Machine-learning study of phase transitions in Ising, Blume-Capel, and Ising-metamagnet models

Vasanth Kumar Babu [*] and Rahul Pandit [†]
Centre for Condensed Matter Theory, Department of Physics, Indian Institute of Science, Bangalore 560012, India

(Received 2 February 2025; accepted 27 May 2025; published 18 June 2025)

We combine machine-learning techniques with Monte Carlo (MC) simulations and finite-size scaling (FSS) to study continuous and first-order phase transitions in Ising, Blume-Capel, and Ising-metamagnet spin models. We go beyond earlier studies that had concentrated on obtaining the correlation-length exponent ν. In particular, we show (a) how to combine neural networks (NNs), trained with data from MC simulations of Ising-type spin models on finite lattices, with FSS to obtain both thermal magnetic exponents $y_t = 1/\nu$ and y_h, respectively, at both critical and tricritical points, (b) how to obtain the NN counterpart of two-scale-factor universality at an Ising-type critical point, and (c) how to get FSS at a first-order transition. We also obtain the FSS forms for the output of our trained NNs as functions of both the temperature and the magnetic field.

DOI: 10.1103/j3lx-rhyw

I. INTRODUCTION

The development of the theory of phase transitions is among the most important advances in theoretical physics over the last 60 years [1–8]. We can now obtain, with great accuracy, the universal critical exponents that characterize the universality classes of different continuous phase transitions. Examples of such transitions include the liquid-gas critical point, the Curie and Néel transitions in ferromagnets and antiferromagnets, respectively, the normal-superconductor transition, and the change from a normal fluid to a superfluid.

Some studies have explored intriguing intersections between statistical physics and machine learning (ML). Ideas from statistical physics are frequently employed to gain insights into ML models, even as the application of ML methods in physics moves apace (see, e.g., Refs. [9–11]). A diverse array of ML techniques are being utilized to explore properties of various models of statistical physics, such as Ising models [12–19], directed percolation [15,20–23], and models for nonequilibrium transitions [24,25]. Furthermore, the ML-aided classification of phases and phase transition has attracted significant attention [12,14–18,20,23,26–31]; both supervised and unsupervised ML methods have been used in such classification. For example, ML-assisted dimensionality reduction, in the vicinity of critical points, has used principal component analysis (PCA) and autoencoders (see, e.g., [27,32,33]).

Phase transitions can occur only in the thermodynamic limit, i.e., when the linear system size $L \to \infty$. However, finite-size-scaling (FSS) analysis provides an effective method for estimating, from finite-size calculations, thermodynamic functions and their singularities, for both continuous and first-order transitions [8,34–37]. Traditionally, FSS is used directly with thermodynamic functions like the magnetization M in a ferromagnet. Recent studies have demonstrated that outputs from machine-learning models, such as neural networks (NNs), can be analyzed with FSS to obtain thermal critical exponents, like ν, at Ising-type critical points [12,14,16–18,20,23]. We extend these studies significantly by working with the two-dimensional Ising, Blume-Capel, and Ising-metamagnet models and demonstrating (a) how to combine NNs, trained with data from Monte Carlo simulations of Ising-type spin models on finite lattices, with FSS to obtain both thermal magnetic exponents $y_t = 1/\nu$ and y_h, respectively, at *both critical and tricritical points*, (b) how to obtain the NN counterpart of two-scale-factor universality [38,39] at an Ising-type critical point, and (c) how to combine NN methods and FSS at a first-order transition. We also use the probability distribution functions (PDFs) of the magnetization M, near the critical point, to obtain general forms for the output of our trained NNs as functions of the temperature and the magnetic field; this has not been attempted hitherto to the best of our knowledge. We also demonstrate, for the Blume-Capel [40–43] and Ising-metamagnet [44–47] models, that NNs, trained near *zero-field* critical points, can successfully uncover scaling properties, in particular critical exponents and scaling functions, in the vicinities of critical points or lines at *nonzero values of these fields*. (We use the word *field* in a generalized sense for parameters in these models; such fields include, e.g., the external magnetic and crystal fields for the Blume-Capel model [40–43], and the external magnetic and staggered-magnetic fields for the Ising-metamagnet model [44–47].) Our method does not yield new physical results for these phase transitions, nor has it achieved, so far, the level of accuracy for critical exponents calculated by high-resolution Monte Carlo simulations, conformal-bootstrap analysis, or series-expansion methods [48–57]. However, our NN-based study is able to exploit *transfer learning*, whereby the zero-field-trained NN suffices to study critical points at nonzero fields.

Our paper is organized as follows: In Sec. II we define the Ising, Blume-Capel, and Ising-metamagnet models, the

[*]Contact author: vasanthb@iisc.ac.in
[†]Contact author: rahul@iisc.ac.in

Monte Carlo and finite-size scaling methods we use, and the neural network architectures we employ. We then present our results in Sec. III. Section IV is devoted to a discussion of our results and conclusions.

II. MODELS AND METHODS

A. Models

We consider the following three spin models for two-dimensional (2D) square lattices: the ferromagnetic Ising model (Sec. II A 1), the Blume-Capel model (Sec. II A 2), and the Ising-metamagnet (Sec. II A 3).

1. Ising ferromagnet

The Ising model is defined by the Hamiltonian

$$\mathcal{H}_I = -J \sum_{\langle i,j \rangle} S_i S_j - H \sum_i S_i, \tag{1}$$

where the Ising spins $S_i = \pm 1$ and $\langle i, j \rangle$ denotes nearest-neighbor pairs of sites [58–61] and we consider a square lattice with $N = L^2$ sites, labeled by i. There is a magnetic field H at every site, and the exchange coupling $J > 0$; i.e., we consider the ferromagnetic case, which has completely aligned spins at temperature $T = 0$ and, at $H = 0$, two coexisting phases: the $\uparrow$ phase with $S_i = 1$, $\forall i$, and the $\downarrow$ phase with $S_i = -1$, $\forall i$. The equilibrium statistical mechanics of this model follows from its intensive bulk free energy f_B, which can be obtained exactly [58–61] if $H = 0$, whence we know that this Ising model exhibits a critical point at $H = 0$ and $T = T_c$, with

$$\sinh\left[\frac{2J}{k_B T_c^{eq}}\right] = 1,$$
$$\Rightarrow k_B T_c = \frac{2J}{\ln(1+\sqrt{2})}, \tag{2}$$

where k_B is the Boltzmann constant. Henceforth, we use units in which k_B and J are set to 1. The *order parameter* for this model is the magnetization per spin $M = \langle \sum_i S_i \rangle / N$, where the angular brackets denote thermal averages like $\langle S_i \rangle = \sum_{\{S_j\}} [S_i \exp(-\beta \mathcal{H}_I)] / \sum_{\{S_j\}} [\exp(-\beta \mathcal{H}_I)]$, where $\sum_{\{S_j\}}$ denotes the sum over all spin states, and $\beta \equiv 1/(k_B T)$. This magnetization and the two-spin correlation length ξ show the following power-law behaviors in the vicinity of the critical point:

$$\begin{aligned} M &\sim |t|^{\beta}, \quad t \to 0^- \text{ and } h = 0, \\ M &\sim |h|^{\frac{1}{\delta}}, \quad h \to 0^{\pm} \text{ and } t = 0, \\ \xi &\sim |t|^{-\nu}, \quad t \to 0^{\pm} \text{ and } h = 0, \\ \xi &\sim |h|^{-\frac{\nu}{\beta\delta}}, \quad h \to 0^{\pm} \text{ and } t = 0, \end{aligned} \tag{3}$$

the reduced temperature $t = \frac{T - T_c}{T_c}$, and $t \to 0^-$ indicates that t approaches 0 from below; $t \to 0^{\pm}$ means that t approaches 0 either from above or below. The exponents $\beta = 1/8$, $\nu = 1$, and $\delta = 15$ for this 2D Ising model (and all models in this universality class [7,8,58–61]).

2. Blume-Capel model

We also study the Blume-Capel model on a 2D square lattice with nearest-neighbor interactions and the Hamiltonian [40–43]

$$\mathcal{H}_{\rm BC} = -J \sum_{\langle i,j \rangle} \mathcal{S}_i \mathcal{S}_j + \Delta \sum_i \mathcal{S}_i^2 - H \sum_i \mathcal{S}_i, \tag{4}$$

where the spins $\mathcal{S}_i = \pm 1, 0$, the ferromagnetic coupling $J > 0$, and Δ and H are, respectively, the crystal field and the magnetic field H. In the Δ-T plane, with the $H = 0$, this model exhibits a line of first-order transitions and another line of 2D Ising-type second-order transitions that meet at a tricritical point [40–43], which has distinct critical exponents. This model (4) has the following two order parameters:

$$M \equiv \frac{\langle \sum_i \mathcal{S}_i \rangle}{N}; \quad Q \equiv \frac{\left\langle \sum_i \mathcal{S}_i^2 \right\rangle}{N}. \tag{5}$$

3. Ising-metamagnet

We consider the Ising-metamagnet on a two-dimensional (2D) square lattice, with $N = L^2$ sites, labeled by i, nearest-neighbor (*nn*) antiferromagnetic interactions $J_1 < 0$, and next-nearest-neighbor (*nnn*) ferromagnetic interactions $J_2 > 0$, and the Hamiltonian [44–47]

$$\begin{aligned} \mathcal{H}_M = &- J_1 \sum_{\langle i,j \rangle_{nn}} S_i S_j - J_2 \sum_{\langle i,j \rangle_{nnn}} S_i S_j \\ &- H \sum_i S_i - H_s \left[\sum_{i \in A} S_i - \sum_{i \in B} S_i \right], \end{aligned} \tag{6}$$

where the Ising spins $S_i = \pm 1$; H and H_s are, respectively, the external magnetic field and the *staggered* magnetic field; and A and B are the two interpenetrating square sublattices that comprise our original bipartite square lattice.

The fields H and H_s are thermodynamically conjugate to the magnetization M and the *staggered* magnetization M_s order parameters for this system; these are defined as follows:

$$\begin{aligned} M_A &\equiv \frac{\sum_{i \in A} \langle S_i \rangle}{N_A}, \quad M_B \equiv \frac{\sum_{i \in B} \langle S_i \rangle}{N_B}, \\ M &\equiv M_A + M_B, \quad M_s \equiv M_A - M_B, \end{aligned} \tag{7}$$

where $N_A = N_B = N/2$.

In the H-T plane, with $H_s = 0$ and $H \geqslant 0$, the model (6) exhibits a line of first-order transition and a line of second-order transitions, in the 2D Ising universality class, that meet at a tricritical point [44–47].

B. Monte Carlo simulations

For arbitrary values of the exchange couplings $J, \ldots$, fields $H, \ldots$, and the temperature T, the bulk free energy f_B and other thermodynamic functions like M cannot be obtained analytically for the models (1), (4), and (6). Here we use standard Metropolis Monte Carlo methods [62], with non-conserved order parameters, to obtain order parameters and their probability distributions. Furthermore, we use various spin configurations, which emerge from our MC simulations at given values of couplings, fields, and the temperature T,

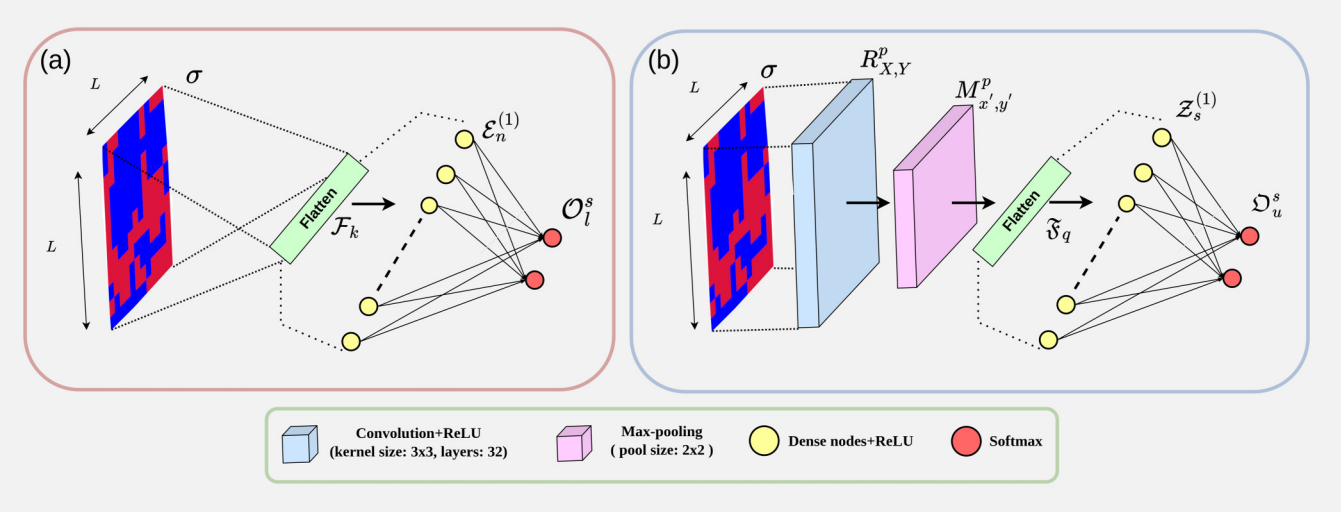

FIG. 1. Schematic diagrams of (a) the fully connected neural nework (FCNN) and (b) the convolutional neural network (CNN), which we train to classify snapshots of spin configurations that we obtain from our Monte Carlo simulations, both above and below the critical temperature T_c.

as inputs for training and testing the convolutional neural networks (CNNs) and fully connected neural networks (FCNNs) (see Sec. II D).

C. Finite-size scaling

Finite-size scaling helps us to extract universal critical exponents and scaling functions, in the vicinity of a critical or a tricritical point, from calculations on small systems, e.g., with N small in models (1), (4), and (6). Strictly speaking, to obtain the intensive bulk free energy f_B, we must take the thermodynamic limit $L \to \infty$, where the linear system size $L = Na$, with a the lattice spacing for the models we consider (we choose $a = 1$). In the vicinity of the critical point, f_L, the finite-size approximation to the f_B, assumes a scaling form [35,63], which is given below for model (1):

$$f_L(t,h) = L^{-d} f^*\left(tL^{y_t}, hL^{y_h}\right), \tag{8}$$

where d is the dimension, the reduced temperature $t \equiv \frac{(T-T_c)}{T_c}$, with T_c the critical temperature, $h \equiv H/(k_B T_c)$, and y_t and y_h are universal scaling exponents that are related to the conventional critical exponents via

$$y_t = \frac{1}{\nu} \quad \text{and} \quad y_h = \beta\delta/\nu \quad \text{whence} \tag{9}$$

$$M = -\left.\frac{\partial f_L}{\partial h}\right|_{h=0} = L^{-\beta/\nu} f_h^*(tL^{1/\nu}, 0), \quad \text{at} \quad h = 0, \tag{10}$$

$$M = -\left.\frac{\partial f_L}{\partial h}\right|_{t=0} = L^{-\beta/\nu} f_h^*(0, hL^{\beta\delta/\nu}), \quad \text{at} \quad t = 0. \tag{11}$$

From Eqs. (10) and (11), we can see that plots of (a) $ML^{\beta/\nu}$ vs $tL^{1/\nu}$, for $h = 0$, and (b) $ML^{\beta/\nu}$ vs $hL^{\beta\delta/\nu}$, for $t = 0$, fall on top of scaling curves for different values of L [these curves are different for cases (a) and (b)]. The best fits for these scaling curves lead to estimates for the critical exponents.

For the 2D Blume-Capel model (4), in the vicinity of tricritical point [40,42,43] at $H = 0$, $T = T_t^{\rm BC}$, and $\Delta = \Delta_t$, the finite-size scaling relation for the free energy is

$$f_L^{\rm BC}(t,g,h) = L^{-d} f^{*BC}\left(tL^{y_t}, gL^{y_g}, hL^{y_h}\right), \tag{12}$$

where $t = \frac{(T-T_t^{\rm BC})}{k_B T_t^{\rm BC}}$ and $g = \frac{(\Delta-\Delta_t)}{(k_B T_t^{\rm BC})} + at$ [40,43], where a is a nonuniversal constant, and g is the deviation from the tricritical point along the tangent to the coexistence curve, in the t-Δ plane.

Similarly, for the 2D metamagnet, the singular part of the free energy in the vicinity of the tricritical point, at $H_s = 0$, $T = T_t^M$, and $H = H_t^M$, has the form (cf. Refs. [45–47])

$$f_L^M(t,g',h_s) = L^{-d} f^{*M}\left(tL^{y_t}, g'L^{y_{g'}}, h_s L^{y_h}\right), \tag{13}$$

where $t = \frac{(T-T_t^M)}{k_B T_t^M}$, and $g' = \frac{(H-H_t^M)}{(k_B T_t^M)} + a't$ [47] is the tangent to the coexistence curve in the T-H plane, $h_s = H_s/(k_B T_t^M)$ is the staggered field, and a' is a nonuniversal coefficient.

D. Neural-network architectures

We employ fully connected neural networks (FCNNs) and convolutional neural networks (CNNs). In Fig. 1(a), we give a schematic diagram of the FCNN we use; the mathematical operations for our FCNN are given below:

$$\begin{aligned}
\text{Input:} \quad & \sigma_{X,Y}\ (L \times L), \quad \text{i.e.,}\ 1 \leqslant X, Y \leqslant L, \\
\text{Flatten:} \quad & \mathcal{F}_k = \sigma_{X,Y}, \quad k = (Y-1)L + X, \\
\text{Dense layer 1:} \quad & E_n^{(1)} = \sum_{k=1}^{L^2} W_{n,k}^{(1)} \mathcal{F}_k + b_n^{(1)}, \ 1 \leqslant n \leqslant 64, \\
\text{ReLU:} \quad & \mathcal{E}_n^{(1)} = \max\left(0, E_n^{(1)}\right), \ 1 \leqslant n \leqslant 64, \\
\text{Dense layer 2:} \quad & E_l^{(2)} = \sum_{m=1}^{64} W_{l,m}^{(2)} \mathcal{E}_m^{(1)} + b_l^{(2)}, \ 1 \leqslant l \leqslant 2, \\
\text{Softmax:} \quad & \mathcal{O}_l^s = \frac{\exp\left(E_l^{(2)}\right)}{\sum_{j=1}^{2} \exp\left(E_j^{(2)}\right)}, \ l = 1, 2.
\end{aligned} \tag{14}$$

Here the input $\sigma_{X,Y}$ is the spin at site (X, Y) [S_i for models (1) and (6) and $\mathcal{S}_i$ for model (4)]. We have a hidden layer with 64 dense nodes with weights $W^{(1)}_{n,k}$, and biases $b^{(1)}_n$, followed by the ReLU activation. In the output, we have two nodes with weights $W^{(2)}_{l,m}$ and biases $b^{(2)}_l$, followed by a softmax activation to obtain the normalized outputs $\mathcal{O}^s_l$.

In Fig. 1(b) we give a schematic diagram of the CNN we use; and the mathematical operations for our CNN are given below:

$$\begin{aligned}
\text{Input:}\quad & \sigma_{X,Y}\ (L\times L),\ \text{ i.e., } 0\leqslant X,Y<L,\\
\text{Padding:}\quad & \tilde{\sigma}_{X,Y},\ \text{ i.e., } 0\leqslant X,Y<L+2,\\
& \tilde{\sigma}_{X,Y}=\sigma_{X-1,Y-1},\ \text{if } 1\leqslant X,\ Y\leqslant L,\\
& \text{else } \tilde{\sigma}_{\tilde{X},\tilde{Y}}=0,\\
\text{Convolution:}\quad & C^{(p)}_{X,Y}=\sum_{i=0}^{2}\sum_{j=0}^{2}\tilde{\sigma}_{X+i,Y+j}F^{(p)}_{i,j}+b^{(p)},\\
& p\in[0,1,\ldots,31],\\
\text{ReLU:}\quad & R^{p}_{X,Y}=\max\bigl(0,C^{(p)}_{X,Y}\bigr),\\
\text{Max-pool:}\quad & M^{(p)}_{x',y'}=\max\bigl(R^{(p)}_{(X:X+1)(Y:Y+1)}\bigr),\\
& 2x'=X,\ 2y'=Y,\ \ 0\leqslant x',y'<\frac{L}{2},\\
\text{Flatten:}\quad & \mathfrak{F}_q=M^{(p)}_{x',y'},\ q=p\left(\frac{L}{2}\right)^2+y'\left(\frac{L}{2}\right)+x',\\
\text{Dense layer 1:}\quad & Z^{(1)}_s=\sum_{r=0}^{32(\frac{L}{2})^2-1}W^{(1)}_{s,r}\mathfrak{F}_r+b^{(1)}_s,\ 1\leqslant s\leqslant 64,\\
\text{ReLU:}\quad & \mathcal{Z}^{(1)}_s=\max\bigl(0,Z^{(1)}_s\bigr),\ 1\leqslant s\leqslant 64,\\
\text{Dense layer 2:}\quad & Z^{(2)}_t=\sum_{s=1}^{64}W^{(2)}_{t,s}\mathcal{Z}^{(1)}_s+b^{(2)}_t,\ 1\leqslant t\leqslant 2,\\
\text{Softmax:}\quad & \mathfrak{O}^s_u=\frac{\exp\left(Z^{(2)}_u\right)}{\sum_{t=1}^{2}\exp\left(Z^{(2)}_t\right)},\ u=1,2.
\end{aligned}\tag{15}$$

We have a convolutional layer, with 32, 3×3 filters $F^{(p)}_{i,j}$, and biases $b^{(p)}$, with ReLU activation (we apply the padding to retain the original spatial dimensions), followed by a max-pooling layer with 2×2 filters and a stride of 2, and a dense layer with 64 nodes, with weights $W^{(1)}_{s,r}$, and biases $b^{(1)}_s$. In the output, we have two nodes with weights $W^{(2)}_{t,s}$ and biases $b^{(2)}_t$, followed by a softmax activation to obtain the normalized outputs $\mathfrak{O}^s_t$.

With the spin configurations $\sigma_{X,Y}$ as the input, we train our neural networks to classify these configurations using binary cross-entropy, with a regularizer term as the loss function,

$$\mathcal{L}=\langle -P(\boldsymbol{\sigma})\log\hat{P}(\boldsymbol{\sigma})-[1-P(\boldsymbol{\sigma})]\log[1-\hat{P}(\boldsymbol{\sigma})]\rangle+\lambda||\mathcal{W}||,\tag{16}$$

where $P(\boldsymbol{\sigma})$ takes the values 1 (or 0) if the spin configuration $\boldsymbol{\sigma}$, from the training data, is below (or above) the critical temperature; and $\hat{P}(\boldsymbol{\sigma})\in[0,1]$ is the output or the prediction of the neural network, which is $\mathcal{O}^s_1$ for our FCNN [Eq. (14)] and $\mathfrak{O}^s_1$ for our CNN [Eq. (15)], $\langle\cdot\rangle$ denotes the average over the training data set, λ is the regularization strength, and

$$||\mathcal{W}||=\sum_{n=1}^{64}\sum_{k=1}^{L^2}W^{(1)2}_{n,k}+\sum_{l=1}^{2}\sum_{m=1}^{64}W^{(2)2}_{l,m},\tag{17}$$

for the FCNN, and

$$||\mathcal{W}||=\sum_{p=0}^{p=31}\sum_{j=0}^{2}\sum_{i=0}^{2}F^{(p)2}_{i,j}+\sum_{s=1}^{64}\sum_{r=0}^{32(\frac{L}{2})^2-1}W^{(1)2}_{s,r}+\sum_{t=1}^{2}\sum_{s=1}^{64}W^{(2)2}_{t,s},\tag{18}$$

for the CNN. We implement our neural networks using Tensorflow [64].

III. RESULTS

We present our results for the scaling forms for neural network outputs in Sec. III A, and results for the ferromagnetic Ising model (1), the Blume-Capel model (4), and the Ising-metamagnet (6) in Secs. III B–III D respectively. We discuss universal scaling functions and scale factors in Sec. III E and FSS in the vicinity of the first-order transition for the Ising model (1) in Sec. III F.

A. Scaling forms for neural network outputs

For a system with linear size L, we define

$$\mathcal{P}_L(t,h)=\langle\hat{P}_L(\boldsymbol{\sigma})\rangle_{\{\boldsymbol{\sigma}\}},\tag{19}$$

where $\langle\cdot\rangle_{\{\boldsymbol{\sigma}\}}$ denotes the average of $\hat{P}_L(\boldsymbol{\sigma})$ [Eq. 16] over the spin configurations $\boldsymbol{\sigma}$ from the test data set obtained at given values of t and h. $\mathcal{P}_L(t,h)$ can be expressed as (cf. Ref. [16] for $h=0$)

$$\mathcal{P}_L(t,h)=\int d\boldsymbol{\sigma}\,\mathfrak{P}_L(\boldsymbol{\sigma},t,h)\hat{P}_L(\boldsymbol{\sigma}),\tag{20}$$

where the $\mathfrak{P}_L(\boldsymbol{\sigma},t,h)$ is the probability distribution function (PDF) of $\boldsymbol{\sigma}$. We now follow Refs. [16,65] to go from $\boldsymbol{\sigma}$ to $|M|$ to obtain

$$\mathcal{P}_L(t,h)=\int d|M|\,[\mathfrak{P}_L(|M|,t,h)\hat{P}_L(|M|)]\tag{21}$$

for a system of linear size L, this has the FSS form [66]

$$\mathfrak{P}_L(|M|,t,h)\equiv L^{\frac{\beta}{\nu}}\mathfrak{P}^*\bigl(|M|L^{\frac{\beta}{\nu}},tL^{y_t},hL^{y_h}\bigr).\tag{22}$$

A heuristic explanation of the expression (22) follows (see Refs. [66–68] for details). Let $\tilde{\mathfrak{P}}_L(M,t)$ be the PDF of M at $h=0$ for a system of linear size L, then the corresponding FSS form can be written as [67,68]

$$\tilde{\mathfrak{P}}_L(M,t)=\frac{1}{\varsigma_M}\tilde{\mathfrak{P}}^*\left(\frac{M}{\varsigma_M},\frac{L}{\xi}\right)\tag{23}$$

in the vicinity of the critical temperature T_c, $\xi\sim|t|^{-\frac{1}{y_t}}$, but ξ is bounded above by L, in a finite-size system, so $\frac{L}{\xi}$ becomes relevant for FSS [35,36,67,68]. In Eq. (23),

$\varsigma_M = \sqrt{\langle M^2\rangle - \langle M\rangle^2} = \sqrt{\langle M^2\rangle}$, whence we obtain $\varsigma_M \sim L^{\frac{-\beta}{\nu}}$; finally we have the following FSS form:

$$\bar{\mathfrak{P}}_L(M,t) = L^{\frac{\beta}{\nu}}\bar{\mathfrak{P}}^*\left(ML^{\frac{\beta}{\nu}}, tL^{y_t}\right); \tag{24}$$

similarly at $t = 0$, from $\xi \sim |h|^{-\frac{1}{y_h}}$, $|M| \sim |h|^{\frac{1}{\delta}}$, we have

$$\bar{\mathfrak{P}}_L(M,h) = L^{\frac{\beta}{\nu}}\bar{\mathfrak{P}}^*\left(ML^{\frac{\beta}{\nu}}, hL^{y_h}\right). \tag{25}$$

We now employ the following FSS form [see Ref. [16] and Fig. 7(a) in Appendix D]

$$\hat{P}_L(|M|) \equiv \hat{P}^*\left(|M|L^{\frac{\beta}{\nu}}\right) \tag{26}$$

and use Eqs. (20), (22), and (26) to obtain

$$\begin{aligned}\mathcal{P}_L(t,h) &= \int d|M|L^{\frac{\beta}{\nu}}\left[\mathfrak{P}^*\left(|M|L^{\frac{\beta}{\nu}}, tL^{y_t}, hL^{y_h}\right)\right.\\ &\quad\left.\times\,\hat{P}^*\left(|M|L^{\frac{\beta}{\nu}}\right)\right]\\ &\equiv \mathcal{P}^*\left(tL^{y_t}, hL^{y_h}\right).\end{aligned} \tag{27}$$

Henceforth, we denote by $\mathcal{P}^*_L(tL^{y_t}, hL^{y_h})$ our numerical approximation of $\mathcal{P}^*(tL^{y_t}, hL^{y_h})$ for systems with linear size L.

Note that the FSS forms (22) and (26) show that tranfer-learning works for our NNs (Secs. III C and III D) because the output of this NN exhibits FSS irrespective of the details of the form of the function $\mathfrak{P}_L(|M|, t, h)$ (see also Ref. [16]).

In Secs. III B–III D, we show how the FSS expression (27) holds for the Ising, Blume-Capel, and Ising-metamagnet models that we consider.

B. Two-dimensional Ising model

We give the specific architectures of the CNN and FCNN that we train to classify the spin configurations of the 2D Ising model (Sec. II A 1), above and below the critical temperature $k_BT_c/J = 2/\ln(1+\sqrt{2})$, in Sec. II [Eqs. (14) and (15)], and the details of training and testing are given in Appendixes A and B.

Figure 2 contains the results from our CNN (15). In Fig. 2(a), we plot $\mathcal{P}_L(t,0)$ vs t, for $L = 12$, 16, 20, and 30; in the inset we give the finite-size-scaling (FSS) plot of $P^*_L(tL^{y_t}, 0)$ vs tL^{y_t} from which we obtain the best-fit exponent $y_t = 1.03(2)$ that is close to the exact value 1 (cf. Refs. [12,14–18] for similar studies at $h = 0$).

We now show how to generalize our discussion above to obtain the magnetic exponent y_h. In Fig. 2(b) we plot $\mathcal{P}_L(0,h)$ vs h; and in the inset we give the FSS plot of $\mathcal{P}^*_L(0, hL^{y_h})$ vs hL^{y_h}; this yields the best-fit magnetic exponent $y_h = 1.84(1)$, which is close to the exact value 1.875. In Appendix B we give the estimates for y_t and y_h, from NNs saved across different training epochs, whence we get error estimates for y_t [$\simeq O(0.01)$] and for y_h [$\simeq O(0.1)$]. Here we use NNs, from which the estimates for y_t and y_h are close to the known values. The plot in Fig. 2(c) shows $\mathcal{P}_L(t,h)$ as a function of t and h for $L = 12$ and 20; in the inset we give the FSS plot of $\mathcal{P}^*_L(tL^{y_t}, hL^{y_h})$ as a function of tL^{y_t} and hL^{y_h}, for $L = 12$, 16, 20, 30, whence we obtain $y_t = 1.07(5)$, and $y_h = 1.83(12)$; in Figs. 2(d) and 2(e), we show different views of the surface onto which the points collapse. In Table I we give the estimates for y_t and y_h that we obtain from scaling-collapse fits for both the CNN (15) and FCNN (14).

TABLE I. Our estimates for the thermal (y_t) and magnetic (y_h) exponents in the vicinity of the second-order transition for the Ising model (1).

NN	y_t [exact:1]	y_h [exact:1.875]
CNN	1.03(2)	1.84(1)
FCNN	1.06(2)	1.85(1)

C. Two-dimensional Blume-Capel model

We train our CNN (15) and FCNN (14) to classify the spin configurations of the 2D Blume-Capel model (4) above and below its critical temperature $T_c \simeq 1.69$ [69], with $\Delta = 0$ and $h = 0$ (details of the training and testing are given in Appendixes A and B).

Figure 3 contains the results from our CNN (15). In Fig. 3(a), we plot $\mathcal{P}_L(t,0)$ vs t, for $L = 12$, 16, 20, and 30 for $\Delta = 0$ and $h = 0$; in the inset we give the finite-size-scaling (FSS) plot of $\mathcal{P}^*_L(tL^{y_t}, 0)$ vs tL^{y_t} from which we obtain the best-fit exponent $y_t = 1.05(2)$. In Fig. 3(b), we plot $\mathcal{P}_L(0,h)$ vs h; and in the inset we give the FSS plot of $\mathcal{P}^*_L(0, hL^{y_h})$ vs hL^{y_h}; this yields the best-fit magnetic exponent $y_h = 1.83(1)$. In Table II we give the estimates for y_t and y_h that we obtain from scaling-collapse fits for both the CNN (15) and FCNN (14).

We now test our CNN (15), trained at $\Delta = 0$ and $h = 0$, for FSS along the second-order transition line and for the 2D Blume-Capel model (4); this transition is in the universality class of the 2D Ising model [40,42,43]. In particular, in Fig. 3(c), we fix $\Delta = 1$, and change T, in the vicinity of $\mathcal{P}_L(t, h = 0) \simeq 0.5$. Now FSS best-fit plots [insets in Figs. 3(c) and 3(d)] yield $y_t = 1.03(1)$ and $y_h = 1.78(1)$. We summarize these results for the CNN (15) and FCNN (14) in Table II.

The Blume-Capel model (4) exhibits a tricritical point at $\Delta \simeq 1.966$ and $T \simeq 0.608$ (see, e.g., Ref. [42]); the exponents at this tricritical point are different from those that characterize the 2D Ising universality class [40–43]. Here we train our CNN (15) and FCNN (14) to classify the spin configurations below and above $T_c = 0.608$ [42], with $\Delta = 1.966$. In Fig. 3(e) we plot $\mathcal{P}_L(t,0)$ vs t, for $L = 10$, 12, 16; in the inset we give the FSS plot of $\mathcal{P}^*_L(tL^{y_t}, 0)$ vs tL^{y_t}, whence we obtain the best-fit exponent $y_t = 1.60(2)$, which is close to $y_t \simeq 1.8$

TABLE II. Estimates for the thermal (y_t) and magnetic (y_h) exponents in the vicinity of the second-order transition for $\Delta = 0$ and $\Delta = 1$ in the Blume-Capel model (4).

NN	y_t [exact:1]	y_h [exact:1.875]
CNN: $\Delta = 0$	1.05(2)	1.83(1)
FCNN: $\Delta = 0$	1.07(2)	1.81(1)
CNN: $\Delta = 1$	1.03(1)	1.78(1)
FCNN: $\Delta = 1$	1.09(1)	2.0(1)

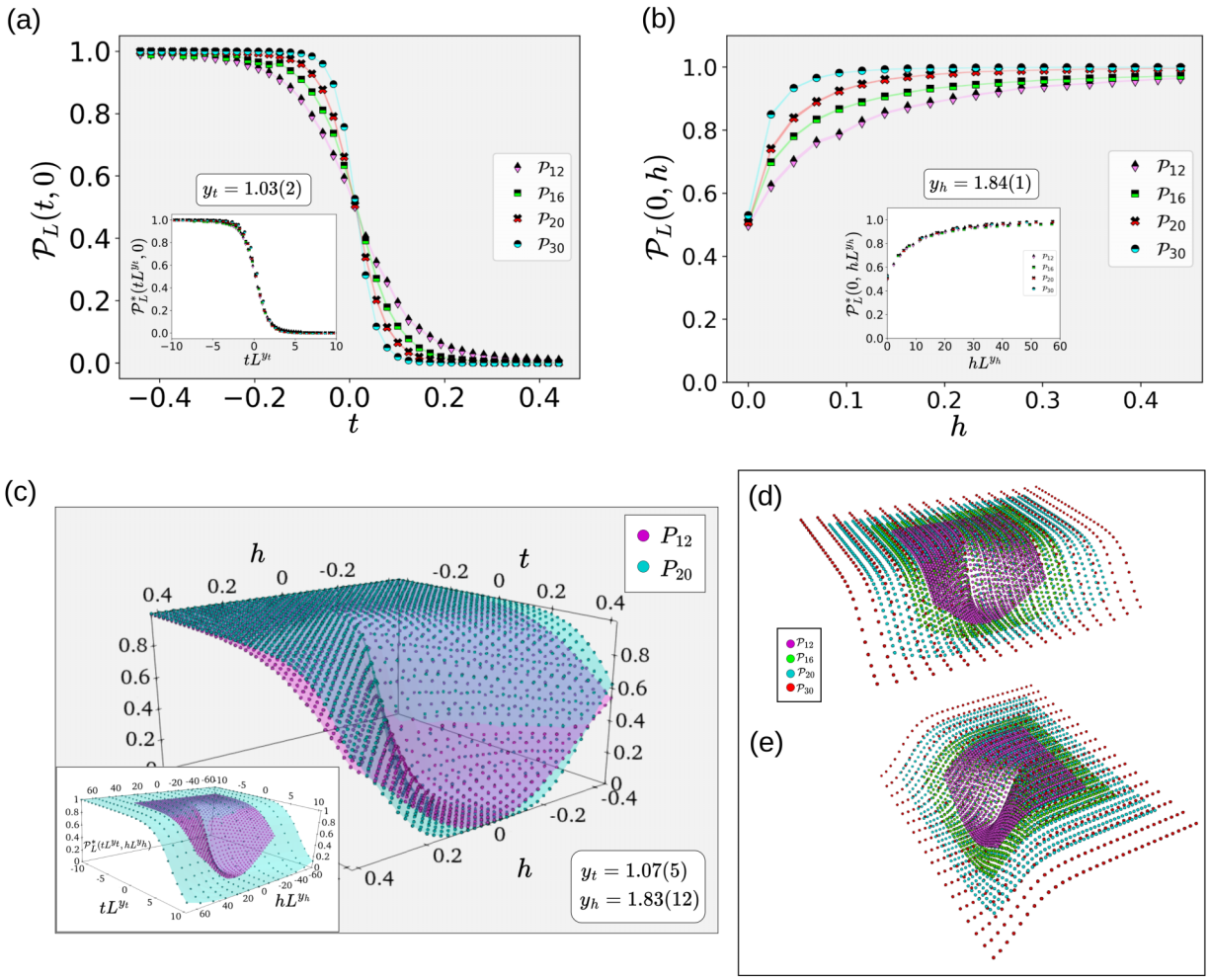

FIG. 2. Plots for the 2D Ising model (1): (a) $\mathcal{P}_L(t,0)$ vs t, for $L = 12$, 16, 20, and 30; in the inset we give the finite-size-scaling (FSS) plot of $\mathcal{P}^*_L(tL^{y_t},0)$ vs tL^{y_t} from which we obtain the best-fit exponent $y_t = 1.03(2)$ that is close to the exact value 1. (b) $\mathcal{P}_L(0,h)$ vs h; and in the inset we give the FSS plot of $P^*_L(0,hL^{y_h})$ vs hL^{y_h}; this yields the best-fit magnetic exponent $y_h = 1.84(1)$, which is close to the exact value 1.875. (c) $\mathcal{P}_L(t,h)$ as a function of t and h for $L = 12$ and 20; in the inset we give the FSS plot of $P^*_L(tL^{y_t}, hL^{y_h})$ as a function of tL^{y_t} and hL^{y_h}, for $L = 12$, 16, 20, 30, whence we obtain $y_t = 1.07(5)$, and $y_h = 1.83(12)$. In panels (d) and (e), we show different views of the surface onto which the points collapse.

[40–43]. In Fig. 3(f) we plot $\mathcal{P}_L(0,h)$ vs h; and in the inset we give the FSS plot of $\mathcal{P}^*_L(0,hL^{y_h})$ vs hL^{y_h}; this yields the best-fit magnetic exponent $y_h = 1.94(4)$, which is close to $y_h \simeq 1.92$ [40–43]. In Table III we give the estimates for y_t and y_h that we obtain from scaling-collapse fits for both the CNN (15) and FCNN (14).

D. Two-dimensional Ising-metamagnet model

We turn now to the 2D Ising-metamagnet model (6). We train our CNN (15) and FCNN (14) to classify the staggered spin configurations of the 2D Ising-metamagnet model (6) above and below its critical temperature $T_c \simeq 5.263$ [45], with $h = 0$ and $h_s = 0$ (details of the training and testing are given in Appendixes A and B).

TABLE III. Estimates for the thermal (y_t) and magnetic (y_h) exponents in the vicinity of the tricritical point for the Blume-Capel model (4).

NN	y_t [exact:1.8]	y_h [exact:1.92]
CNN: $\Delta = 1.966$	1.60(2)	1.94(4)
FCNN: $\Delta = 1.966$	1.87(3)	1.94(3)

Figure 4 contains the results from our CNN (15). In Fig. 4(a) we plot $\mathcal{P}_L(t,0)$ vs t, for $L = 12$, 16, 20, and 30 for $h = 0$ and $h_s = 0$; in the inset we give the FSS plot of $P^*_L(tL^{y_t},0)$ vs tL^{y_t} from which we obtain the best-fit exponent $y_t = 1.06(1)$. In Fig. 4(b) we plot $\mathcal{P}_L(0,h_s)$ vs h_s, and in the inset we give the FSS plot of $\mathcal{P}^*_L(0,h_sL^{y_h})$ vs $h_sL^{y_h}$; this yields the best-fit magnetic exponent $y_h = 1.90(1)$. In Table IV we give the estimates for y_t and y_h that we obtain from scaling-collapse fits for both the CNN (15) and FCNN (14).

We now test our CNN (15), trained at $h = 0$ and $h_s = 0$, for FSS along the second-order transition line; for the 2D Ising-metamagnet model (6), this transition is in the universality class of the 2D Ising model [45]. In particular, in Fig. 4(c), we fix $H = 1$, and change T, in the vicinity of $\mathcal{P}_L(t,h=0) \simeq 0.5$. Now the FSS best-fit plots [insets in Figs. 4(c) and 4(d)]

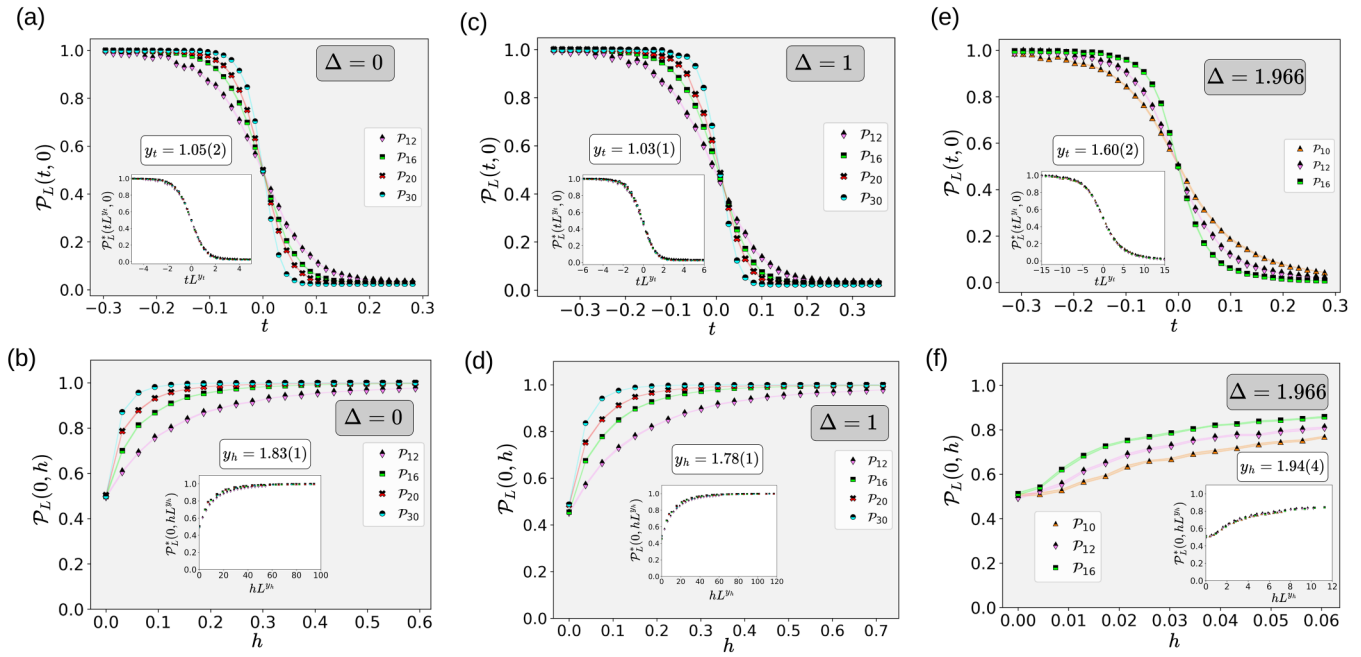

FIG. 3. Plots for the 2D Blume-Capel model (4): (a) $\mathcal{P}_L(t,0)$ vs t, for $L = 12,\ 16,\ 20$, and 30 for $\Delta = 0$ and $h = 0$; in the inset we give the finite-size-scaling (FSS) plot of $P_L^*(tL^{y_t},0)$ vs tL^{y_t} from which we obtain the best-fit exponent $y_t = 1.05(2)$. (b) $\mathcal{P}_L(0,h)$ vs h; and in the inset we give the FSS plot of $P_L^*(0,hL^{y_h})$ vs hL^{y_h}; this yields the best-fit magnetic exponent $y_h = 1.83(1)$. Panels (c) and (d) are the counterparts of (a) and (b), but with $\Delta = 1$. (e) and (f) are the counterparts of (a) and (b), but in the vicinity of tricritical point at $\Delta = 1.966$.

yield $y_t = 1.07(2)$ and $y_{h_s} = 1.85(2)$. We summarize these results for the CNN (15) and FCNN (14) in Table IV.

The 2D Ising-metamagnet model (6) exhibits a tricritical point at $h_s = 0$, $H \simeq 3.927$, and $T \simeq 2.41$ (see, e.g., [45]); the exponents at this tricritical point [45–47] are different from those that characterize the 2D Ising universality class. Here we train our CNN (15) and FCNN (14) to classify the spin configurations below and above $T_c = 2.41$ [45], with $H_s = 0$

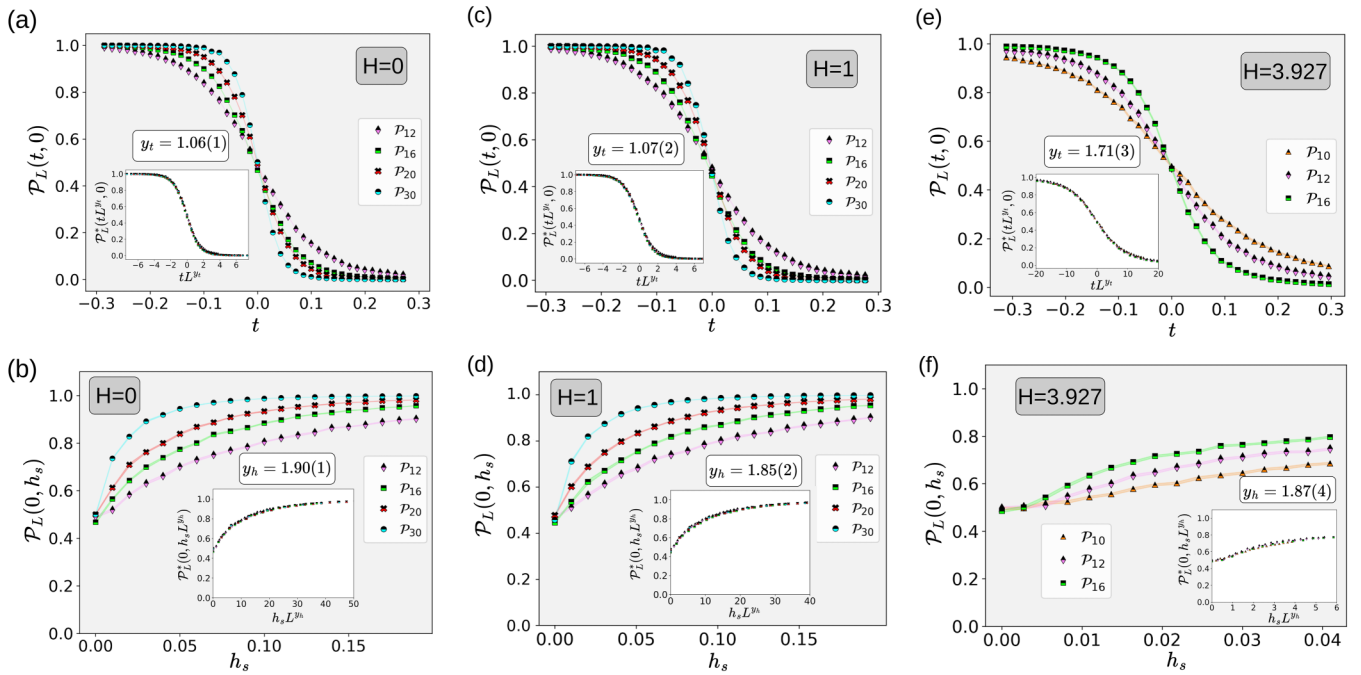

FIG. 4. Plots for the 2D Ising-metamagnet model (6): (a) $\mathcal{P}_L(t,0)$ vs t, for $L = 12,\ 16,\ 20$, and 30 for $h = 0$ and $h_s = 0$; in the inset we give the FSS plot of $P_L^*(tL^{y_t},0)$ vs tL^{y_t} from which we obtain the best-fit exponent $y_t = 1.06(1)$. (b) $\mathcal{P}_L(0,h_s)$ vs h_s; and in the inset we give the FSS plot of $P_L^*(0,h_sL^{y_h})$ vs $h_sL^{y_h}$; this yields the best-fit magnetic exponent $y_h = 1.90(1)$. Panels (c) and (d) are the counterparts of (a) and (b), but with $H = 1$. Panels (e) and (f) are the counterparts of (a) and (b), but in the vicinity of the tricritical point at $H = 3.927$.

TABLE IV. Estimates for the thermal (y_t) and magnetic (y_h) exponents in the vicinity of the second-order transitions for $H = 0$ and $H = 1$ for the Ising-metamagnet model (6).

NN	y_t [exact:1]	y_h [exact:1.875]
CNN: $H = 0$	1.06(1)	1.90(1)
FCNN: $H = 0$	1.09(1)	1.85(2)
CNN: $H = 1$	1.07(2)	1.85(2)
FCNN: $H = 1$	1.092(1)	1.93(2)

and $H = 3.927$. In Fig. 4(e) we plot $\mathcal{P}_L(t, 0)$ vs t, for $L = 10,\ 12,\ 16$; in the inset we give the FSS plot of $\mathcal{P}_L^*(tL^{y_t}, 0)$ vs tL^{y_t}, whence we obtain the best-fit exponent $y_t = 1.71(3)$, which is close to $y_t \simeq 1.8$ [45]. In Fig. 4(f), we plot $\mathcal{P}_L(0, h_s)$ vs h_s; and in the inset we give the FSS plot of $P_L^*(0, h_s L^{y_h})$ vs $h_s L^{y_h}$; this yields the best-fit magnetic exponent $y_h = 1.87(4)$, which is close to $y_h \simeq 1.92$. In Table V we give the estimates for y_t and y_h that we obtain from scaling-collapse fits for both the CNN (15) and FCNN (14).

E. Universal scaling functions and scale factors

We now investigate two-scale factor universality [38,39,70] by comparing the scaling functions $\mathcal{P}_L^{*\alpha}(tL^{y_t}, hL^{y_h})$ obtained from our neural networks, where the label α is I, BC, and M for the 2D Ising (1), 2D Blume-Capel (4), and 2D Ising-metamagnet (6) models, respectively. The plots in Fig. 5(a) of $\mathcal{P}_L^I((T - T_c^I), 0)$ vs $(T - T_c^I)$, $\mathcal{P}_L^{\rm BC}((T - T_c^{\rm BC}), 0)$ vs $(T - T_c^{\rm BC})$, and $\mathcal{P}_L^M((T - T_c^M), 0)$ vs $(T - T_c^M)$, for $L = 12,\ 16,\ 20$, and 30 show that these curves are different for these three models. We now show that, if we scale tL^{y_t} by model-dependent scale factors, then all these curves collapse onto one universal curve. In particular, if we use the scaled variables

$$f_t^{\rm BC} tL^{y_t} \quad \text{and} \quad f_t^M tL^{y_t} \tag{28}$$

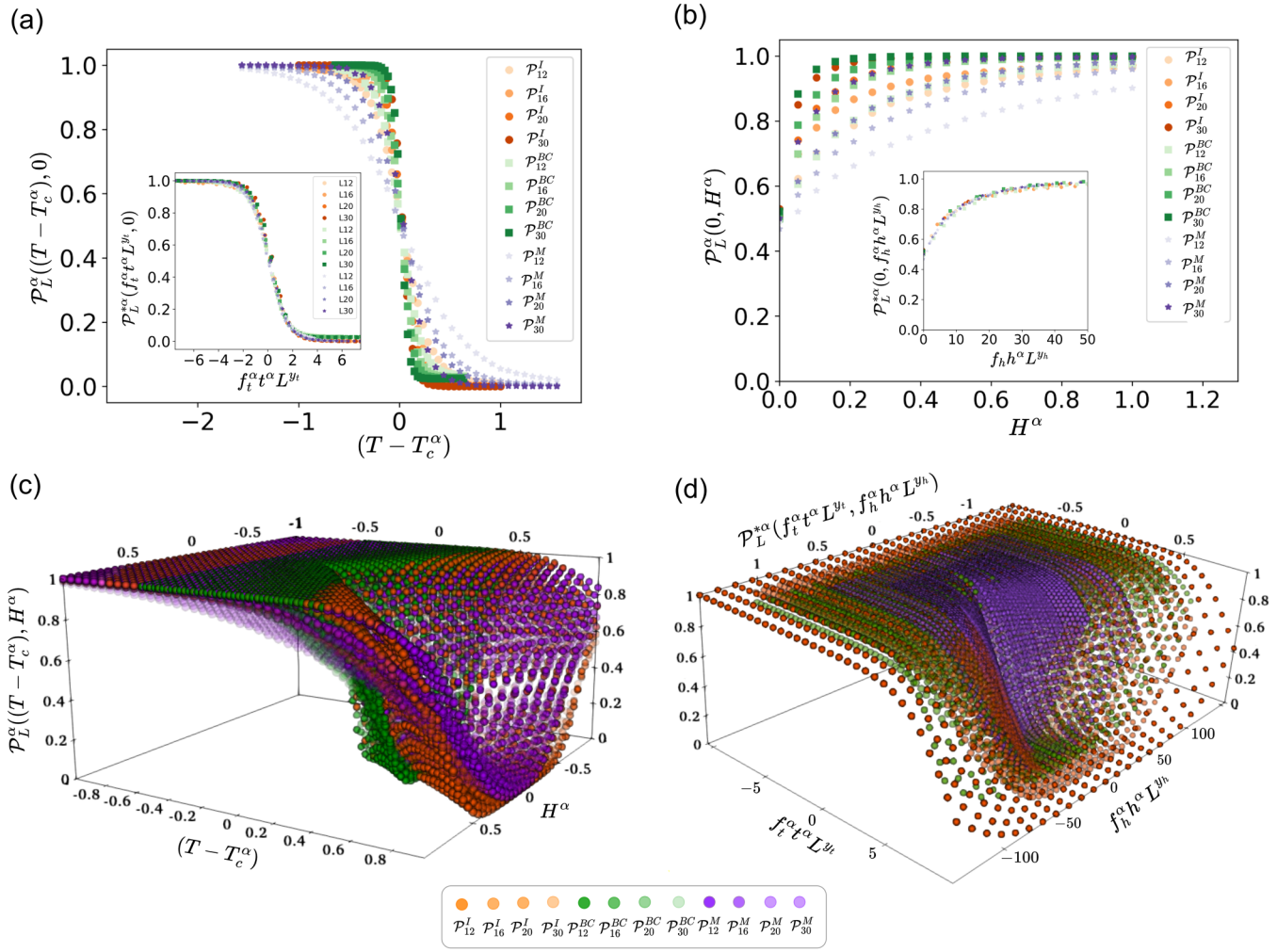

FIG. 5. (a) Plots of $P_L^I((T - T_c^I), 0)$ vs $(T - T_c^I)$, $P_L^{\rm BC}((T - T_c^{\rm BC}), 0)$ vs $(T - T_c^{\rm BC})$, and $P_L^M((T - T_c^M), 0)$ vs $(T - T_c^M)$, for $L = 12,\ 16,\ 20$, and 30; these curves are different for the three models we consider. (b) $P_L^I(0, H)$ vs H, $P_L^{\rm BC}(0, H)$ vs H, and $P_L^M(0, H_s)$ vs H_s, for $L = 12,\ 16,\ 20$, and 30. (c) $P_L^\alpha((T - T_c^\alpha), H)$ and $P_L^\alpha((T - T_c^\alpha), H_s)$ vs $T - T_c^\alpha$ and H or H_s, for all three models, $\alpha = I$, BC, M. (d) These surfaces collapse onto one universal scaling surface, if we use the scaled variables $f_t^{\rm BC} tL^{y_t}$ and $f_h^{\rm BC} hL^{y_h}$, for the Blume-Capel model (4), and $f_t^M tL^{y_t}$ and $f_h^M h_s L^{y_h}$, for the Ising-metamagnet model (6).

TABLE V. Estimates for the thermal (y_t) and magnetic (y_h) exponents in the vicinity of the tricritical point for the Ising-metamagnet model (6).

NN	y_t [exact:1.80]	y_h [exact:1.92]
CNN	1.71(3)	1.87(4)
FCNN	1.69(2)	2.03(5)

for the Blume-Capel and the Ising-metamagnet models, respectively, then we get the universal curve shown in the inset of Fig. 5(a) and the best fit yields the model-dependent scale factors $f_t^{\rm BC} \simeq 1.27(6)$ and $f_t^M \simeq 1.1(4)$. Similarly, the plots in Fig. 5(b) of $P_L^I(0,H)$ vs H, $P_L^{\rm BC}(0,H)$ vs H, and $P_L^M(0,H_s)$ vs H_s, for $L = 12$, 16, 20, and 30, demonstrate that these curves are different for these three models; but, if we use the scaled variables

$$f_h^{\rm BC} h L^{y_h} \quad \text{and} \quad f_h^M h_s L^{y_h} \tag{29}$$

for the Blume-Capel and the Ising-metamagnet models, respectively, then we get the universal curve shown in the inset of Fig. 5(b) and the best fit yields the model-dependent scale factors $f_h^{\rm BC} \simeq 0.95(10)$ and $f_h^M \simeq 1.02(5)$. The plots in the insets of Figs. 5(a) and 5(b) are a result of two-scale factor universality [38,39,70] in our NN study. In Fig. 5(c) we give plots of $P_L^\alpha((T - T_c^\alpha), H$ or$H_s)$ vs $T - T_c^\alpha$ and H or H_s, for all three models, i.e., α is I, BC, and M; these surfaces collapse onto one universal scaling surface [Fig. 5(d)], if we use the scaled variables $f_t^{\rm BC} t L^{y_t}$ and $f_h^{\rm BC} h L^{y_h}$, for the Blume-Capel model, and $f_t^M t L^{y_t}$ and $f_h^M h_s L^{y_h}$, for the Ising-metamagnet model.

If N_L^α is the number of spin configurations that are correctly classified out of a total N configurations for a system of linear size L, say, below T_c^α and in the temperature range $[T_0, T_c^\alpha]$, then we can show (see Appendix D)

$$\frac{\left(N - N_L^\alpha\right)}{N} \propto \frac{1}{f_t^\alpha \left|t_0^\alpha\right| L^{y_t}}, \tag{30}$$

where $t_0^\alpha \equiv \frac{T_0 - T_c^\alpha}{T_c}$. The plot of $\frac{(N - N_L^\alpha)}{N}$ vs $f^\alpha |t_0^\alpha| L^{y_t}$ in Fig. 7(b) in Appendix D verifies this relation.

F. First-order phase boundary in the 2D Ising model

To investigate the FSS of the CNNs in the vicinity of the first-order transition, we follow the theoretical ideas given in Refs. [36,37,71–73]. In a renormalization-group (RG) treatment, the jump of the order parameter at a first-order transition is governed by the flows of the RG recursion relations in the vicinity of a discontinuity fixed point (see, e.g., Ref. [73]); such a fixed point occurs in the subspace of even-spin couplings (i.e., h and all odd-spin couplings vanish) and lies out at infinity (i.e., at $J/(k_B T)$ and many other even-spin couplings $= \infty$) [74]; the dominant eigenvalues at this discontinuity fixed point are such that $y_h = d$ and $y_t = (d - 1)$, where d is the spatial dimension [71–73]. Therefore, in the vicinity of the first-order transition [37] in the Ising model (1) we have the following: for $\frac{H M_{sp} L^d}{k_b T} \gg 1$, where M_{sp} is the infinite-system magnetization at $h \to 0^+$ [37], the magnetization for the system of linear size L is M_L, which $\to M_0$, the equilibrium magnetization value for the infinite L and h [see Fig. 6(a)]. If $\frac{H M_{sp} L^d}{k_b T} \ll 1$, we have $M_L \propto h L^d$ [37].

Transfer learning allows us to use the CNNs, which we have trained in the vicinity of the critical point of the Ising model (1), to obtain plots of $P_L(T = 2.1, h)$, as we change h from positive to negative values to cross the first-order boundary at $T = 2.1 < T_c$ [Fig. 6(b)]; we show plots for $L = 5$, 6, and 9. From these plots we see that $\mathcal{P}_L$ tends to $\mathcal{P}_L(|M_0|)$, for large values of h; and $\mathcal{P}_L[|M|_L(h = 0)]$ at $h = 0$ depends on the value of L. In Fig. 6(c) we present plots of $[\mathcal{P}_L(T = 2.1, h) - \mathcal{P}_L(T = 2.1, h = 0)]$ vs h; in the inset we show that these curves collapse onto one curve if we plot $[\mathcal{P}_L^*(T = 2.1, hL^d) - \mathcal{P}_L(T = 2.1, h = 0)]$ vs hL^d, with $d = 2$.

IV. DISCUSSION AND CONCLUSIONS

Earlier studies [12,14,16–18,20,23] of neural-network-aided methods for the determination of critical exponents have considered only simple critical points, such as the one in the Ising model (1) at $h = 0$. Our work goes well beyond these earlier investigations by developing a full framework for the scaling forms of NN outputs in the vicinities of critical and tricritical points and obtaining the following: (a) both the thermal and magnetic exponents, y_t and y_h, not only at simple Ising-type critical points, but also at tricritical points in the Blume-Capel (4) and the Ising-metamagnet (6) models; (b) the full scaling form for the NN outputs $\mathcal{P}_L$ given in Eq. (27); (c) the nonuniversal scale factors that are required for two-scale-factor universality; and (d) the CNN manifesta-

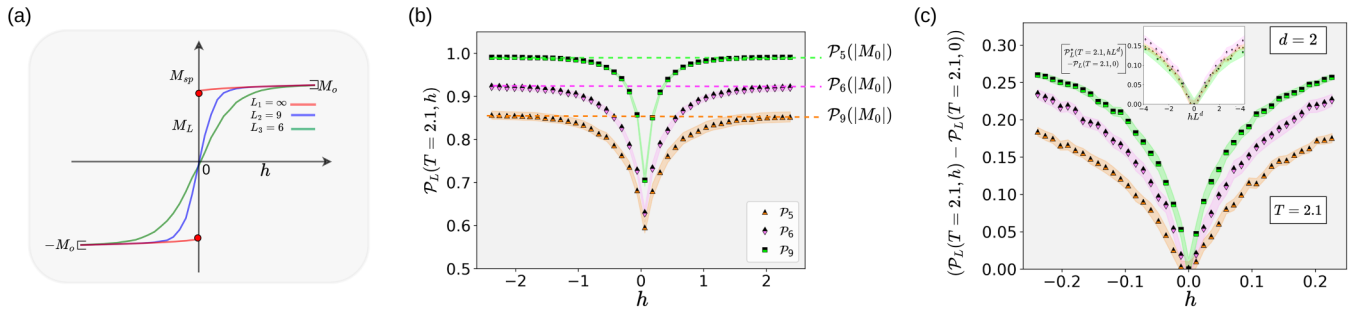

FIG. 6. (a) Plots of M_L vs h for the Ising first-order transition in the vicinity of $T_c = 2.1$. (b) $\mathcal{P}_L(T = 2.1, h)$ vs h (c) $[\mathcal{P}_L(T = 2.1, h) - \mathcal{P}_L(T = 2.1, h = 0)]$ vs h, and in the inset is the collapse plot of $[\mathcal{P}_L^*(T = 2.1, hL^d) - \mathcal{P}_L(T = 2.1, h = 0)]$ vs hL^d.

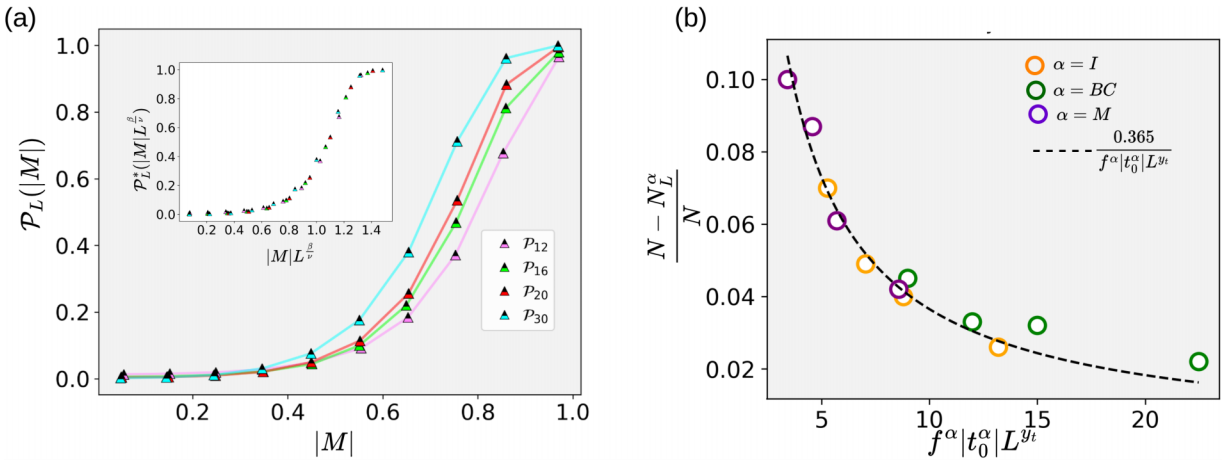

FIG. 7. (a) Plot of $\mathcal{P}_L(|M|)$ vs $|M|$, and in the inset, a plot of $\mathcal{P}_L^*(|M|L^{\frac{\beta}{\nu}})$ vs $|M|L^{\frac{\beta}{\nu}}$, for the Ising model (1). (b) Plot of $\frac{N-N_L^\alpha}{N}$ vs $f^\alpha|t_0^\alpha|L^{y_t}$ [see Eq. (D8)] for the Ising (I) (1), Blume-Capel (BC) (4), and Ising-metamagnet (M) (6) models.

tion of finite-size scaling at the Ising-model first-order phase boundary.

From the machine-learning point of view, our study provides the following interesting applications of transfer learning: We first train our CNNs and FCNNs in the vicinities of the zero-field critical points, at $h = 0$ of the Ising model (1), at $\Delta = 0$, $h = 0$ in the Blume-Capel model (4), and at $h_s = 0$, $h = 0$ for the Ising-metamagnet (6); and then we use these trained NNs to uncover scaling and critical exponents at critical points that occur at $h \neq 0$ for the Ising model (1), $h \neq 0$ and $\Delta \neq 0$ for the Blume-Capel (4) model, and $h \neq 0$, $h_s \neq 0$ for the Ising-metamagnet model (6). Our studies are valuable because they help us to go beyond the mere classification of phase by understanding the behavior and characteristics of the NN outputs $\mathcal{P}_L(t, h)$ and then extract, from these characteristics, other useful quantities, e.g., the scaling form for the number of spin configurations that are correctly classified [see Eq. (30)]. Our NN method has an additional computational overhead because of training, compared to traditional FSS methods that work directly with the order parameter M obtained from MC simulations. However, once the NN has been trained at, say, the Ising-model critical point, it can be used for other critical points in this university class because of transfer learning. Furthermore, our NN method provides $\hat{P}_L(\boldsymbol{\sigma})$.

The methods that we have developed can be extended, via transfer learning, to study phase transitions in other models. For example, the well-known Ising-lattice-gas mapping relates the Ising spins S_i to lattice-gas variables $n_i = [S_i + 1]/2$; given that $S_i = \pm 1$, we have $n_i = 0$ or 1 and the up-spin (down-spin) of the Ising model maps onto the high-density (low-density) phase of the lattice gas; in general, the liquid-gas critical point in a d-dimensional continuum fluid should be in the universality class of the d-dimensional Ising model. Therefore, it is interesting to use our trained NNs first to study Ising-model criticality in $d = 3$ and then employ transfer learning to examine the liquid-gas critical point in a continuum fluid described, e.g., the Lennard-Jones potential [8,75] as we will show in future work.

Although machine-learning methods, e.g., those that we have employed here, have, so far, not yielded values of critical exponents that are as accurate as those obtained by high-resolution MC, conformal-bootstrap, and series-expansions methods (e.g., see Refs. [48–57] for the three-dimensional Ising model). We expect that, as these ML methods are refined, by building on the framework that we have outlined here, the ML-based determination of critical exponents will move apace and achieve accuracies comparable to those attained by the other methods mentioned above. In future work, we hope to use the Wolff [76] or Swendsen-Wang [77] algorithms along with our NN study to overcome problems posed by critical slowing down, sampling bias, and large time correlations. It would also be interesting to explore if the estimates for the tricritical exponents (see the estimates of y_t in Tables III and V), improve with training data obtained from such sampling algorithms. Another interesting direction for future work is to employ our NN method in conjunction with spin configurations sampled from generative models, such as variational autoregressive networks [78] and restricted Boltzmann machines [79].

ACKNOWLEDGMENTS

We thank M. Barma for discussions, the Anusandhan National Research Foundation(ANRF), the Science and Engineering Research Board (SERB), and the National Supercomputing Mission (NSM), India, for support, and the Supercomputer Education and Research Centre (IISc) for computational resources.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

APPENDICES

In Appendix A we give the details of the training datasets and neural-network training. This is followed by Appendix B, in which we give the details of the datasets and neural networks that we use for testing and error estimates. In Appendix C, we give the details of our explorations with neural networks into regularization strength and larger system sizes. Finally, in Appendix D we give a detailed derivation of Eq. (30).

APPENDIX A: DATASETS AND NEURAL NETWORK TRAINING

For the 2D Ising model (1) in the vicinity of second-order transition, we carry out Monte Carlo (MC) simulations, for systems with linear sizes $L = 12,\ 16,\ 20,$ and 30, at 100 different values of the temperature T, spaced at intervals of 0.01 in the range $[T_c - 0.5, T_c + 0.5]$, where $T_c \simeq 2.269$ is the critical temperature. For each value of T, we discard the first 10^6 Monte Carlo steps per spin (MCS/S) and include 2000 spin configurations, from the subsequent 10^6 MCS/S; thus, we obtain a total of 2×10^5 spin-configuration snapshots, which we then use to train our neural networks (NNs). We carry out similar MC simulations to obtain training data for the Blume-Capel model (4), with training data in the range $[T_c - 0.5, T_c + 0.5]$, where $T_c \simeq 1.693$, and for the Ising-metamagnet model (6), with training data in the range $[T_c - 1, T_c + 1]$, where $T_c \simeq 5.263$. For the 2D Blume-Capel model (4) and the Ising-metamagnet model (6), while obtaining the spin configurations to train NNs in the vicinity of their tricritical points, we discard the first 10^7 MCS/S, and include 2000 spin configurations from the next 10^7 steps for each of the 100 values of temperature in the ranges $[T_c - 0.125, T_c + 0.125]$ ($T_c \simeq 0.608$) and $[T_c - 0.75, T_c + 0.75]$ ($T_c \simeq 2.41$), respectively.

To test the NNs in the vicinity of the first-order transition for the 2D Ising model (1), we train systems of linear size $L = 5,\ 6,\ 9$, where we obtain the training spin configurations from MC simulations, in the temperature range $[T_c - 1, T_c + 1]$ (see discussion above).

We train our NNs for 9000–10 000 epochs and save the NNs, after every 100 epochs; finally, we obtain 10 NNs, for each system size, which we use on our test data set (Appendix B). We train our NNs with the Adam Optimizer [80], with an initial learning rate of 10^{-3} and a batch size of 256. To prevent the overfitting of NNs, we set the regularizer strength $\lambda = 0.005$ in Eq. (16).

APPENDIX B: NEURAL NETWORKS AND DATASETS FOR TESTING AND ERROR ESTIMATIONS

In the vicinity of a second-order transition, we calculate P_L for the test data as follows: We discard the first 10^6 MCS/S, at each temperature, and consider 5000 snapshots from the next 10^6 MCS/S. We draw 2000 snapshots randomly, five times, from these 5000 snapshots. This yields five values of P_L, per NN, so, for 10 NNs (see Appendix A), we obtain 50 values from which we obtain the mean and the error for P_L. From the 50 estimates for P_L, obtained for the Ising model (1), we obtain $y_t = 1.06(7)$, and $y_h = 1.73(26)$ for the CNNs. Similarly, for the FCNNs, we obtain $y_t = 1.0(3)$ and $y_h = 1.80(30)$. In Fig. 2 we use the CNN set for which $\mathcal{P}_L(0, 0)$ is close to 0.5. In Table I we use the NN set that gives the estimates of y_t and y_h which are close to the known values; to obtain the error estimates for this NN set, we draw 2000 spin-configurations randomly, from the 5000 spin configurations; we repeat this 10 times and obtain the mean and the standard deviations for the best fits for the exponents (we use the same procedure for error estimation for all the models in the vicinity of a second-order transition).

We use the procedure, described in the previous paragraph, for the Blume-Capel (4) model at $\Delta = 0$. This yields $y_t = 1.05(6)$ and $y_h = 1.91(30)$, for the CNNs, and $y_t = 1.06(4)$ and $y_h = 1.77(25)$, for the FCNNs. In Figs. 3(a) and 3(b), we used the CNNs for which $P_L(0, 0)$ is close to 0.5. In Table II we use the CNN and FCNN sets that yield estimates close to known values of y_t and y_h. For $\Delta = 1$, when we use the CNNs trained for $\Delta = 0$, this yields $y_t = 1.05(5)$ and $y_h = 1.76(23)$; similarly, for FCNNs, we obtain $y_t = 1.03(9)$ and $y_h = 2.00(30)$. For $\Delta = 1$, in Figs. 3(c) and 3(d) and Table II, we use the CNN and FCNN set that we have used for $\Delta = 0$ [Figs. 3(a) and 3(b)].

A similar procedure for the Ising metagmagnet (6) model at $H = 0$ yields $y_t = 1.07(4)$ and $y_h = 1.74(27)$, for the CNNs, and $y_t = 1.08(4)$ and $y_h = 1.85(39)$, for the FCNNs. In Figs. 4(a) and 4(b) and Table IV, we use the CNN and the FCNN sets that yield estimates close to known values of y_t and y_h. For $H = 1$, the CNNs trained with $H = 0$ yield $y_t = 1.09(3)$ and $y_h = 1.70(30)$; similarly, for the FCNNs, we obtain $y_t = 1.1(5)$ and $y_h = 1.89(31)$. In Figs. 4(c) and 4(d), and Table IV, for $H = 1$, we use the NN set that we employ for $H = 0$ [Figs. 4(a) and 4(b)].

In the vicinity of the tricritical point, we calculate P_L for the test data as follows: We discard the first 10^7 MCS/S, at each temperature, and consider 5000 snapshots from the next 10^7 MCS/S. We draw 2000 snapshots randomly, five times, from these 5000 snapshots. This yields five values of P_L, per NN, so, for 10 NNs (see Appendix A), we obtain 50 values from which we obtain the mean and the error for P_L. From the 50 estimates for P_L, obtained for the Blume-Capel model (4), we obtain $y_t = 1.51(8)$, and $y_h = 1.96(44)$ for the CNNs. Similarly, for the FCNNs, we obtain $y_t = 2.00(15)$ and $y_h = 2.08(27)$. In Figs. 3(e) and 3(f), and Table III, we use the NN set that gives the estimates of y_t and y_h which are close to the known values; to obtain the error estimates for this NN set, we draw 2000 spin configurations randomly, from the 5000 spin configurations; we repeat this 50 times and obtain the mean and the standard deviations for the best fits for the exponents [we use the same procedure for error estimation for the Ising-metamagnet model (6) in the vicinity of its tricritical point].

We use the procedure, described in the previous paragraph for the Ising-metamagnet (6) model in the vicinity of the tricritical point. This yields $y_t = 1.66(3)$ and $y_h = 2.64(75)$ for the CNNs, and $y_t = 1.67(9)$ and $y_h = 2.21(73)$ for the FCNNs. In Figs. 4(e) and 4(f), and Table V, we use the CNN and FCNN sets that yield estimates close to known values of y_t and y_h.

In the vicinity of the first-order boundary for the Ising model (1), we obtain the $P_L(T = 2.1, h)$ in Fig. 6, by averaging over all 10 CNNs, for each L (see the discussion in Appendix A).

TABLE VI. Loss (16) and classification accuracy for the training and validation data, for the 2D Ising model (1) with $L = 12,\ 16,\ 20,$ and 30 and for different values of λ (17) for our FCNN (14).

Metric	λ	$L = 12$		$L = 16$		$L = 20$		$L = 30$	
		Training	Validation	Training	Validation	Training	Validation	Training	Validation
Loss	0	0.279(1)	0.361(1)	0.167(2)	0.381(2)	0.045(4)	1.04(2)	0.019(2)	0.338(3)
Accuracy	0	0.880(1)	0.848(1)	0.937(1)	0.869(1)	0.982(2)	0.879(1)	0.990(1)	0.929(1)
Loss	10^{-4}	0.299(3)	0.365(2)	0.204(1)	0.354(2)	0.129(4)	0.374(4)	0.074(3)	0.302(3)
Accuracy	10^{-4}	0.877(1)	0.850(1)	0.929(1)	0.8769(4)	0.926(2)	0.908(1)	0.983(1)	0.926(1)
Loss	10^{-3}	0.326(3)	0.359(2)	0.256(9)	0.30(1)	0.199(3)	0.244(3)	0.140(3)	0.154(3)
Accuracy	10^{-3}	0.866(1)	0.853(1)	0.905(4)	0.885(4)	0.927(2)	0.909(2)	0.940(1)	0.939(1)
Loss	10^{-2}	0.343(3)	0.347(3)	0.269(1)	0.271(2)	0.221(5)	0.224(4)	0.157(1)	0.161(1)
Accuracy	10^{-2}	0.855(2)	0.854(2)	0.894(1)	0.891(1)	0.911(4)	0.911(2)	0.940(1)	0.939(1)
Loss	10^{-1}	0.355(5)	0.358(5)	0.287(7)	0.286(7)	0.233(4)	0.235(4)	0.137(2)	0.134(7)
Accuracy	10^{-1}	0.853(2)	0.852(3)	0.889(2)	0.888(2)	0.912(1)	0.911(1)	0.939(1)	0.938(1)
Loss	1	0.372(7)	0.373(8)	0.302(8)	0.302(7)	0.248(1)	0.248(4)	0.152(5)	0.142(7)
Accuracy	1	0.850(3)	0.849(3)	0.887(3)	0.885(3)	0.908(3)	0.909(4)	0.936(4)	0.935(5)

In Fig. 7(a) we obtain $\mathcal{P}_L(|M|)$ as follows: we calculate $|M|$ using $\boldsymbol{\sigma}$ from the test data (see Appendix B) and consider the interval $[0, 1]$, dividing it into 10 bins, i.e., $[0, 0.1], [0.1, 0.2], \ldots, [0.9, 1]$; for $\boldsymbol{\sigma}$ with $|M|$ lying in each of these bins, we calculate the average of $\hat{P}(\sigma)$ and plot it versus the average value of $|M|$.

APPENDIX C: NEURAL NETWORK EXPLORATIONS

In Table VI we show how the regularization strength λ affects the loss (16) and the classification accuracy of our FCNN (14) for the training and validation data obtained from our MC simulations of the 2D Ising model (1). From Table VI we observe that, small values of λ (i.e., $\lambda \in [0, 10^{-3})$) lead to overfitting, i.e., the loss (16) (accuracy) for the training data is higher (lower) than that for the validation data. For $\lambda \in [10^{-3}, 10^{-2}]$, we see that (a) the training- and validation-data loss (16) (accuracy) are comparable and (b) there is an improvement in the validation-data loss (16) (accuracy) compared to that for low values of λ; hence, this is a suitable range for λ for better generalization. For $\lambda \in (10^{-2}, 1]$, we find underfitting, i.e., there is a increase (decrease) in the overall loss (16) (accuracy) for both the training and validation data.

In Table VII we obtain the training data from our MC simulations of the 2D Blume-Capel model (4) at $\Delta = 0$ and the validation data from such simulations for the 2D Blume-Capel model at $\Delta = 1$, and we present the classification accuracies of our FCNN (14) for different values of λ. Here we observe that the dependence of the generalization capabilities on λ of our FCNN is similar to that mentioned in the previous paragraph for the 2D Ising model.

In Table VIII we give the loss (16) and classification accuracy for the FCNN, which has the same architecture as in Eq. (14) but with 128 nodes, and for the CNN with the same architecture as in Eq. (15) but with 64 filters and 128 nodes. Here we set $\lambda = 0.05$; the loss (16) and classification accuracy in Table VIII are similar to those in Table VI with $\lambda \in [10^{-3}, 10^{-2}]$.

In Tables VI–VIII, we trained the FCNNs for 1000 epochs for each value of λ. The training and validation data each consist of 2×10^5 snapshots (spin configurations). In Tables VI–VIII, we present the mean and our error estimates obtained from 10 FCNNs saved every 10 epochs, from the 900th to the 1000th training epoch.

The training time for our neural networks increases with system size; for example, the training time for our FCNN increases from $\simeq 5 \times 10^3$ seconds to $\simeq 7.5 \times 10^3$ seconds (on a single NVIDIA A100 GPU) for system sizes $L = 12$ to $L = 30$ for the Ising model (1) (similarly, the training time for the CNN increases from $\simeq 7.5 \times 10^3$ seconds to $\simeq 10 \times 10^3$ seconds). Furthermore, when we train our FCNN with $L = 50$, the training time increases to $\simeq 12 \times 10^3$ seconds ($\simeq 15 \times 10^3$ seconds for the CNN]) Including $L = 40$ and $L = 50$ in our FSS yielded the best-fit exponents $y_t = 1.06(2)$ and $y_h = 1.82(1)$ for the CNN (15), and $y_t = 1.04(1)$ and $y_h = 1.85(1)$ for the FCNN. It would be interesting to investigate our method for larger system sizes, employing sampling methods such as Wolff [76] or Swendsen-Wang [77] algorithms.

APPENDIX D: RELATION BETWEEN SYSTEM SIZE, TWO-SCALE FACTOR, REDUCED TEMPERATURE, AND CLASSIFICATION ACCURACY

In Eq. (30) we had shown that, if N_L^α is the number of spin configurations that are correctly classified out of a total N configurations for a system of linear size L, say, below T_c^α and in the temperature range $[T_0^\alpha, T_c^\alpha]$, then

$$\frac{\left(N - N_L^\alpha\right)}{N} \propto \frac{1}{f_t^\alpha \left|t_0^\alpha\right| L^{y_t}}, \tag{D1}$$

where $t_0^\alpha \equiv \frac{T_0 - T_c^\alpha}{T_c}$. We derive this relation now.

Let n_s be the number of spin configurations (henceforth, snapshots) that we use in our testing data set for reduced temperatures that lie between t and $t + dt$, where $t = \frac{T - T_c}{T_c}$. Out of these n_s snapshots, let $n_L(t)$ be the number of snapshots that are correctly classified (e.g., as lying below T_c), for a system with linear size L. By definition,

$$n_L(t) = \mathcal{F}_L(t) n_s, \tag{D2}$$

TABLE VII. Loss (16) and classification accuracy for the FCNN (14), whose training data were obtained for the 2D Blume-Capel (4) at $\Delta = 0$, and validation data obtained at $\Delta = 1$, for $L = 12,\ 16,\ 20,$ and 30 with different values of λ (17).

Metric	λ	$L=12$		$L=16$		$L=20$		$L=30$	
		Training	Validation	Training	Validation	Training	Validation	Training	Validation
Loss	0	0.182(1)	0.269(4)	0.089(2)	0.368(2)	0.0041(4)	2.24(3)	0.001(1)	1.53(4)
Accuracy	0	0.927(1)	0.905(1)	0.967(1)	0.911(1)	0.999(1)	0.908(1)	0.9997(2)	0.9404(4)
Loss	10^{-4}	0.203(5)	0.274(5)	0.128(4)	0.306(3)	0.129(4)	0.374(4)	0.054(6)	0.416(6)
Accuracy	10^{-4}	0.923(2)	0.908(2)	0.958(2)	0.917(1)	0.983(5)	0.915(3)	0.9929(2)	0.935(1)
Loss	10^{-3}	0.228(3)	0.25(1)	0.179(1)	0.201(5)	0.199(3)	0.244(3)	0.123(7)	0.128(5)
Accuracy	10^{-3}	0.913(2)	0.909(7)	0.936(1)	0.930(2)	0.949(1)	0.944(2)	0.961(1)	0.959(2)
Loss	10^{-2}	0.242(3)	0.235(6)	0.195(3)	0.183(6)	0.221(4)	0.224(4)	0.130(2)	0.128(7)
Accuracy	10^{-2}	0.907(1)	0.910(4)	0.929(1)	0.933(4)	0.942(1)	0.943(2)	0.9586(4)	0.961(2)
Loss	10^{-1}	0.248(3)	0.247(9)	0.203(5)	0.194(1)	0.233(4)	0.234(4)	0.137(2)	0.134(7)
Accuracy	10^{-1}	0.906(1)	0.912(3)	0.928(2)	0.930(5)	0.957(1)	0.960(1)	0.958(1)	0.960(1)
Loss	1	0.259(9)	0.248(9)	0.216(9)	0.198(7)	0.198(7)	0.248(4)	0.152(5)	0.142(7)
Accuracy	1	0.905(2)	0.912(3)	0.926(3)	0.935(2)	0.955(2)	0.960(1)	0.955(4)	0.960(5)

where $\mathcal{F}_L(t) = \int d\boldsymbol{\sigma}\, \mathfrak{P}_L(\boldsymbol{\sigma}, t, 0) H[\hat{P}_L(\boldsymbol{\sigma}) - 0.5]$, the Heaviside function $H(x) = 0$, if $x < 0$, and $H(x) = 1$, if $x \geqslant 0$, and we use 0.5 as the threshold for classifying the spin configurations. Arguments similar to those used in Eqs. (22), (26), and (27) give the same FSS form for $\mathcal{F}_L(t)$ as that for $\mathcal{P}_L(t, 0)$.

If n is the number of sets of such (n_s) snapshots, in the intervals $[t_i, t_i + dt]$, where $i \in [1, 2, \ldots, n]$, and total number of snapshots is $N = n_s * n$, then the total number of snapshots correctly classified (e.g., as lying below T_c) is

$$N_L = \sum_{i=1}^{n} n_L(t_i) = \sum_{i=1}^{n} \mathcal{F}_L(t_i) \frac{N}{n}. \tag{D3}$$

Using $\frac{T_c - T_0}{T_c} = ndt = |t_0|$ in Eq. (D3) we get

$$\begin{aligned} \frac{N_L}{N} &= \frac{1}{|t_0|} \sum_{i=1}^{n} \mathcal{F}_L(t_i) dt, \text{ and, if } n \to \infty, \\ &= \frac{1}{|t_0|} \int_{t_0}^{0} \mathcal{F}_L(t) dt \\ &= \frac{1}{|t_0| L^{y_t}} \int_{x_L^0}^{0} \mathcal{F}_L^*(x) dx, \end{aligned} \tag{D4}$$

where we have used $\mathcal{F}_L(t) \equiv \mathcal{F}_L^*(tL^{y_t}) = \mathcal{F}_L^*(x)$ and changed the variable of integration from t to $x = tL^{y_t}$. From Eq. (D4) and using

$$\begin{aligned} \int_{x_L^o}^{0} \mathcal{F}_L^*(x) dx &= \int_{x_L^0}^{X^0} \mathcal{F}_L^*(x) dx + \int_{X^0}^{0} \mathcal{F}_L^*(x) dx \\ &\approx X^0 - x_L^0 + \int_{X^0}^{0} \mathcal{F}_L^*(x) dx \\ &= X^0 - t_0 L^{y_t} + \int_{X^0}^{0} \mathcal{F}_L^*(x) dx \\ &= a + |t_0| L^{y_t}, \end{aligned} \tag{D5}$$

where $\mathcal{F}^*(x)$ is asymptotically close to 1 $\forall x_L^0 < X^0$, we get

$$\frac{N_L}{N} \approx 1 + \frac{a}{|t_0| L^{y_t}} \Rightarrow \frac{(N - N_L)}{N} \approx \frac{A}{|t_0| L^{y_t}}, \tag{D6}$$

where $A = -a$. If we now consider $\mathcal{F}_L^{*\alpha}(x)$, for the Ising ($\alpha = I$), Blume-Capel ($\alpha = \mathrm{BC}$), and metamagnet ($\alpha = M$) models (1), (4), and (6), we must include the scale factors f_t^α [see Eq. (28)], and use $\mathcal{F}_L^{*\alpha}(f_t^\alpha x) \equiv \mathcal{F}_L^*(x)$ to obtain

$$\begin{aligned} a^\alpha &= X^{0,\alpha} + \int_{X^{0,\alpha}}^{0} \mathcal{F}_L^{*\alpha}(x') dx' \\ &= \frac{X^0}{f_t^\alpha} + \frac{\int_{X^0}^{0} \mathcal{F}_L^*(x) dx}{f_t^\alpha} = \frac{a}{f_t^\alpha}. \end{aligned} \tag{D7}$$

TABLE VIII. Loss (16) and classification accuracy for the training and validation data, for the 2D Ising model (1) with $L = 12,\ 16,\ 20,$ and 30, for the FCNN [similar to Eq. (14)] with 128 hidden nodes, and for the CNN [similar to Eq. (15)] with 128 hidden nodes and 64 filters.

NN	Metric	$L=12$		$L=16$		$L=20$		$L=30$	
		Training	Validation	Training	Validation	Training	Validation	Training	Validation
FCNN	Loss	0.339(3)	0.345(3)	0.265(1)	0.268(2)	0.216(2)	0.222(3)	0.153(2)	0.158(2)
FCNN	Accuracy	0.856(1)	0.854(1)	0.895(1)	0.892(1)	0.916(1)	0.914(1)	0.941(1)	0.939(1)
CNN	Loss	0.322(1)	0.323(1)	0.247(1)	0.244(1)	0.188(2)	0.190(2)	0.122(1)	0.125(1)
CNN	Accuracy	0.862(1)	0.862(1)	0.900(1)	0.898(1)	0.921(1)	0.921(1)	0.949(1)	0.948(1)

Then the relation for the number of correctly classified snapshots for $t_0^\alpha \leqslant t \leqslant 0$, N_L^α is

$$\frac{\left(N - N_L^\alpha\right)}{N} = \frac{A}{f_t^\alpha \left|t_0^\alpha\right| L^{y_t}}. \tag{D8}$$

We verify this relation in Fig. 7(b), where we plot $\frac{(N-N_L^\alpha)}{N}$ vs $f_t^\alpha |t_0^\alpha| L^{y_t}$ for the Ising (I) (1), Blume-Capel (BC) (4) and Ising-metamagnet (M) (6) models, for systems with linear sizes $L = 12$, 16, 20, and 30 [in Fig. 7(b), we use the temperature range $[T_c - 1, T_c]$ for the Ising (1) and Blume-Capel (4) models, and $[T_c - 1.5, T_c]$ for the Ising-metamagnet model (6)].

[1] M. E. Fisher, The renormalization group in the theory of critical behavior, Rev. Mod. Phys. **46**, 597 (1974).

[2] M. E. Fisher, Scaling, universality and renormalization group theory, Critical Phenomena (Stellenbosch 1982) 1 (1983).

[3] L. P. Kadanoff, Critical behavior. universality and scaling, in *From Order To Chaos: Essays: Critical, Chaotic and Otherwise*, edited by F. J. W. Hahne (World Scientific, Singapore, 1993), pp. 222–239.

[4] M. E. Fisher, Renormalization group theory: Its basis and formulation in statistical physics, Rev. Mod. Phys. **70**, 653 (1998).

[5] C. Domb, *Phase Transitions and Critical Phenomena, Vol. 6*, Phase Transitions and Critical Phenomena (Elsevier, 2000).

[6] L. P. Kadanoff, *Statistical Physics: Statics, Dynamics and Renormalization* (World Scientific, 2000).

[7] M. Kardar, *Statistical Physics of Fields* (Cambridge University Press, 2007).

[8] N. Goldenfeld, *Lectures on Phase Transitions and the Renormalization Group* (CRC Press, 2018).

[9] M. Mezard and A. Montanari, *Information, Physics, and Computation* (Oxford University Press, 2009).

[10] 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).

[11] P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, A high-bias, low-variance introduction to machine learning for physicists, Phys. Rep. **810**, 1 (2019).

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

[13] A. Morningstar and R. G. Melko, Deep learning the Ising model near criticality, J. Mach. Learn. Res. **18**, 1 (2018).

[14] Z. Li, M. Luo, and X. Wan, Extracting critical exponents by finite-size scaling with convolutional neural networks, Phys. Rev. B **99**, 075418 (2019).

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

[16] D. Kim and D.-H. Kim, Emergence of a finite-size-scaling function in the supervised learning of the Ising phase transition, J. Stat. Mech. (2021) 023202.

[17] K. Shiina, H. Mori, Y. Okabe, and H. K. Lee, Machine-learning studies on spin models, Sci. Rep. **10**, 2177 (2020).

[18] V. Chertenkov, E. Burovski, and L. Shchur, Finite-size analysis in neural network classification of critical phenomena, Phys. Rev. E **108**, L032102 (2023).

[19] D. Bayo, B. Çivitcioğlu, J. J. Webb, A. Honecker, and R. A. Römer, Machine learning of phases and structures for model systems in physics, J. Phys. Soc. Jpn. **94**, 031002 (2025).

[20] J. Shen, W. Li, S. Deng, and T. Zhang, Supervised and unsupervised learning of directed percolation, Phys. Rev. E **103**, 052140 (2021).

[21] D. Bayo, A. Honecker, and R. A. Römer, Machine learning the 2D percolation model, in *Journal of Physics: Conference Series*, Vol. 2207 (IOP Publishing, 2022), p. 012057.

[22] D. Bayo, A. Honecker, and R. A. Römer, The percolating cluster is invisible to image recognition with deep learning, New J. Phys. **25**, 113041 (2023).

[23] Y. Wang, W. Li, F. Liu, and J. Shen, Supervised and unsupervised learning of-dimensional even-offspring branching annihilating random walks, Mach. Learn.: Sci. Tech. **5**, 015033 (2024).

[24] J. Venderley, V. Khemani, and E.-A. Kim, Machine learning out-of-equilibrium phases of matter, Phys. Rev. Lett. **120**, 257204 (2018).

[25] Y. Tang, J. Liu, J. Zhang, and P. Zhang, Learning nonequilibrium statistical mechanics and dynamical phase transitions, Nat. Commun. **15**, 1117 (2024).

[26] A. Tanaka and A. Tomiya, Detection of phase transition via convolutional neural networks, J. Phys. Soc. Jpn. **86**, 063001 (2017).

[27] 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).

[28] 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).

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

[30] R. Zhang, B. Wei, D. Zhang, J.-J. Zhu, and K. Chang, Few-shot machine learning in the three-dimensional Ising model, Phys. Rev. B **99**, 094427 (2019).

[31] Z. Tian, S. Zhang, and G.-W. Chern, Machine learning for structure-property mapping of Ising models: Scalability and limitations, Phys. Rev. E **108**, 065304 (2023).

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

[33] 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).

[34] M. E. Fisher and M. N. Barber, Scaling theory for finite-size effects in the critical region, Phys. Rev. Lett. **28**, 1516 (1972).
[35] M. N. Barber, Finite-size scaling, in *Phase Transitions Critical Phenomena*, edited by C. Domb and J. L. Lebowitz (Academic Press, London, 1983), pp. 146–266.
[36] M. E. Fisher and A. N. Berker, Scaling for first-order phase transitions in thermodynamic and finite systems, Phys. Rev. B **26**, 2507 (1982).
[37] K. Binder and D. P. Landau, Finite-size scaling at first-order phase transitions, Phys. Rev. B **30**, 1477 (1984).
[38] D. Stauffer, M. Ferer, and M. Wortis, Universality of second-order phase transitions: The scale factor for the correlation length, Phys. Rev. Lett. **29**, 345 (1972).
[39] A. Aharony, Two-scale-factor universality and the ϵ expansion, Phys. Rev. B **9**, 2107 (1974).
[40] P. D. Beale, Finite-size scaling study of the two-dimensional Blume-Capel model, Phys. Rev. B **33**, 1717 (1986).
[41] N. B. Wilding and P. Nielaba, Tricritical universality in a two-dimensional spin fluid, Phys. Rev. E **53**, 926 (1996).
[42] W. Kwak, J. Jeong, J. Lee, and D.-H. Kim, First-order phase transition and tricritical scaling behavior of the Blume-Capel model: A Wang-Landau sampling approach, Phys. Rev. E **92**, 022134 (2015).
[43] L. Moueddene, N. G. Fytas, Y. Holovatch, R. Kenna, and B. Berche, Critical and tricritical singularities from small-scale Monte Carlo simulations: The Blume–Capel model in two dimensions, J. Stat. Mech. (2024) 023206.
[44] P. A. Rikvold, W. Kinzel, J. D. Gunton, and K. Kaski, Finite-size-scaling study of a two-dimensional lattice-gas model with a tricritical point, Phys. Rev. B **28**, 2686 (1983).
[45] P. Beale, Finite-size scaling at an Ising tricritical point, J. Phys. A: Math. Gen. **17**, L335 (1984).
[46] D. P. Landau and R. H. Swendsen, Tricritical universality in two dimensions, Phys. Rev. Lett. **46**, 1437 (1981).
[47] H. Herrmann, Finite size scaling approach to a metamagnetic model in two dimensions, Phys. Lett. A **100**, 256 (1984).
[48] A. M. Ferrenberg, J. Xu, and D. P. Landau, Pushing the limits of monte carlo simulations for the three-dimensional Ising model, Phys. Rev. E **97**, 043301 (2018).
[49] J. Xu, A. M. Ferrenberg, and D. P. Landau, High-resolution Monte Carlo study of the order-parameter distribution of the three-dimensional Ising model, Phys. Rev. E **101**, 023315 (2020).
[50] P. Butera and M. Comi, Critical universality and hyperscaling revisited for Ising models of general spin using extended high-temperature series, Phys. Rev. B **65**, 144431 (2002).
[51] H. W. Blöte, L. N. Shchur, and A. L. Talapov, The cluster processor: New results, Int. J. Mod. Phys. C **10**, 1137 (1999).
[52] Y. Deng and H. W. J. Blöte, Simultaneous analysis of several models in the three-dimensional Ising universality class, Phys. Rev. E **68**, 036125 (2003).
[53] Y. Ozeki and N. Ito, Nonequilibrium relaxation method, J. Phys. A: Math. Theor. **40**, R149 (2007).
[54] M. Weigel and W. Janke, Error estimation and reduction with cross correlations, Phys. Rev. E **81**, 066701 (2010).
[55] M. Hasenbusch, Finite size scaling study of lattice models in the three-dimensional Ising universality class, Phys. Rev. B **82**, 174433 (2010).
[56] F. Kos, D. Poland, D. Simmons-Duffin, and A. Vichi, Precision islands in the Ising and $O(N)$ models, J. High Energy Phys. 08 (2016) 036.
[57] S. Wang, Z.-Y. Xie, J. Chen, B. Normand, and T. Xiang, Phase transitions of ferromagnetic Potts models on the simple cubic lattice, Chin. Phys. Lett. **31**, 070503 (2014).
[58] L. Onsager, Crystal statistics. I. A two-dimensional model with an order-disorder transition, Phys. Rev. **65**, 117 (1944).
[59] B. M. McCoy and T. T. Wu, *The Two-Dimensional Ising Model* (Harvard University Press, 1973).
[60] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Elsevier, 2016).
[61] C. J. Thompson, *Mathematical Statistical Mechanics* (Princeton University Press, 2015).
[62] D. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, 2021).
[63] V. Privman, *Finite Size Scaling and Numerical Simulation of Statistical Systems* (World Scientific, 1990).
[64] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, TensorFlow: A system for Large-Scale machine learning, in *12th USENIX Symposium on Operating systems Design and Implementation (OSDI 16)* (USENIX, Savannah, GA, USA, 2016), pp. 265–283.
[65] K. Binder, Finite size scaling analysis of Ising model block distribution functions, Z. Phys. B **43**, 119 (1981).
[66] N. Wilding and A. Bruce, Density fluctuations and field mixing in the critical fluid, J. Phys.: Condens. Matter **4**, 3087 (1992).
[67] S. T. Bramwell, J.-Y. Fortin, P. C. W. Holdsworth, S. Peysson, J.-F. Pinton, B. Portelli, and M. Sellitto, Magnetic fluctuations in the classical *XY* model: The origin of an exponential tail in a complex system, Phys. Rev. E **63**, 041106 (2001).
[68] D. Lopes Cardozo and P. C. W. Holdsworth, Finite-size scaling of the magnetization probability density for the critical Ising model in slab geometry, J. Phys.: Condens. Matter **28**, 166007 (2016).
[69] V. Mozolenko and L. Shchur, Blume-Capel model analysis with a microcanonical population annealing method, Phys. Rev. E **109**, 045306 (2024).
[70] P. C. Hohenberg, A. Aharony, B. I. Halperin, and E. D. Siggia, Two-scale-factor universality and the renormalization group, Phys. Rev. B **13**, 2986 (1976).
[71] B. Nienhuis and M. Nauenberg, Renormalization-group theory and calculations of tricritical behavior, Phys. Rev. B **13**, 2021 (1976).
[72] W. Klein, D. Wallace, and R. Zia, Essential singularities at first-order phase transitions, Phys. Rev. Lett. **37**, 639 (1976).
[73] A. N. Berker and M. Wortis, Blume-emery-griffiths-potts model in two dimensions: Phase diagram and critical properties from a position-space renormalization group, Phys. Rev. B **14**, 4946 (1976).
[74] A nonperturbative RG, of the type used in Ref. [73], is essential for obtaining such fixed points that lie at infinite values of couplings.
[75] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Elsevier, 2023).
[76] U. Wolff, Collective Monte Carlo updating for spin systems, Phys. Rev. Lett. **62**, 361 (1989).

[77] R. H. Swendsen and J.-S. Wang, Nonuniversal critical dynamics in Monte Carlo simulations, Phys. Rev. Lett. **58**, 86 (1987).

[78] D. Wu, L. Wang, and P. Zhang, Solving statistical mechanics using variational autoregressive networks, Phys. Rev. Lett. **122**, 080602 (2019).

[79] F. D'Angelo and L. Böttcher, Learning the Ising model with generative neural networks, Phys. Rev. Res. **2**, 023266 (2020).

[80] D. P. Kingma, Adam: A method for stochastic optimization, arXiv:1412.6980.