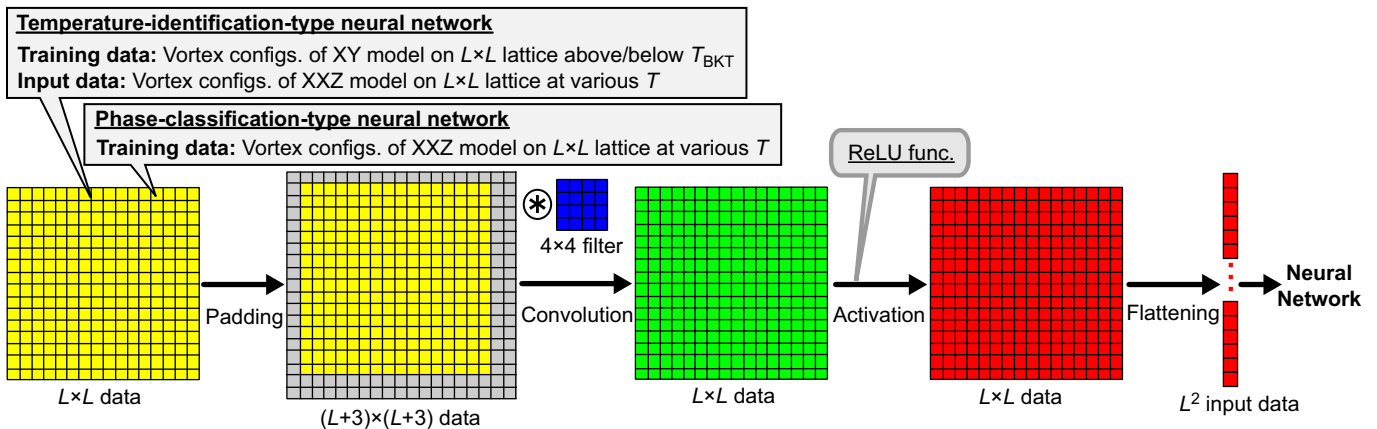
FIG. 3: Data convolution process in the convolutional neural network for detection of the BKT transition in the XY-like XXZ model with $\Delta < 1$ on the $L \times L$ lattice. Spatial vortex-configuration data with L^2 components are made from the spin-configuration data with $3L^2$ components generated by the Monte Carlo thermalization. The whole process of the convolution layer contains four procedures, i.e., padding, convolution, activation, and flattening. The L^2 data are transformed to $(L+3)^2$ data by adding zeros as additional components in the padding procedure. A 4×4 filter is applied to the $(L+3)^2$ data to obtain L^2 convoluted data in the convolution procedure, while the ReLU function is adopted for the activation procedure. The processed L^2 data are fed to a neural network, e.g., the phase-classification-type neural network or the temperature-identification-type neural network, which are basically the same as those in Fig. 4(a) and (b), respectively, although the number of input nodes is different. For the phase-classification-type neural network, the vortex configurations for the XY model are used as training data, while those of the XXZ model is used as input data. This neural network output probabilities P_{PM} and P_{BKT} that the input vortex configuration belongs to the paramagnetic phase and the BKT phase, respectively, where their sum $P_{\mathrm{PM}} + P_{\mathrm{BKT}}$ is normalized to be unity. The weight matrix in this neural network as well as 4^2 components of the filter are trained so as to assign the inputted vortex configurations to correct phases. For the temperature-identification-type neural network, the vortex configurations of the XXZ model are used to train the weight matrix and the filter so as to output correct temperatures at which the input vortex configurations are generated by the Monte Carlo thermalization.

data. For the lth input data ($l = 1, 2, \cdots, 100$) generated at temperature T, the probabilities $P_1^l(T)$ and $P_2^l(T)$ are output from the first and the second nodes in the output layer, respectively. From the 100 outputs of $P_\alpha^l(T)$ ($\alpha = 1, 2$), we calculate averaged probabilities $\bar{P}_\alpha(T)$ at each temperature point as,

$$\bar{P}_\alpha(T) = \frac{1}{N_l} \sum_{l=1}^{N_l} P_\alpha^l(T) \qquad (N_l = 100). \tag{3}$$

At the critical temperature, the neural network is confused in the phase classification because pronounced thermal-fluctuation effect obscures the difference between two phases. Thus, we determine the critical temperature by the crossing point of $\bar{P}_1(T)$ and $\bar{P}_2(T)$, at which both $\bar{P}_1(T)$ and $\bar{P}_2(T)$ take 0.5.

B. Temperature-identification method

We next implement a neural network shown in Fig. 2(b). The multiclass classification of temperatures is carried out by this neural network. Inputs to this neural network are spin or vortex configurations of the XXZ model prepared by the Monte Carlo thermalization at 200 temperature points ranged from $T_{n=1} = 0.01J$ to

$T_{n=200} = 2.00J$ with constant intervals of $\Delta T = 0.01J$. The outputs are vectors with 200 components, where the nth component represents a probability that the input spin or vortex configuration is generated at $T_n = n\Delta T$. This neural network is referred to as the temperature-identification type hereafter.

For detection of the second-order phase transition in the Ising-like XXZ model with $\Delta = 1.05$, we simply use this neural network, which comprises one input layer, three hidden layers, and one output layer. The input layer has $3L^2$ nodes corresponding to all the spin components on the $L \times L$ lattice, whereas the output layer has 200 nodes corresponding to the 200 temperature points. The number of the hidden nodes in each hidden layer is determined so as to encode the $3L^2$-component inputs into 200-component outputs with the same compression ratio at each layer. Adjacent layers are connected by weight matrices. The ReLU function is employed as an activation function for the first and third hidden layers, while the softmax function is employed for the second hidden layer and the output layer.

On the other hand, we use the temperature-identification-type convolutional neural network [2] for detecting the BKT transition in the XY-like XXZ model with $\Delta = 0.95$. The convolutional neural network comprises one input layer, three hidden layers, and one output layer, where the first hidden layer is the convolu-

tion layer. As activation functions, the ReLU function is employed for the first and the third hidden layer, the softmax function is employed for the second hidden layer and the output layer. The same data processing as in the convolution layer of phase-classification-type convolutional neural network is performed in the convolution layer. In the convolutional neural network, input data is processed as follows: First, two-dimensional input data is transformed into one-dimensional data in the convolution layer. The convolution layer connects the second hidden layer in Fig. 2(b). The rest processes are the same as the neural network in Fig. 2(b). Here, the output layer has 200 nodes corresponding to the 200 temperature points, while the number of the hidden nodes is determined as the same as $\Delta = 1.05$.

We prepare sets of spin configurations of the Ising-like XXZ model with $\Delta = 1.05$ or vortex configurations of the XY-like XXZ model with $\Delta = 0.95$ and corresponding temperature labels as training data. Therefore, the neural networks are trained to correctly estimate the temperature at which the input spin or vortex configurations are generated. We use the cross-entropy error function as a cost function. In the training process, using the optimization algorithm called Adam [56], we minimize the loss function by tuning components of weight matrices.

After the training procedure, we focus on the weight matrix $W_{j,n}$ connecting the jth node in the last hidden layer with $j = 1, 2, \cdots 200$ and the nth node in the output layer corresponding to the nth temperature point $T_n = n\Delta T$ with $n = 1, 2, \cdots 200$. A gray-scale plot of the matrix components $W_{j,n}$ in plane of j and T_n is called heat map. Heat maps tend to exhibit distinct patterns for different phases, which can be exploited to detect phase transitions.

We evaluate critical temperatures by quantitative analysis of the pattern changes in heat maps. For this purpose, we introduce special kinds of correlation function $C_W(T)$ and variance $V_W(T)$. The correlation function $C_W(T)$ is defined by,

$$C_W(T_n) = \frac{1}{N_\mathrm{h}} \sum_{m=1}^{n-1} \sum_{j=1}^{N_\mathrm{h}} W_{j,m} W_{j,m+1}, \qquad (4)$$

where $N_h (= 200)$ is the number of nodes in the last hidden layer. The product $W_{j,m} W_{j,m+1}$ quantifies similarity or difference between adjacent components. On the other hand, the variance $V_W(T)$ is defined by,

$$V_W(T_n) = \frac{1}{n-1} \sum_{j=1}^{N_h} \sum_{m=1}^{n-1} (W_{j,m} - \bar{W}_j)^2,$$

$$\bar{W}_j = \frac{1}{n-1} \sum_{m=1}^{n-1} W_{j,m}. \qquad (5)$$

For $T_n < T_\mathrm{c}$, heat maps tend to exhibit a stripe pattern reflecting an ordered phase. Consequently, the components $W_{j,n}$ take similar values within each row specified by j and thus the variance $V_W(T_n)$ is small. On the
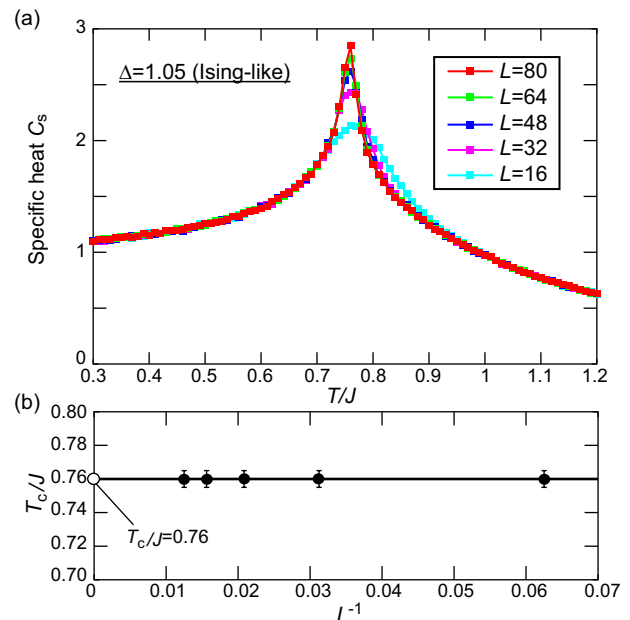

FIG. 4: (a) Temperature profiles of specific heat for the Ising-like XXZ model with $\Delta = 1.05$ for various system sizes, each of which exhibits a peak at the critical temperature T_c of the second-order phase transition from the paramagnetic to ferromagnetic phases. (b) Finite-size scaling analysis of T_c. In the thermodynamic limit of $L \to \infty$, T_c is evaluated to be $T_\mathrm{c}/J = 0.76$ by extrapolation.

TABLE I: Critical temperatures of the Ising-like XXZ model with $\Delta = 1.05$ for various system sizes in units of J. The case of $L \to \infty$ denote results in the thermodynamic limit.

L	16	32	48	64	80	∞
$T_\mathrm{c}^\mathrm{MC}/J$	0.760(5)	0.760(5)	0.760(5)	0.760(5)	0.760(5)	0.760(0)
$T_\mathrm{c}^\mathrm{PC}/J$	0.724(6)	0.735(3)	0.749(4)	0.755(3)	0.763(6)	0.766(5)

contrary, for $T_n > T_\mathrm{c}$, heat maps tend to exhibit a sandstorm pattern reflecting a disordered phase, and the components $W_{j,n}$ take scattered values along each row. As a result, the variance $V_W(T_n)$ increases from T_c with increasing T_n. Pattern changes in the heat map manifests themselves in the T_n-profiles of $C_W(T_n)$ and $V_W(T_n)$ in distinct manners.

IV. RESULTS FOR $\Delta = 1.05$ (ISING-LIKE CASE)

A. Monte Carlo calculation

We first determine the critical temperature T_c of the second-order phase transition in the Ising-like XXZ model with $\Delta = 1.05$ from specific heat calculated by the Monte Carlo calculation. Figure 4(a) shows the calculated temperature profiles of specific heat for several lattice sizes. For each lattice size, the specific heat has

a peak at T_c. A finite-size scaling analysis is performed to evaluate T_c in the thermodynamic limit. We derive the size dependence of T_c. The spin correlation length ξ behaves as $\xi \propto |T - T_c|$ near the critical temperature because the second-order phase transition belongs to the Ising universality class. Because the correlation length ξ satisfies $\xi = L$ at the critical temperature $T_c(L)$ for finite-size systems, the size dependence of the critical temperature is given by,

$$T_c(L) = T_c(\infty) + \frac{a}{L}. \tag{6}$$

Here $T_c(\infty)$ is the critical temperature in the thermodynamic limit, and a is a constant. Figure 4(b) presents the size dependence of the critical temperature $T_c(L)/J$. From the extrapolation to $L^{-1} \to 0$, the critical temperature in the thermodynamic limit is evaluated as $T_c^{MC}/J = 0.76$. The critical temperatures for several system sizes are summarized in Table I.

B. Spin configurations used as training data

We prepare many spin configurations of the Ising model and the Ising-like XXZ model with $\Delta = 1.05$ for several lattice sizes $L \times L$ ($L = 16, 32, 48, 64, 80$) as training data of the neural network. The spin configurations are generated by the Monte Carlo thermalization with 12000 iterative steps based on the single-flip Metropolis algorithm. For the Ising model, 100 spin configurations are prepared at each of 400 temperature points from $T = 0.01J$ to $T = 4.00J$ at intervals of $\Delta T = 0.01J$. For the Ising-like XXZ model, on the other hand, 200 spin configurations are prepared at each of 200 temperature points from $T = 0.01J$ to $T = 2.00J$ at intervals of $\Delta T = 0.01J$.

The temperature phase diagram and the examples of spin configurations for the Ising model are presented in Figs. 5(a)-(c). The blue (red) arrows in Figs. 5(b) and (c) denote $+1$ (-1) spins. Most of the spins are pointing in the same direction in the ferromagnetic phase [Fig. 5(b)], whereas $+1$ and -1 spins are randomly distributed in the paramagnetic phase [Fig. 5(c)]. On the other hand, a schematic temperature phase diagram and the examples of spin configurations for the Ising-like XXZ model are presented in Figs. 5(d)-(f). The arrows in Figs. 5(e) and (f) denote the in-plane components of the spins, while the colors of pixels represent their out-of-plane components. The image for the ferromagnetic phase [Fig. 5(e)] is almost unicolored by red or blue because most of the ordered spins are pointing along the z axis, while the red and blue pixels are randomly distributed in the image for the paramagnetic phase [Fig. 5(f)]. The existence of in-plane spin components might make the situation complicated for detection of the phase transition in the XXZ model.
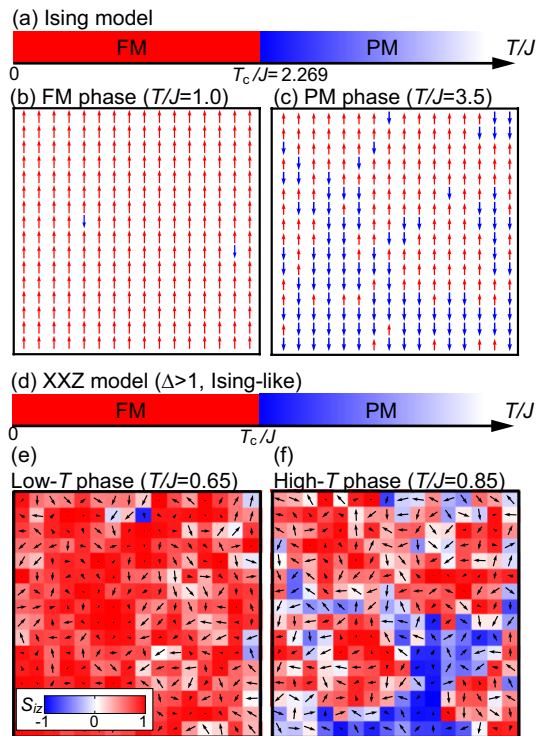

FIG. 5: Spin configurations of the Ising model on square lattices are used for data to train the neural network aiming for detection of the second-order phase transition in the XXZ model with $\Delta > 1$. (a) Schematic temperature phase diagram of the square-lattice Ising model, which exhibits the second-order phase transition from the paramagnetic (PM) phase to the ferromagnetic (FM) phase at $T_c/J = 2.269$. Spin configurations generated by the Monte Carlo thermalization in the temperature range $0.01 \le T/J \le 2.01$ ($2.52 \le T/J \le 4.00$) are used for the training data for the FM (PM) phase. (b), (c) Typical spin configurations of (b) FM phase ($T/J = 1.0$, $L = 16$) and (c) PM phase ($T/J = 3.5$, $L = 16$). Spin configurations of the XXZ model with $\Delta = 1.05$ at various temperatures are entered to the trained neural network as input data. (d) Schematic temperature phase diagram of the Ising-like XXZ model with $\Delta > 1$. (e), (f) Typical spin configurations at (e) lower temperatures ($T/J = 0.65$) and (f) higher temperatures ($T/J = 0.85$) of the XXZ model with $\Delta = 1.05$. It turns out that the neural network trained by the binary-spin configurations of the Ising model is capable of detecting the phase transition of the three-dimensional vector-spin states in the XXZ model.

C. Phase-classification method

We attempt to detect the second-order phase transition in the Ising-like XXZ model with $\Delta = 1.05$ by the phase-classification-type neural network shown in Fig. 2(a). As the training dataset, 100 ferromagnetic spin configurations of the Ising model (instead of the XXZ model) are prepared at each temperature point ranged from $T = 0.01$ to $T = 2.01$, while 100 paramagnetic spin configurations are prepared within a temperature range from $T = 2.52$ to $T = 4.00$ at intervals of $\Delta T = 0.01$. Here,
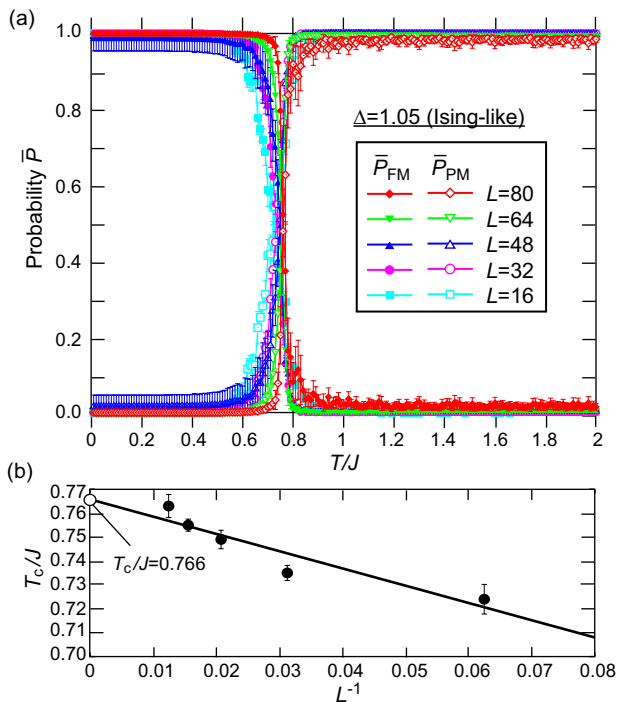
FIG. 6: (a) Temperature profiles of the averaged probabilities $\bar{P}_{\text{PM}}$ and $\bar{P}_{\text{FM}}$ of the Ising-like XXZ model with $\Delta = 1.05$ obtained by the phase-classification-type neural network for various system sizes. Here, $\bar{P}_{\text{PM}}(T)$ and $\bar{P}_{\text{FM}}(T)$ are averages of the probabilities that a set of the spin configurations generated at temperature T by the Monte Carlo thermalization are classified to the paramagnetic phase and the ferromagnetic phase, respectively. The summation $\bar{P}_{\text{PM}} + \bar{P}_{\text{FM}}$ is always normalized to be unity. Each crossing point of $\bar{P}_{\text{PM}}$ and $\bar{P}_{\text{FM}}$, at which both probabilities take 0.5, is assigned to a transition temperature for the corresponding system size. (b) Finite-size scaling analysis of T_c. T_c in the thermodynamic limit of $L \to \infty$ is evaluated to be $T_c/J = 0.766$ by extrapolation, which coincides well with $T_c/J = 0.76$ evaluated by the conventional Monte Carlo analysis.

the spin configurations around $T_c = 2.269$ are excluded from the training data. Each spin configuration of the Ising model on the $L \times L$ square lattice has only L^2 components. We, therefore, transform the binary spin variables $S_i (= \pm 1)$ into three-component vectors $(0, 0, S_i)$. This transformation is justified by the uniaxial magnetic anisotropy in the XXZ models.

The vectors $(0, 0, S_1, \cdots, 0, 0, S_i, \cdots, 0, 0, S_{L^2})$, which is arrays of the vectors $(0, 0, S_i)$ at all sites, is fed to the neural network. On the other hand, the answer labels for the training data are vectors with two components, which represent probabilities that the input spin configuration belongs to the paramagnetic phase and the ferromagnetic phase, respectively. The vector $(1, 0)$ is given as the answer labels for the input spin configurations generated below T_c, while the vector $(0, 1)$ is given for those generated above T_c. The former (latter) vector mean that the input is absolutely for the ferromagnetic (paramagnetic)

phase. The neural network is trained so as to correctly guess phases to which the input spin configurations belong.

After completing the training procedure, we input the spin configurations of the XXZ model to the trained neural network, where each input is an array of the spin vectors, i.e., $(S_1^x, S_1^y, S_1^z, \cdots, S_i^x, S_i^y, S_i^z, \cdots, S_{L^2}^x, S_{L^2}^y, S_{L^2}^z)$. As the iput dataset, 100 spin configurations are prepared at each temperature point within a range from $T = 0.01J$ to $T = 2.00J$ at intervals of $\Delta T = 0.01J$. For the 100 spin configurations generated at T, 100 output vectors $\left(P_{\text{FM}}^l(T), P_{\text{PM}}^l(T) \right)$ are obtained, where l $(= 1, 2, \cdots, 100)$ is the index of the input data. We calculate the averages of $P_{\text{FM}}^l(T)$ and $P_{\text{PM}}^l(T)$ over 100 input data to obtain the averaged probabilities $\bar{P}_{\text{FM}}(T)$ and $\bar{P}_{\text{PM}}(T)$.

Figure 5(a) shows the temperature profiles of $\bar{P}_{\text{FM}}(T)$ and $\bar{P}_{\text{PM}}(T)$. We find that $\bar{P}_{\text{FM}}(T) \approx 1$ and $\bar{P}_{\text{PM}}(T) \approx 0$ at lower temperatures, whereas $\bar{P}_{\text{FM}}(T) \approx 0$ and $\bar{P}_{\text{PM}}(T) \approx 1$ at higher temperatures. This indicates that the neural network correctly guesses phases to which the input data belong. The critical temperature T_c for each lattice size is determined from a crossing point of the two profiles. Moreover, T_c in the thermodynamic limit is evaluated to be $T_c^{\text{PC}}/J = 0.766$ by the finite-size scaling analysis using Eq. (6). This value coincides well with the value $T_c^{\text{MC}}/J = 0.76$ obtained by the Monte Carlo calculation. The values of T_c for several lattice sizes evaluated by the present phase-classification method are summarized in Table I. Now it has been successfully demonstrated that the phase-classification-type neural network can detect the phases and the second-order phase transition in the Ising-like XXZ model.

D. Temperature-identification method

Next we attempt to detect the second-order phase transition in the Ising-like XXZ model with $\Delta = 1.05$ by the temperature-identification-type neural network shown in Fig. 2(b). As the training dataset, 100 spin configurations of the XXZ model are prepared at each temperature point $T_n = n\Delta T$ ($n = 1, 2, \cdots, 200$) ranged from $T = 0.01J$ to $T = 2.00J$ at intervals of $\Delta T = 0.01J$. The vectors $(S_1^x, S_1^y, S_1^z, \cdots S_i^x, S_i^y, S_i^z, \cdots, S_{L^2}^x, S_{L^2}^y, S_{L^2}^z)$, which are arrays of the spin vectors (S_i^x, S_i^y, S_i^z) at all sites i $(= 1, 2, \cdots, L^2)$, are fed to the neural network as inputs for training. As the answer labels, we adopt temperatures T_n, at which the input spin configurations are generated. Here, the answer label is represented by a vector with 200 components. When the answer is $T/J = T_n/J$, only the nth component of the vector is set to be unity, while the other components are set to be zero. The neural network is trained so as to correctly guess temperatures at which the input spin configurations are generated.

After the training procedure, we focus on the weight matrix connecting the last hidden layer and the output
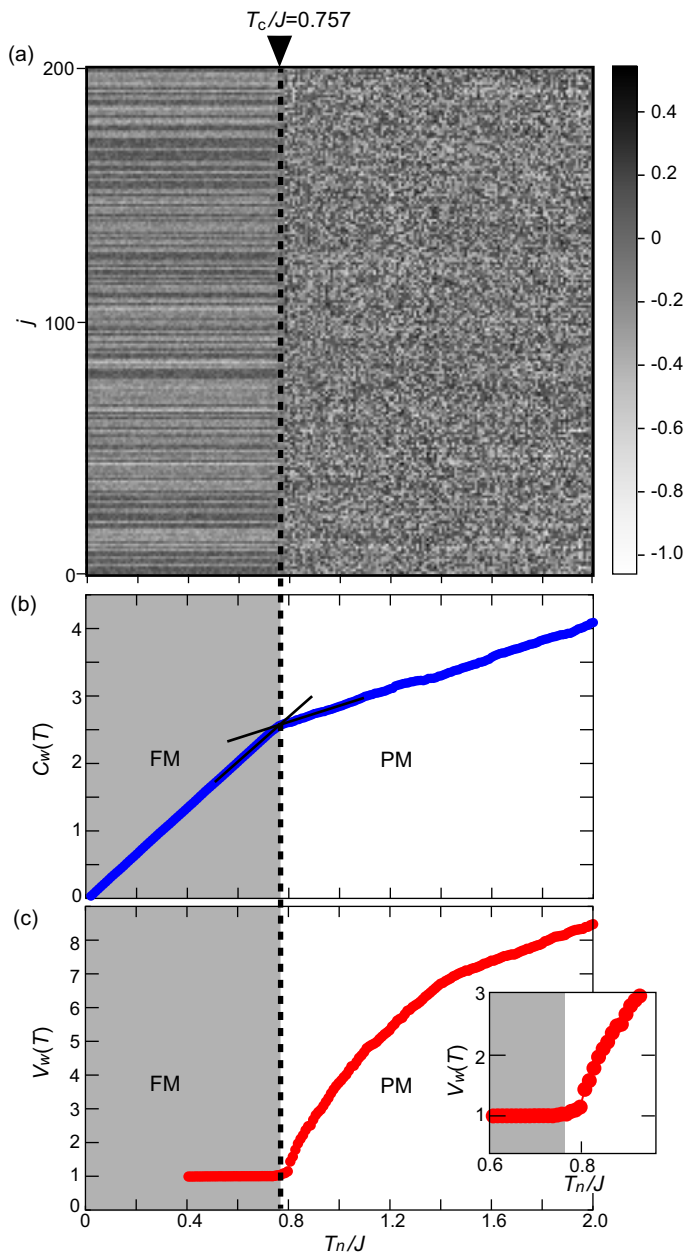
FIG. 7: (a) Heat map of the weight-matrix components $W_{j,n}$ in the temperature-identification-type neural network, which connect the jth node in the last hidden layer and the nth node in the output layer in the plane of j $(= 1, 2, \cdots, 200)$ and $T_n = n\Delta T$ $(n = 1, 2, \cdots, 200)$ for the Ising-like XXZ model with $\Delta = 1.05$ on the square lattice of $L = 80$. A clear change in pattern is seen at $T_c/J = 0.757$, which is ascribed to the phase transition from the paramagnetic (PM) phase to the ferromagnetic (FM) phase at this temperature. (b) Correlation function $C_W(T)$ of the weight-matrix components $W_{j,n}$ defined by Eq. (3) as a function of T_n, which clearly shows a change in slope at T_c. (c) Variance $V_W(T)$ of the weight-matrix components $W_{j,n}$ defined by Eq. (4) as a function of T_n, which shows an abrupt change at T_c.

layer. Components of the weight matrix are visualized as a heat map in Fig. 7(a). Here, the horizontal axis represents the index n of output nodes or corresponding temperature $T_n/J (= n\Delta T/J)$, and the vertical axis represents the index j of hidden nodes. A clear change in pattern is observed between the regimes above and below $T \approx 0.76$, where a stripe (sandstorm) pattern appears in the lower (higher) temperature regime. To analyze this pattern change more quantitatively, we calculate the correlation function $C_W(T)$ and variance $V_W(T)$, which are, respectively, defined by Eq. (4) and Eq. (5). Figures 7(b) and (c) show the temperature profiles of $C_W(T)$ and $V_W(T)$, respectively. The plot of $C_W(T)$ shows a clear change in slope associated with the phase transition at $T_c/J = 0.757$. The plot of $V_W(T)$ also shows a clear difference in behavior between the regimes above and below T_c. It is constant below T_c, starts increasing at T_c, and monotonically increases above T_c. The evaluated critical temperature coincides well with that obtained by the conventional Monte Carlo analysis.

Here it should be mentioned that the variance $V_W(T_n)$ is calculated using the weight-matrix components $W_{j,m}$ at discretized temperature points T_m ($m = 1, 2, \cdots, n - 1$) below T_n. When T_n is small, the number of the temperature points that can be used for the variance calculation, i.e., $n - 1$, is small so that the calculated variance $V_W(T_n)$ might contain a certain error. Therefore, the variances $V_W(T_n)$ in Fig. 7(c) are calculated only for $T_n/J > 0.4$ to suppress the errors. The calculated profile of $V_W(T_n)$ is very smooth, and the errors are negligibly small even around $T_n/J \sim 0.4$. We think that this variance-based analysis is applicable even to much lower temperatures.

V. RESULTS FOR $\Delta = 0.95$ (XY-LIKE CASE)

A. Monte Carlo calculation

We determine the critical temperature T_{BKT} of the BKT transition in the XY-like XXZ model from helicity modulus γ calculated by the Monte Carlo calculation. The helicity modulus of the XXZ model is calculated using the following formula given in the form of thermal average,

$$
\gamma_{\mathrm{XXZ}} = \frac{J}{L^2} \left\langle \sum_{\langle i,j \rangle_x} \sqrt{1 - S_{iz}^2} \sqrt{1 - S_{jz}^2} \cos(\phi_i - \phi_j) \right\rangle
$$
$$
- \frac{\beta J^2}{L^2} \left\langle \left(\sum_{\langle i,j \rangle_x} \sqrt{1 - S_{iz}^2} \sqrt{1 - S_{jz}^2} \sin(\phi_i - \phi_j) \right)^2 \right\rangle.
$$
$$(7)$$

For the derivation of this formula, see Appendix. This quantity exhibits a universal jump from $\gamma = 0$ to $\gamma = 2T_{\mathrm{BKT}}/\pi$ at T_{BKT}. For each lattice size, T_{BKT} is determined by the crossing point between the temperature

TABLE II: Critical temperatures of the XY-like XXZ model with $\Delta = 0.95$ for various system sizes. The case of $L \to \infty$ denote results in the thermodynamic limit.

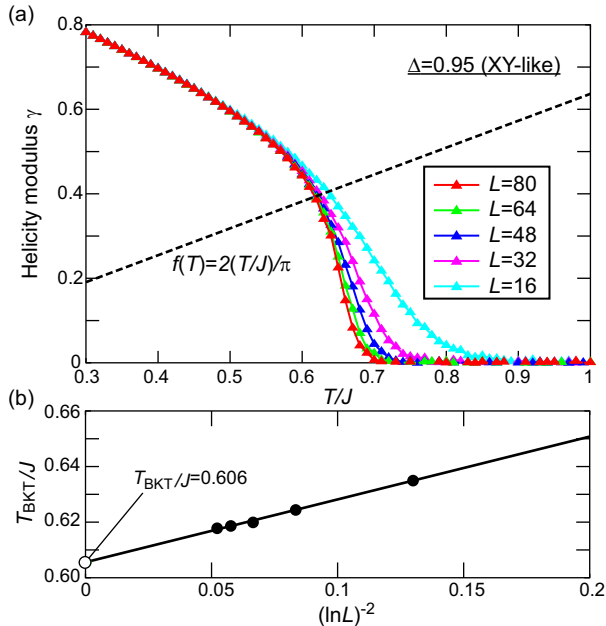| L	16	32	48	64	80	96	112	∞
$T_{\mathrm{BKT}}^{\mathrm{MC}}/J$	0.6351	0.6243	0.6201	0.6188	0.6177	-	-	0.606
$T_{\mathrm{BKT}}^{\mathrm{PC}}/J$	-	-	0.574(5)	0.576(2)	0.578(5)	0.582(9)	0.584(7)	0.603(5)

(a)

(b)

FIG. 8: (a) Temperature profiles of helicity modulus for the XY-like XXZ model with $\Delta = 0.95$ for various system sizes. A point at which the curve and the dashed line cross is assigned to the transition temperature T_{BKT} from the paramagnetic phase to the BKT phase for each system size. The dashed line represents a linear function $f(T) = 2(T/J)/\pi$. (b) Finite-size scaling analysis of T_{BKT}. In the thermodynamic limit of $L \to \infty$, T_{BKT} is evaluated to be $T_{\mathrm{BKT}}/J = 0.606$ by extrapolation.

profiles of γ and the line $f(T) = 2T/\pi$. Fig. 8(a) shows the temperature profiles of γ_{XXZ} for the XY-like XXZ model with $\Delta = 0.95$ for several lattice sizes.

We evaluate the BKT transition temperature in the thermodynamic limit $T_{\mathrm{BKT}}(\infty)$ by a finite-size scaling analysis. Kosterlitz and Thouless revealed that the correlation length behaves as $\exp(c/\sqrt{|t|})$ at temperatures slightly above $T_{\mathrm{BKT}}(\infty)$. Here c is a constant, and t is a relative temperature defined by,

$$t = \frac{T - T_{\mathrm{BKT}}(\infty)}{T_{\mathrm{BKT}}(\infty)}. \tag{8}$$

For finite-size systems, ξ equals to the lattice size L at T_{BKT}. Therefore, $T_{\mathrm{BKT}}(L)$ is given as a function of the system size L,

$$T_{\mathrm{BKT}}(L) = T_{\mathrm{BKT}}(\infty) + \frac{c'}{(\ln L)^2}. \tag{9}$$

The size dependence of $T_{\mathrm{BKT}}(L)$ for the XY-like XXZ model with $\Delta = 0.95$ is shown in Fig. 4(b). We evaluate $T_{\mathrm{BKT}}^{\mathrm{MC}}/J = 0.606$ in the thermodynamic limit from the extrapolation. The obtained values of $T_{\mathrm{BKT}}(L)$ for several sizes are summarized in Table II.

B. Vortex configurations used as training data

We prepare many vortex configurations of the XY model and the XY-like XXZ model with $\Delta = 0.95$ for several lattice sizes $L \times L$ ($L = 16, 32, 48, 64, 80$) as training data of the neural network, which are made from the spin configurations by calculating the vorticity on each plaquette. Here the spin configurations are generated by the Monte Carlo thermalization with 10^5 iterative steps based on the single-flip Metropolis algorithm. For the XY model and the XY-like XXZ model, vortex configurations are prepared at each temperature point.

The temperature phase diagram and the examples of vortex configurations for the XY model are presented in Figs. 9(a)-(c). The red circles and blue crosses indicate vortices and antivortices, respectively, in Figs. 9(b) and (c). Corresponding spin configurations are also shown in these figures where the arrows represent the in-plane components of spin vectors. One vortex-antivortex bound pair is seen in Fig. 9(b) for the BKT phase, whereas six free vortices and antivortex are seen in Fig. 9(c) for the paramagnetic phase. On the other hand, a schematic temperature phase diagram and the examples of vortex configurations for the XY-like XXZ model are presented in Figs. 9(d)-(f). Two vortex-antivortex bound pairs are seen in Fig. 9(e) for the BKT phase, whereas there are many free vortices and antivortices in Fig. 9(f) for the paramagnetic phase. Here the colors of pixels represent their out-of-plane components or the z components.

Here it should be mentioned that we have also examined the bare spin configurations as input data instead of the vortex configurations. We have found that the machine learning method often fails to detect the BKT transition in this case. The neural network tends to detect the growth of global magnetization component more sensitively than the evolution of vortex configuration, which makes it difficult to detect the BKT transition characterized by the bound and unbound states of vortex-antivortex pairs. In fact, this aspect has been pointed out in previous work [25]. This problem might be more crucial for the weakly easy-plane XXZ models with three-dimensional continuous vector spins than the
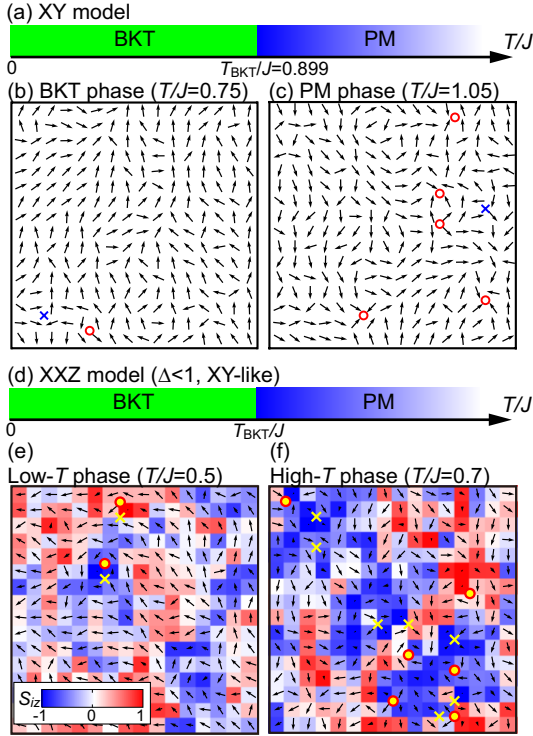
FIG. 9: Vortex configurations of the XY models on square lattices are used for data to train the neural network aiming for detection of the BKT transition in the XXZ model with $\Delta < 1$. (a) Schematic temperature phase diagram of the square-lattice XY model, which exhibits the BKT transition from the paramagnetic (PM) phase to the BKT phase at $T_{\rm BKT}/J = 0.899$. Vortex configurations generated by the Monte Carlo thermalization in the temperature range $0.75 \lesssim T/J \leq 0.89$ ($0.90 \leq T/J \lesssim 1.05$) are used for the training data for the BKT (PM) phase. (b), (c) Typical spin and vortex configurations of (b) BKT phase ($T/J = 0.75$, $L = 16$) and (c) PM phase ($T/J = 1.05$, $L = 16$). Locations of vortices and antivortices are indicated by solid circles and cross symbols. Spin configurations of the XXZ model with $\Delta = 0.95$ at various temperatures are entered to the trained neural network as input data. (d) Schematic temperature phase diagram of the XY-like XXZ model with $\Delta < 1$. (e), (f) Typical spin configurations at (e) lower temperatures ($T/J = 0.5$) and (f) higher temperatures ($T/J = 0.7$) of the XXZ model with $\Delta = 0.95$. It turns out that the neural network trained by the two-dimensional planar-spin configurations of the XY model can detect the phase transition of the three-dimensional vector-spin states in the XXZ model.

XY model with pure planar vector spins. In our previous work [35], we examined the q-state clock model and found that the machine-learning method can detect the BKT transition successfully even with bare spin configurations as input data, which might be attributable to the two-dimensional discrete spin degrees of freedom.
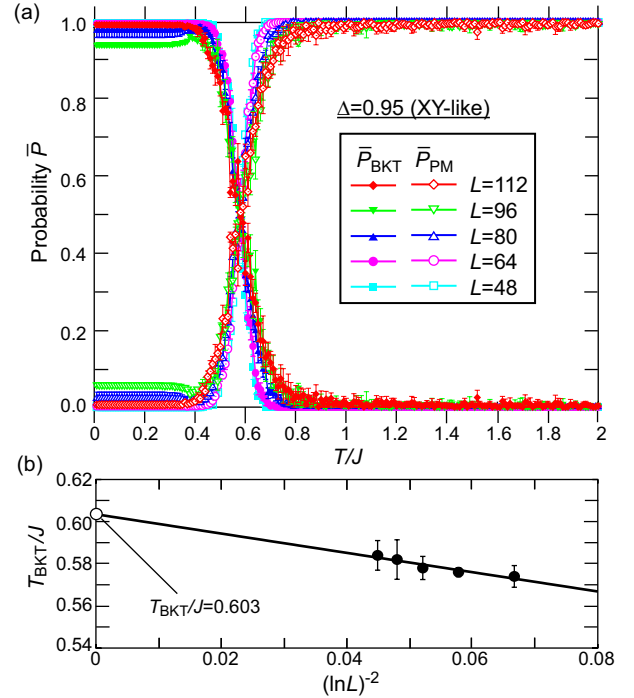

FIG. 10: (a) Temperature profiles of the averaged probabilities $\bar{P}_{\rm PM}$ and $\bar{P}_{\rm BKT}$ of the XY-like XXZ model with $\Delta = 0.95$ obtained by the phase-classification-type neural network for various system sizes, where the summation $\bar{P}_{\rm PM} + \bar{P}_{\rm BKT}$ is always normalized to be unity. Each crossing point of $\bar{P}_{\rm PM}$ and $\bar{P}_{\rm BKT}$ is assigned to a transition temperature for the corresponding system size. (b) Finite-size scaling analysis of $T_{\rm BKT}$. $T_{\rm BKT}$ in the thermodynamic limit of $L \to \infty$ is evaluated to be $T_{\rm BKT}/J = 0.603$ by extrapolation, which coincides well with $T_{\rm BKT}/J = 0.606$ evaluated by the conventional Monte Carlo analysis.

C. Phase-classification method

We attempt to detect the BKT phase transition in the XY-like XXZ model with $\Delta = 0.95$ by the phase-classification-type neural network [Fig. 2(a)] combined with the convolutional neural network [Fig. 3]. As the training dataset, 100 vortex configurations for the BKT (paramagnetic) phase in the XY model (instead of the XXZ model) are prepared at each temperature point within a range from $T \sim 0.75J$ to $T = 0.89J$ (from $T = 0.90J$ to $T \sim 1.05J$) at intervals of $\Delta T = 0.01J$.

We input vectors with L^2 components, which are arrays of the vorticities V_i on all plaquettes in the square lattice with periodic boundary conditions, to the neural network. On the other hand, the answer labels for the training data are vectors with two components, which represent probabilities that the input vortex configuration belongs to the BKT phase and the paramagnetic phase, respectively. More specifically, $(1,0)$ is given as answer labels for the input vortex configurations prepared at $T < T_{\rm BKT}$, whereas $(0,1)$ is given for those prepared at $T > T_{\rm BKT}$. The neural network is trained

so as to correctly guess phases to which the input vortex configurations belong.

After completing the training procedure, we input the vortex configurations of the XXZ model to the trained neural network, where each input is an array of the L^2 vorticities. As the input dataset, 100 vortex configurations are prepared at each temperature point within a range from $T = 0.01J$ to $T = 2.00J$ at intervals of $\Delta T = 0.01J$. For the 100 vortex configurations generated at T, 100 output vectors $(P_{\mathrm{BKT}}^l(T),\ P_{\mathrm{PM}}^l(T))$ are obtained, where $l\ (= 1, 2, \cdots, 100)$ is the index of the input data. We calculate the averages of $P_{\mathrm{BKT}}^l(T)$ and $P_{\mathrm{PM}}^l(T)$ over 100 input data to obtain the averaged probabilities $\bar{P}_{\mathrm{BKT}}(T)$ and $\bar{P}_{\mathrm{PM}}(T)$.

Figure 10(a) shows the temperature profiles of $\bar{P}_{\mathrm{BKT}}(T)$ and $\bar{P}_{\mathrm{PM}}(T)$. In these plots, $\bar{P}_{\mathrm{BKT}}(T) \approx 1$ and $\bar{P}_{\mathrm{PM}}(T) \approx 0$ at lower temperatures, whereas $\bar{P}_{\mathrm{BKT}}(T) \approx 0$ and $\bar{P}_{\mathrm{PM}}(T) \approx 1$ at higher temperatures, indicating that the neural network correctly guess phases to which the input data belong. The critical temperature T_{BKT} for each lattice size is determined from a crossing point of the two profiles. Moreover, T_{BKT} in the thermodynamic limit is evaluated as $T_{\mathrm{BKT}}^{\mathrm{PC}}/J = 0.603$ by the finite-size scaling analysis using Eq. (9). This value coincides well with the value $T_{\mathrm{BKT}}^{\mathrm{MC}}/J = 0.606$ obtained by the Monte Carlo calculation. The values of T_{BKT} for several lattice sizes evaluated by the present phase-classification method are summarized in Table II. Now we have succeeded in demonstrating the detections of the BKT transition in the XY-like XXZ model by the phase-classification-type neural network.

D. Temperature-identification method

Finally we attempt to detect the BKT transition in the XY-like XXZ model with $\Delta = 0.95$ by the temperature-identification-type neural network combined with the convolutional neural network. As the training dataset, 100 vortex configurations of the XXZ model are prepared at each temperature point ranged from $T = 0.01J$ to $T = 2.00J$ at intervals of $\Delta T = 0.01J$. We input the vortex configurations to the neural network in the form of vectors with L^2 components. As the answer labels, we adopt temperatures T_n, at which the input vortex configurations are generated. Again the neural network is trained so as to correctly guess temperatures at which the input vortex configurations are generated.

After the training procedure, we again focus on the weight matrix of the neural network connecting the last hidden layer and the output layer, components of which are visualized as a heat map in Fig. 11(a). We find that a stripe (sandstorm) pattern appears in the lower (higher) temperature regime. However, their boundary is difficult to identify by eyes because the stripe and sandstorm patterns are mixed in the in-between regime of $0.3 < T_n/J < 0.7$. To analyze the pattern change more quantitatively, we calculate the correlation func-
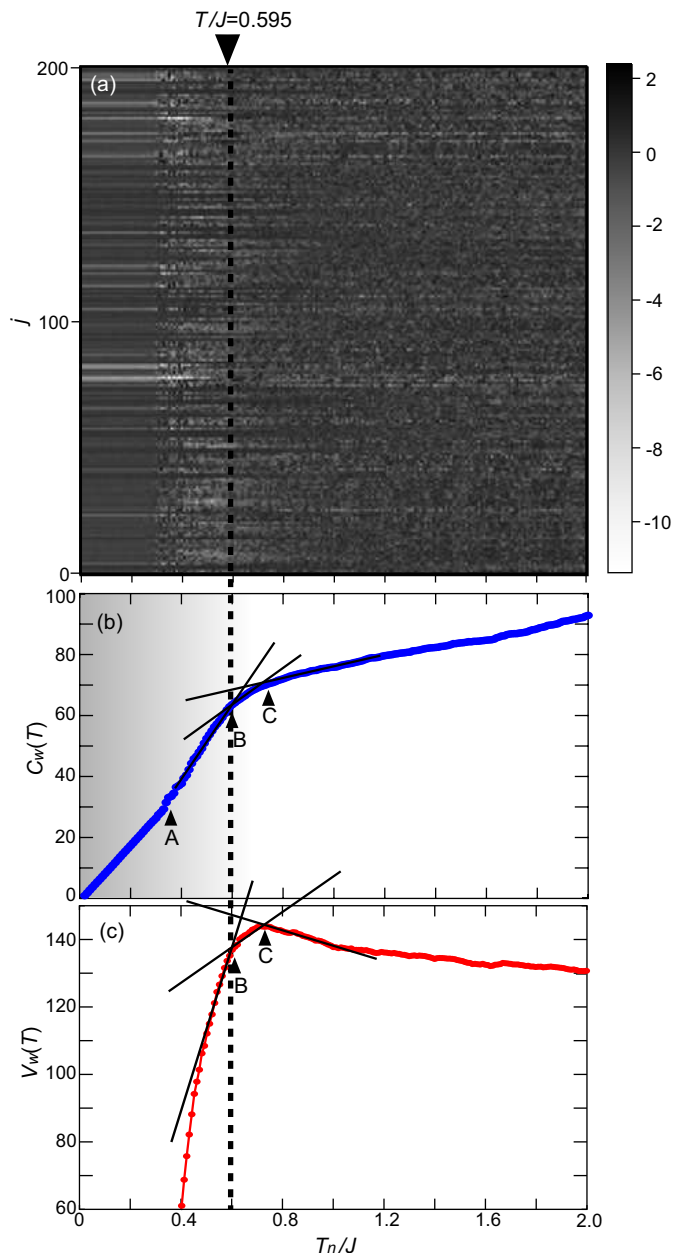

FIG. 11: (a) Heat map of the weight-matrix components $W_{j,n}$ in the temperature-identification-type neural network for the XY-like XXZ model with $\Delta = 0.95$ on the square lattice of $L = 80$. (b), (c) Temperature profiles of (b) correlation function $C_W(T)$ and (c) variance $V_W(T)$ of the weight-matrix components $W_{j,n}$. These profiles exhibit changes in slope at several points (A, B, and C), but A and C do not correspond to the BKT transition points. On the other hand, B is located almost at the real transition point of $T_{\mathrm{BKT}}/J = 0.606$.

tion $C_W(T)$ and variance $V_W(T)$ defined by Eq. (4) and Eq. (5). Again we calculated $V_W(T_n)$ for $T_n/J > 0.4$.

Figs. 11(b) and (c) present the temperature profiles of $C_W(T)$ and $V_W(T)$, respectively. We find that $C_W(T)$ exhibits changes in slope three times at points A, B, and C, whereas $V_W(T)$ exhibits abrupt changes in be-
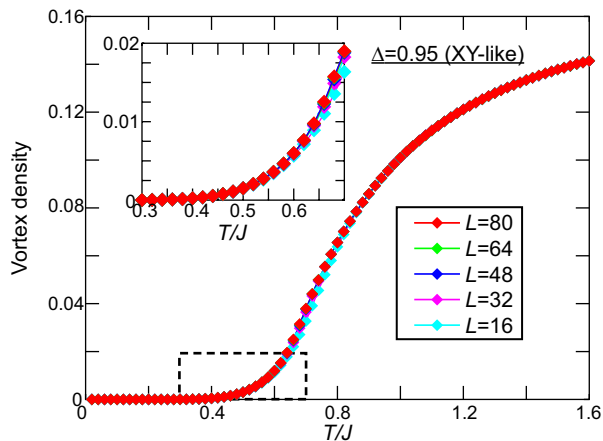
FIG. 12: Temperature profiles of the vortex density in the XY-like XXZ model with $\Delta = 0.95$ on square lattices for various system sizes. Inset magnifies an area indicated by the dashed rectangle, around which the vortex density starts increasing.

havior twice at B and C. To understand possible origins of these anomalies, we calculate temperature profiles of the vortex density for several lattice sizes of $L \times L$ with $L = 16, 32, 48, 64, 80$ (Fig. 12). Here, the vortex density is calculated as the total number of vortices and antivortices per one plaquette. In these profiles, we find that A corresponds to a temperature at which the profiles start to increase, whereas C corresponds to its inflection point. At the point A, the neural network might detect the emergence of vortices and antivortices. On the other hand, the anomaly C might be related with a peak of the specific heat. We indeed observed the specific-heat peak around this temperature in the Monte Carlo calculations (not shown). Because the specific heat and the vortex density are related with each other through the spin configurations, they could have anomalies (i.e., a peak and an inflection) simultaneously when a certain crossover occurs. The neural network might detect an anomaly associated with the crossover around the point C. On the contrary, the anomaly B seems to correspond to T_{BKT}. Accordingly, we conclude that the temperature-identification-type convolutional neural network seemingly captures the BKT transition in the XY-like XXZ model, but it fails to determine the transition point uniquely.

VI. CONCLUSION AND DISCUSSION

In summary, we have examined two machine learning methods, which are, respectively, referred to as the phase-classification method and the temperature-identification method, for detecting the second-order transition in the Ising-like XXZ model with a weak easy-axis anisotropy of $\Delta = 1.05$ and the BKT transition in the XY-like XXZ model with a weak easy-plane anisotropy $\Delta = 0.95$.

In the phase-classification method, the neural network trained by the spin configurations of the Ising model is used to classify the spin configurations of the Ising-like XXZ model into the ferromagnetic phase and the paramagnetic phase. On the other hand, the neural network trained by the vortex configurations of the XY model is used to classify the vortex configurations of the XY-like XXZ model into the BKT phase and the paramagnetic phase. For both cases, we have evaluated the critical temperatures for several system sizes and have performed finite-size scaling analyses. The obtained critical temperatures in the thermodynamic limit coincide well with those evaluated in the previous studies for both cases.

In the temperature-identification method, the neural network is trained so as to correctly answer the temperatures at which the spin (vortex) configurations of the XXZ model with $\Delta = 1.05$ ($\Delta = 0.95$) are generated by the Monte Carlo thermalization. After the training is completed, we focus on the weight matrix connecting the last hidden layer and the output layer to evaluate the critical temperatures. We have introduced two quatities, i.e., the correlation function $C_W(T)$ and the variance $V_W(T)$ to quantitatively analyze the pattern changes in the heat map that visualizes the weight-matrix components. The evaluated critical temperature for the second-order transition in the XXZ model with $\Delta = 1.05$ coincide with that obtained in the previous study. On the contrary, in the case of the XXZ model with $\Delta = 0.95$, the neural network detects not only the BKT transition but also other cross-over behaviors and thus fails to determine the BKT transition temperature uniquely.

The phase transitions in theoretical spin models have been usually studied by statistical analyses based on the Monte Carlo technique. Our machine learning methods have some advantages over this traditional technique. First, the computational cost can be significantly reduced. Both of our machine learning methods require the spin or vortex configurations generated by the Monte Carlo thermalization, but they do not necessarily have to be thermal equilibrium states. Namely even imperfectly thermalized spin or vortex configurations can work for the detection of the phase transitions, which enables us to save the numerical cost significantly. This is in striking contrast to the Monte Carlo method, which requires the huge number of samplings after sufficient thermalization procedure to improve the accuracy of the thermal averages. We have confirmed that the analysis which takes approximately an hour by our machine learning methods typically takes several hours or even a few days if it is done by the Monte Carlo method.

Second, our machine learning methods do not require the derivation of model-sensitive expression of the helicity modulus for detection of the BKT transition. As mentioned in the introduction section, the physical quantity called helicity modulus is usually used to detect the BKT transition in the Monte Carlo analysis. The expression of this quantity does not have a general form

but has a form specific to each spin model, which must be derived by hand for each case. Its derivation becomes difficult when the theoretical spin model is not simple. Our methods can avoid this difficulty and thus can be applied to general spin models. In addition, in the case of the phase-classification method, the neural network can be applied to other models without any extra training procedure, if it is once trained.

We should also mention that several machine learning methods have been applied to theoretical spin models as argued in the introduction section, most of which are based on supervised learning techniques. Our proposals have several advantages over them. One advantage is that our methods require no or less information of the model in advance. Conventional machine learning methods often require the prior knowledge of fundamental properties of the model, e.g, the number of phases, the approximate transition temperature(s), and even the order parameter(s) to assign the answer labels to the training data. Thus, they include a contradiction that the properties of the model we are about to examine must be known in advance. In contrast, both of our methods require the prior knowledge of neither the transition temperatures nor the order parameters. In the case of the temperature-identification method, we do not need to know even the number of phases.

For detection of phase transitions in theoretical spin models, the previously proposed machine learning methods often require preprocessing of input data to extract the features of the phases and phase transitions. In particular, for detection of the BKT phase, the methods using the vortex configurations, the spin-direction histogram, and some correlation functions have been proposed, which are made from the bare spin-configuration data. Among these quantities, the vortex configurations are relatively natural to be employed because the BKT behavior is characterized by the bound and un-bound states of the vortex-antivortex pairs. Moreover, they are easily made from the spin-configuration data. In this sense, the least requirement of the feature engineering is another advantage of our methods. Here we note that in Ref. [35], we applied the temperature-identification method to the q-state clock model with $q = 8$ and demonstrated that the BKT phase and the successive BKT transitions in this model can be successfully detected even using the bare spin configurations as input data. This might be attributed to the fact that less degrees of freedom of the two-dimensional discretized spin vectors of the clock models. On the contrary, the three-dimensional continuous spin vectors of the weakly easy-plane XXZ models have larger degrees of freedom, which might make the detection of the BKT behavior more difficult.

Finally, we discuss possible future challenges for further development of our work. First, the examination of the XXZ models with different anisotropy values might be worth doing. In the present work, we have examined only two cases with $\Delta = 1.05$ and $\Delta = 0.95$, which are, respectively, the Ising-like case with a weak easy-axis anisotropy and the XY-like case with a weak easy-plane anisotropy on the verge of their boundary. We chose these rather difficult cases to test the efficiency of our methods. However, it must be interesting to examine how the performances of our machine learning methods depend on the distance from the boundary at $\Delta = 1$, which is expected to scale with the clarity of phase transitions. Second, the application of the phase-classification method to other spin models that can host the BKT phase is also an issue of interest. Through this research, we might be able to clarify what features contained in the input data are indeed detected by the neural network and to establish a guiding principle of the systematic machine learning scheme for detection of phase transitions in spin models.

VII. APPENDIX A: HELICITY MODULUS FOR THE XXZ MODEL

The helicity modulus is a physical quantity which describes the degree of increase in free energy when a global twist is applied to the system. Specifically, in a system with lattice size of L in one direction, we consider the gradual twisting of spins with an equivalent twisting angle of δ/L from the left end to the right end so as to achieve a situation that the rightmost spin is twisted by an infinitesimal angle δ relative to the leftmost spin. When the Hamiltonian contains only terms symmetric with respect to the spin exchange, the increment of the free energy F in this case does not depend on the sense of the twist, i.e., the sign of δ, and thus can be expanded by the square of δ. The helicity modulus γ is defined as the coefficient of the leading term of the expansion as,

$$F(\delta) - F(0) = \frac{\gamma}{2}\delta^2 + \mathcal{O}(\delta^4). \tag{10}$$

Through transforming into the second-order differencial form and taking the limit of $\delta \to 0$, we see that γ is given by the second-order derivative of the free energy F with respect to δ as,

$$\gamma \approx 2\frac{F(\delta) - F(0)}{\delta^2} = \frac{1}{\delta}\left(\frac{F(\delta) - F(0)}{\delta} - \frac{F(0) - F(-\delta)}{\delta}\right) \approx \left.\frac{\partial^2 F}{\partial \delta^2}\right|_{\delta=0}. \tag{11}$$

The three-dimensional spin vector $\boldsymbol{S}_i$ can be written with S_i^z and ϕ_i as,

$$\boldsymbol{S}_i = (S_i^x, S_i^y, S_i^z) = \left(\sqrt{1 - S_i^{z\,2}} \cos \phi_i, \sqrt{1 - S_i^{z\,2}} \sin \phi_i, S_i^z \right). \tag{12}$$

Using this expression, the distribution function $Z(\delta)$ of the twisted system is given by,

$$Z(\delta) = \int_{-2\pi}^{-2\pi} d\phi_1 \int_{-2\pi}^{-2\pi} d\phi_2 \cdots \int_{-2\pi}^{-2\pi} d\phi_{L^2} \; e^{-\beta \tilde{\mathcal{H}}(\delta)} \equiv \int \mathcal{D}\phi \; e^{-\beta \tilde{\mathcal{H}}(\delta)}. \tag{13}$$

Here $\tilde{\mathcal{H}}(\delta)$ is the Hamiltonian of the twisted system. The helicity modulus γ is calculated as,

$$\begin{aligned}
\gamma &= \left. \frac{\partial^2 F}{\partial \delta^2} \right|_{\delta=0} = \left. \frac{\partial^2}{\partial \delta^2} \left(-\frac{1}{\beta} \ln Z(\delta) \right) \right|_{\delta=0} \\
&= \left. \frac{\int \mathcal{D}\phi \, \frac{\partial^2 \tilde{\mathcal{H}}}{\partial \delta^2} e^{-\beta \tilde{\mathcal{H}}}}{Z(\delta)} \right|_{\delta=0} - \beta \left. \frac{\int \mathcal{D}\phi \left(\frac{\partial \tilde{\mathcal{H}}}{\partial \delta} \right)^2 e^{-\beta \tilde{\mathcal{H}}}}{Z(\delta)} \right|_{\delta=0} + \beta \left. \left(\frac{\int \mathcal{D}\phi \, \frac{\partial \tilde{\mathcal{H}}}{\partial \delta} e^{-\beta \tilde{\mathcal{H}}}}{Z(\delta)} \right)^2 \right|_{\delta=0} \\
&= \left. \left\langle \frac{\partial^2 \tilde{\mathcal{H}}}{\partial \delta^2} \right\rangle \right|_{\delta=0} - \beta \left. \left\langle \left(\frac{\partial \tilde{\mathcal{H}}}{\partial \delta} \right)^2 \right\rangle \right|_{\delta=0} + \beta \left. \left\langle \frac{\partial \tilde{\mathcal{H}}}{\partial \delta} \right\rangle^2 \right|_{\delta=0}.
\end{aligned} \tag{14}$$

When the system contains only symmetric exchange interactions associated with $\boldsymbol{S}_i \cdot \boldsymbol{S}_j$, the following relation holds,

$$\left. \left\langle \frac{\partial \tilde{\mathcal{H}}}{\partial \delta} \right\rangle \right|_{\delta=0} = \left. \left\langle \frac{\partial F}{\partial \delta} \right\rangle \right|_{\delta=0} = 0. \tag{15}$$

Consequently, the last term in Eq. (14) vanishes, and we obtain the following expression for a general form of the helicity modulus,

$$\gamma = \left. \left\langle \frac{\partial^2 \tilde{\mathcal{H}}}{\partial \delta^2} \right\rangle \right|_{\delta=0} - \beta \left. \left\langle \left(\frac{\partial \tilde{\mathcal{H}}}{\partial \delta} \right)^2 \right\rangle \right|_{\delta=0}. \tag{16}$$

Now we discuss the derivation of the formula of the helicity modulus for the XXZ model. The Hamiltonian of the XXZ model can be written using S_i^z and ϕ_i as,

$$\begin{aligned}
\mathcal{H} &= -J \sum_{<i,j>} \left\{ \sqrt{1 - S_i^{z\,2}} \sqrt{1 - S_j^{z\,2}} (\cos \phi_i \cos \phi_j + \sin \phi_i \sin \phi_j) + \Delta S_i^z S_j^z \right\} \\
&= -J \sum_{<i,j>} \left\{ \sqrt{1 - S_i^{z\,2}} \sqrt{1 - S_j^{z\,2}} \cos(\phi_i - \phi_j) + \Delta S_i^z S_j^z \right\}.
\end{aligned} \tag{17}$$

We consider the twisting of spin alignment in the x direction. Subsequently, the term in the Hamiltonian associated with the bonds along the x axis is modified as,

$$\tilde{\mathcal{H}}_x(\delta) = -J \sum_{<i,j>_x} \left\{ \sqrt{1 - S_i^{z\,2}} \sqrt{1 - S_j^{z\,2}} \cos \left(\phi_i - \phi_j - \frac{\delta}{L} \right) + \Delta S_i^z S_j^z \right\}. \tag{18}$$

Note that the term associated with the bonds along the y axis does not change. For the above Hamiltonian, we obtain,

$$\frac{\partial^2 \tilde{\mathcal{H}}}{\partial \delta^2} = \frac{\partial^2 \tilde{\mathcal{H}}_x}{\partial \delta^2} = \frac{J}{L^2} \sum_{<i,j>_x} \sqrt{1 - S_i^{z\,2}} \sqrt{1 - S_j^{z\,2}} \cos \left(\phi_i - \phi_j - \frac{\delta}{L} \right), \tag{19}$$

$$\frac{\partial \tilde{\mathcal{H}}}{\partial \delta} = \frac{\partial \tilde{\mathcal{H}}_x}{\partial \delta} = \frac{J}{L} \sum_{<i,j>_x} \sqrt{1 - S_i^{z\,2}} \sqrt{1 - S_j^{z\,2}} \sin \left(\phi_i - \phi_j - \frac{\delta}{L} \right). \tag{20}$$

Eventually, we obtain the following expression of the helicity modulus for the XXZ model,

$$
\gamma_{\mathrm{XXZ}} = \left\langle \frac{\partial^2 \tilde{\mathcal{H}}}{\partial \delta^2} \right\rangle \bigg|_{\delta=0} - \beta \left\langle \left(\frac{\partial \tilde{\mathcal{H}}}{\partial \delta} \right)^2 \right\rangle \bigg|_{\delta=0}
$$

$$
= \frac{J}{L^2} \left\langle \sum_{<i,j>_x} \sqrt{1 - S_i^{z\,2}} \sqrt{1 - S_j^{z\,2}} \cos(\phi_i - \phi_j) \right\rangle - \frac{\beta J^2}{L^2} \left\langle \left(\sum_{<i,j>_x} \sqrt{1 - S_i^{z\,2}} \sqrt{1 - S_j^{z\,2}} \sin(\phi_i - \phi_j) \right)^2 \right\rangle. \quad (21)
$$

Sustituting $S_i^z = 0$ into this expression, we obtain the expression for the XY model as,

$$
\gamma_{\mathrm{XY}} = \frac{J}{L^2} \left\langle \sum_{<i,j>_x} \cos(\phi_i - \phi_j) \right\rangle - \frac{\beta J^2}{L^2} \left\langle \left(\sum_{<i,j>_x} \sin(\phi_i - \phi_j) \right)^2 \right\rangle. \quad (22)
$$

VIII.   ACKNOWLEDGEMENT

[1] Y. LeCun, Y. Bengio, and G. Hinton, *Deep learning*, Nature **521**, 436 (2015).

[2] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (Adaptive Computation and Machine Learning series) (MIT Press, Cambridge, MA, 2016).

[3] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, *Machine learning and the physical sciences*, Rev. Mod. Phys. **91**, 045002 (2019).

[4] E. Bedolla, L. C. Padierna, and R. Castañeda-Priego, *Machine learning for condensed matter physics*, J. Phys.:Condens. Matter **33**, 053001 (2021).

[5] G. Carleo and M. Troyer, *Solving the quantum many-body problem with artificial neural networks*, Science **355**, 602 (2017).

[6] Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, *Restricted Boltzmann machine learning for solving strongly correlated quantum systems*, Phys. Rev. B **96**, 205152 (2017).

[7] Y. Nomura and M. Imada, *Dirac-Type Nodal Spin Liquid Revealed by Refined Quantum Many-Body Solver Using Neural-Network Wave Function, Correlation Ratio, and Level Spectroscopy*, Phys. Rev. X **11**, 031034 (2021).

[8] T. Ohtsuki and T. Mano, *Drawing Phase Diagrams of Random Quantum Systems by Deep Learning the Wave Functions*, J. Phys. Soc. Jpn. **89**, 022001 (2020).

[9] A. Tanaka and A. Tomiya, *Detection of Phase Transition via Convolutional Neural Networks*, J. Phys. Soc. Jpn. **86**, 063001 (2017).

[10] J. Carrasquilla and R.G. Melko, *Machine learning phases of matter*, Nat. Phys. **13**, 431 (2017).

[11] S. Arai, M. Ohzeki, and K. Tanaka, *Deep Neural Network Detects Quantum Phase Transition*, J. Phys. Soc. Jpn. **87**, 033001 (2018).

[12] P. Suchsland and S. Wessel, *Parameter diagnostics of phases and phase transition learning by neural networks*, Phys. Rev. B **97**, 174435 (2018).

[13] W. Hu, R. R. P. Singh, and R. T. Scalettar, *Discovering phases, phase transitions, and crossovers through unsupervised machine learning: A critical examination*, Phys. Rev. E **95**, 062122 (2017).

[14] S. J. Wetzel, *Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders*, Phys. Rev. E **96**, 022140 (2017).

[15] L. Wang, *Discovering phase transitions with unsupervised learning*, Phys. Rev. B **94**, 195105 (2016).

[16] P. Ponte and R. G. Melko, *Kernel methods for interpretable machine learning of order parameters*, Phys. Rev. B **96**, 205146 (2017).

[17] E. P. L. van Nieuwenburg, Y. H. Liu, and S. D. Huber, *Learning phase transitions by confusion*, Nat. Phys. **13**, 435 (2017).

[18] Y. H. Liu and E. P. L. van Nieuwenburg, *Discriminative Cooperative Networks for Detecting Phase Transitions*, Phys. Rev. Lett. **120**, 176401 (2018).

[19] C. Giannetti, B. Lucini, and D. Vadacchino, *Machine Learning as a universal tool for quantitative investigations of phase transitions*, Nucl. Phys. B **944**, 114639 (2019).

[20] D. Bachtis, G. Aarts, and B. Lucini, *Extending machine learning classification capabilities with histogram reweighting*, Phys. Rev. E **102**, 033303 (2020).

[21] N. D. Mermin and H. Wagner, *Absence of Ferromagnetism or Antiferromagnetism in One- or Two-Dimensional Isotropic Heisenberg Models* , Phys. Rev. Lett. **17**, 1133 (1966).

[22] J. M. Kosterlitz and D. J. Thouless, *Ordering, metastability and phase transitions in two-dimensional systems*, J. Phys. C **6**, 1181 (1973).

[23] J. M. Kosterlitz, *The critical properties of the two-dimensional xy model*, J. Phys. C **7**, 1046 (1974).

[24] M. Richter-Laskowska, H. Khan, N. Trivedi, and M. M.

Maśka, *A machine learning approach to the Berezinskii-Kosterlitz-Thouless transition in classical and quantum models* , Cond. Mat. Phys. **21**, 33602 (2018).

[25] M. J. S. Beach, A. Golubeva, and R. G. Melko, *Machine learning vortices at the Kosterlitz-Thouless transition*, Phys. Rev. B **97**, 045207 (2018).

[26] W. Zhang, J. Liu, and T. C. Wei, *Machine learning phase transitions in the percolation and XY models*, Phys. Rev. E **99**, 032142 (2019).

[27] J. F. Rodriguez-Nieva and M. S. Scheurer, *Identifying topological order through unsupervised machine learning*, Nat. Phys. **15**, 790 (2019).

[28] Q. H. Tran, M. Chen, and Y. Hasegawa, *Topological persistence machine of phase transitions*, Phys. Rev. E **103**, 052127 (2021).

[29] J. Wang, W. Zhang, T. Hua, and T. C. Wei, *Unsupervised learning of topological phase transitions using the Calinski-Harabaz index*, Phys. Rev. Res. **3**, 013074 (2021)

[30] T. Mendes-Santos, X. Turkeshi, M. Dalmonte, and A. Rodriguez, *Unsupervised Learning Universal Critical Behavior via the Intrinsic Dimension*, Phys. Rev. X **11**, 011040 (2021).

[31] J. Singh, M. S. Scheurer, and V. Arora, *Conditional generative models for sampling and phase transition indication in spin systems*, SciPost Phys. **11**, 043 (2021).

[32] S. Haldar, S. S. Rahaman, and M. Kumar, *Study of the Berezinskii-Kosterlitz-Thouless transition: An unsupervised machine learning approach*, arXiv:2205.15151.

[33] K. Shiina, H. Mori, Y. Okabe, and H. K. Lee, *Machine-Learning Studies on Spin Models*, Sci. Rep. **10**, 2177 (2020).

[34] Y. Tomita, K. Shiina, Y. Okabe, and H. K. Lee, *Machine-learning study using improved correlation configuration and application to quantum Monte Carlo simulation*, Phys. Rev. E **102**, 021302(R) (2020).

[35] Y. Miyajima, Y. Murata, Y. Tanaka, and M. Mochizuki, *Machine learning detection of Berezinskii-Kosterlitz-Thouless transitions in q-state clock models*, Phys. Rev. B **104**, 075114 (2021).

[36] Z. Hu, Z. Ma, Y.-D. Liao, H. Li, C. Ma, Yi Cui, Y. Shangguan, Z. Huang, Y. Qi, W. Li, Zi Y. Meng, J. Wen, and W. Yu, *Evidence of the Berezinskii-Kosterlitz-Thouless phase in a frustrated magnet*, Nat. Commun. **11**, 5631 (2020).

[37] U. Tutsch, B. Wolf, S. Wessel, L. Postulka, Y. Tsui, H. O. Jeschke, I. Opahle, T. Saha-Dasgupta, R. Valentí, A. Brühl, K. Remović-Langer, T. Kretz, H.-W. Lerner, M. Wagner, and M. Lang, *Evidence of a field-induced Berezinskii-Kosterlitz-Thouless scenario in a two-dimensional spin-dimer system*, Nat. Commun. **5**, 5169 (2014).

[38] Y. Kohama, M. Jaime, O. E. Ayala-Valenzuela, R. D. McDonald, E. D. Mun, J. F. Corbey, and J. L. Manson, *Field-induced XY and Ising ground states in a quasi-two-dimensional S=1/2 Heisenberg antiferromagnet*, Phys. Rev. B **84**, 184402 (2011).

[39] A. Cuccoli, T. Roscilde, R. Vaia, and P. Verrucchi, *Detection of XY Behavior in Weakly Anisotropic Quantum Antiferromagnets on the Square Lattice*, Phys. Rev. Lett. **90**, 167205 (2003).

[40] A. Cuccoli, T. Roscilde, R. Vaia, and P. Verrucchi, *Field-induced XY behavior in the S=1/2 antiferromagnet on the square lattice*, Phys. Rev. B **68**, 060402(R) (2003).

[41] D. Kim and D.-H. Kim, *Smallest neural network to learn the Ising criticality*, Phys. Rev. E **98**, 022138 (2018).

[42] D. Bachtis, G. Aarts, and B. Lucini, *Mapping distinct phase transitions to a neural network*, Phys. Rev. E **102**, 053306 (2020).

[43] K. Fukushima and K. Sakai, *Can a CNN trained on the Ising model detect the phase transition of the q-state Potts model?*, Prog. Theor. Exp. Phys. **2021**, 061A01 (2021).

[44] S. Hikami and T. Tsuneto, *Phase Transition of Quasi-Two Dimensional Planar System*, Prog. Theor. Phys. **63**, 387 (1980).

[45] C. Kawabata and A. R. Bishop, *Monte Carlo simulation of the two-dimensional classical Heisenberg model with easy-plane anisotropy*, Solid State Commun. **42**, 595 (1982).

[46] P. A. Serena, N. García, and A. Levanyuk, *Monte Carlo calculations on the two-dimensional anisotropic Heisenberg model*, Phys. Rev. B **47**, 5027 (1993).

[47] A. Cuccoli, V. Tognetti, and R. Vaia, *Two-dimensional XXZ model on a square lattice: A Monte Carlo simulation*, Phys. Rev. B **52**, 10221 (1995).

[48] A. S. T. Pires, *Classical two-dimensional XXZ model: A test of a generalized self-consistent harmonic approximation*, Phys. Rev. B **54**, 6081 (1996).

[49] K. W. Lee and C. E. Lee, *Monte Carlo study of the Kosterlitz-Thouless transition in the Heisenberg model with antisymmetric exchange interactions*, Phys. Rev. B **72**, 054439 (2005).

[50] K. Aoyama and H. Kawamura, *Effects of magnetic anisotropy on spin and thermal transport in classical antiferromagnets on the square lattice*, Phys. Rev. B **100**, 144416 (2019).

[51] A. A. Shirinyan, V. K. Kozin, J. Hellsvik, M. Pereiro, O. Eriksson, and D. Yudin, *Self-organizing maps as a method for detecting phase transitions and phase identification*, Phys. Rev. B **99**, 041108(R) (2019).

[52] M. E. Fisher, M. N. Barber, and D. Jasnow, *Helicity Modulus, Superfluidity, and Scaling in Isotropic Systems*, Phys. Rev. A **8**, 1111 (1973).

[53] D. R. Nelson and J. M. Kosterlitz, *Universal Jump in the Superfluid Density of Two-Dimensional Superfluids*, Phys. Rev. Lett. **39**, 1201 (1977).

[54] H. Weber and P. Minnhagen, *Monte Carlo determination of the critical temperature for the two-dimensional XY model*, Phys. Rev. B **37**, 5986 (1988).

[55] F. Chollet *et al.*, *KERAS* (2015), https://github.com/keras-team/keras

[56] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, arXiv:1412.6980.