Detecting the three-dimensional Ising model phase transition with a ground-state-trained autoencoder

Ahmed Abuali [1,*] David A. Clarke [2,†] Morten Hjorth-Jensen [3,‡] Ioannis Konstantinidis [4] Claudia Ratti [1] and Jianyi Yang [4]

[1] *Physics Department, University of Houston, Houston, Texas 77204, USA*

[2] *Fakultät für Physik, Universität Bielefeld, D-33615 Bielefeld, Germany*

[3] *Department of Physics and Center for Computing in Science Education, University of Oslo, N-0316 Oslo, Norway*

[4] *Computer Science Department, University of Houston, Houston, Texas 77204, USA*

(Received 31 March 2026; revised 8 July 2026; accepted 16 July 2026; published 27 July 2026)

We develop a one-class, deep-learning framework to detect the phase transition and recover critical behavior of the three-dimensional (3D) Ising model. A 3D convolutional autoencoder (CAE) is trained on ground-state configurations only, without prior knowledge of the critical temperature, the Hamiltonian, or the order parameter. After training, the model is applied to Monte Carlo configurations across a wide temperature range and different lattice sizes. The mean-square reconstruction error is shown to be sensitive to the transition. Finite-size scaling of the peak location for the reconstruction error susceptibility yields the critical temperature $T_c = 4.5128(58)$, in excellent agreement with the known value. Additionally, we obtain an estimate of the correlation-length critical exponent, $\nu = 0.63(27)$, also consistent with results from the literature. Our results show that a one-class CAE, trained on zero-temperature configurations only, can recover nontrivial critical behavior of the 3D Ising model.

DOI: 10.1103/h8n4-94g4

I. INTRODUCTION

In recent years, machine-learning (ML) techniques have emerged as powerful tools for discovering structure in many-body systems directly from microscopic data. Supervised neural networks have been successfully trained to distinguish phases when supplied with labeled data [1–9], while unsupervised techniques, including but not limited to principal component analysis, clustering algorithms, and autoencoders, reveal that the enormous space of spin configurations actually contains structures that correspond to physical ordering [10–17]. Generically speaking, supervised ML models are trained on many temperatures above and below the transition. In Refs. [2,5,8,9], minimal training set models, i.e., models training on one or two out-of-distribution temperatures, were shown to be capable of recognizing phases and, in some cases, extracting critical parameters.

*Contact author: amabuali@uh.edu; ahmed.abuali1975@gmail.com

†Contact author: clarke.davida@gmail.com

‡Contact author: mhjensen@uio.no

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Choosing one or two sets with unambiguous phases is a strategy that allows one to train even in cases where labeling is challenging, for example, in systems with crossovers, and it has the computational advantage of significantly reducing the training time.

Given the analytic tractability of the two-dimensional (2D) Ising model, studies showcasing the capabilities of ML to detect its phase transitions are plentiful; in addition to some of the above studies, one finds numerous examples in, e.g., Ref. [18] and references therein. Such studies for the three-dimensional (3D) Ising model, meanwhile, remain comparatively sparse [17,19–24]. In these studies, when training is employed, it is carried out across many temperatures in a range covering the critical temperature.

This work extends ML studies of the 3D Ising model by applying a fundamentally different strategy, employing an autoencoder trained on $T = 0$ configurations only, without exposure to the critical region or explicit physical knowledge such as the Hamiltonian or order parameter. The trained autoencoder is then evaluated on independent Monte Carlo configurations spanning the full temperature range to determine whether signatures of the phase transition emerge solely from its learned representation. While the envisioned use case of such minimal training strategies is systems without a clear order parameter—such as QCD, where identifying the crossover width remains a fundamental challenge—we choose here, as with our previous investigation of the 2D Ising model [9], a well-studied system with a second-order transition, as it offers known critical parameters that can be used as fiducial points. If the approach can reproduce the known T_c, it is encouraging for application to crossover systems to define a meaningful pseudocritical temperature; if it can additionally reproduce some of the critical exponents, it demonstrates that minimally trained neural networks can recover further nontrivial, physics-relevant patterns in spin configurations.

We adopt a similar approach as our previous study of the 2D Ising model [9]; however, in this work we use an architecture with convolutional layers trained on only $T = 0$ configurations, rather than training on both $T = 0$ and $T = \infty$ configurations. Moving from two to three spatial dimensions considerably increases the complexity of the input; therefore, instead of a convolutional neural network binary classifier, we employ a convolutional autoencoder (CAE) to find simplified representations of the configurations. Despite this extremely restricted training regimen, we find that the CAE output responds to changes in phase and is capable of accurately extracting T_c and ν.

The outline of this paper is as follows. We start in Sec. II with a brief summary of the 3D Ising model, along with results from statistical physics studies that will be used to benchmark our CAE. In Sec. III, we describe our Monte Carlo data, ML architecture, and training strategy. The output of the CAE is presented in Sec. IV, showing that this approach can successfully extract T_c and ν. We close with Sec. V, where we discuss our findings and the outlook.

II. 3D ISING MODEL AND FINITE-SIZE SCALING

The 3D Ising model consists of spins $\sigma_i = \pm 1$ residing on the sites of a lattice of volume $V = L^3$ with Hamiltonian

$$H = -J \sum_{\langle i\, j \rangle} \sigma_i \sigma_j, \tag{1}$$

where the sum runs over nearest-neighbor pairs and J is the interaction strength. For the remainder of this paper, we work in units $J = k_B = 1$. The order parameter is the absolute value of the magnetization $|m|$, where

$$m = \frac{1}{V} \sum_i \sigma_i. \tag{2}$$

As the system approaches the critical temperature T_c, the magnetic susceptibility

$$\chi = \frac{V}{T} (\langle |m|^2 \rangle - \langle |m| \rangle^2) \tag{3}$$

diverges. For finite L, the divergence in χ softens to a pronounced peak, whose location in T can be used to define a pseudocritical temperature $T_c(L)$. This pseudocritical temperature approaches T_c according to the critical exponent ν as

$$|T_c(L) - T_c| \sim L^{-1/\nu}. \tag{4}$$

Although the 2D Ising model admits an exact solution [25], no closed-form solution exists in 3D. A common strategy to glean quantitative understanding of the 3D Ising model is to use Markov chain Monte Carlo (MCMC) simulations, which is also the method we use in this study to generate configurations. Other methods include, but are not limited to, high-temperature series expansions and the conformal bootstrap. These methods have delivered robust results for the critical parameters over the past several decades. Taking a weighted average of many literature results [26–41] yields a benchmark $T_c = 4.511\,514(27)$; meanwhile, a weighted average for ν [26,32,33,35,37–40,42–53] delivers a benchmark 0.630 36(39) [54]. These averages are weighted by the uncertainty; in particular, for each quantity we compute the variance σ_W^2 and weighted mean μ_W as

$$\sigma_W^2 = \left(\sum_i \frac{1}{\sigma_i^2}\right)^{-1}, \quad \mu_W = \frac{1}{\sigma_W^2}\sum_i \frac{\mu_i}{\sigma_i^2}, \tag{5}$$

where the sums run over results from each study. Results from different papers are treated as statistically independent. The software that computes these averages is available in the AnalysisToolbox [55,56].

III. ML MODEL AND MCMC SET UP

In our previous study [9], we were interested in binary classification using training approaches that could be employed also when ensemble labels are not readily available, for example, when one does not know the location of T_c *a priori*. This motivated us to adopt a training regimen where the network is only shown classes at $T = 0$ and $T = \infty$, where the only required knowledge is the form of maximally ordered and maximally disordered configurations. Previous investigations of minimally trained networks have shown that only a single training class, in particular, the ground state, is sufficient for supervised models to detect phase changes in statistical physics systems [2,5]. Moreover, while training sets at $T = 0$ are exact, sets at $T = \infty$, realized as configurations with half up spins and half down spins, are only a very rough approximation, intended primarily to show the network what a maximally disordered configuration looks like. With this in mind, and motivated by the one-class studies, we choose to employ a one-class training regimen here.

The network we used for our 2D Ising study had one convolutional layer, because another 2D Ising study found that the same architecture was well optimized for the 2D Ising model when trained on multiple temperatures above and below T_c [6]. While we found that this architecture performed reasonably well for a two-set regimen in our previous study, it cannot be expected to succeed when applied to a one-set regimen with an additional spatial dimension. To help tackle the increased complexity when moving to 3D, we wanted an approach that can find reduced representations of the input configurations to expose their salient features. Moreover, using an autoencoder allows us to avoid an explicit label for our training set; hence, the approach taken in this study is unsupervised.

Therefore, we use a one-class 3D CAE, which is trained exclusively on ground-state configurations, i.e., configurations at $T = 0$. The autoencoder consists of an encoder e and a decoder d, which together define the reconstruction of an input configuration C as

$$\bar{C} = d(e(C)), \tag{6}$$

where $\bar{C}$ denotes the reconstructed spin configuration. The encoder progressively reduces the spatial resolution of the input configuration using strided 3D convolutions, thereby compressing the spin configuration into a lower-resolution representation. The decoder then reconstructs the configuration by applying a sequence of 3D transpose convolutions, which increase the spatial resolution in a

learnable manner until the original lattice size is recovered. Autoencoders are well suited for this task because they learn the statistical structure of a dataset directly from training sets without requiring labels or predefined physical observables. Details of the architecture are discussed in Appendix A.

To quantify the quality of the reconstruction $\bar{C}$, we employ the mean-square error (MSE)

$$\mathrm{MSE}(\bar{C}, C) \equiv \frac{1}{V}||\bar{C} - C||_2^2 = \frac{1}{V}\sum_i (\bar{\sigma}_i - \sigma_i)^2, \tag{7}$$

where the sum runs over all sites and $\bar{\sigma}_i$ is the reconstructed spin value of $\bar{C}$ at site i. Since the training is performed only on ground-state configurations, the MSE evaluated on $T > 0$ configurations measures how well $C(T)$ can be represented by the structure learned from the ground-state data. Configurations similar to those seen during training produce small MSEs, while thermally disordered configurations produce larger errors.

Spin configurations of size L^3 with $50 \leqslant L \leqslant 130$ for the 3D Ising model are generated using MCMC. For each L, we probe 84 temperatures in the range $T \in [3.0, 6.0]$. The temperatures are not evenly spaced, but rather denser close to our benchmark value of T_c. Simulations are adapted from the Metropolis checkerboard algorithm implementation described in Ref. [57]. For each lattice size, the system is first thermalized using a number of warm-up sweeps equal to 20% of the total number of Monte Carlo sweeps. For each temperature, we generate 3000 configurations; for temperatures near T_c, we generate 4000 configurations to enhance statistical resolution. After thermalization, measurements are separated by L^2 sweeps. For $L \leqslant 60$, configurations near T_c are instead separated by L^3 sweeps to reduce autocorrelation between successive measurements. This choice reflects the increase of the autocorrelation times with system size, particularly near the critical temperature. All MCMC configurations are evaluated by the trained neural network without further preprocessing. These configurations are used exclusively for evaluation; no $T > 0$ data are included in the training of the autoencoder.

The autoencoder is trained on 2000 ground-state $T = 0$ configurations, which consist of 1000 fully spin-up configurations and 1000 fully spin-down configurations. Including both ordered states ensures that the training set respects the global $\mathbb{Z}_2$ symmetry of the Ising model. After training, the model is applied to the MCMC configurations discussed above. Each configuration is represented as a 3D binary field, which serves as the input to the machine-learning model. No physical observables, temperature labels, or Hamiltonian parameters are provided to the network. In these respects, the CAE is completely physics agnostic.

Statistical analysis, curve fitting, and plotting were performed using AnalysisToolbox software [55,56].

IV. RESULTS

After applying our CAE to our MCMC data, we estimate $\langle \mathrm{MSE} \rangle$ and its statistical uncertainty for each temperature using a standard jackknife with 40 bins. In Fig. 1 (left), we show results for $\langle \mathrm{MSE} \rangle$. To check the stability of the CAE, we tested two independent (i.e., different starting seeds) CAEs on the same Markov chains, which we call CAE1 and CAE2, indicated in purple and blue. We show results from our $L = 50$ ensembles (solid lines) and our $L = 130$ ensembles (dotted lines). The inset shows the region $4.40 < T < 4.55$. We see that $\langle \mathrm{MSE} \rangle$ generically increases with temperature, which is the expected behavior, since the CAE is trained on ground-state configurations, and increasing the temperature drives configurations to be less ordered, i.e.; they bear less resemblance to the ground state. In the inset, we can see that $\langle \mathrm{MSE} \rangle$ changes between CAE runs, though this discrepancy seems to be smaller than the dependence on L. We also see that the increase of $\langle \mathrm{MSE} \rangle$ with T is not strictly monotonic: In the region near T_c, the curves become somewhat jagged, and this nonmonotonicity increases for $L = 130$. Similar jaggedness has been observed when using a quantity derived from the MSE in the 2D J_1-J_2 Ising model [16], so it may also be that this quantity

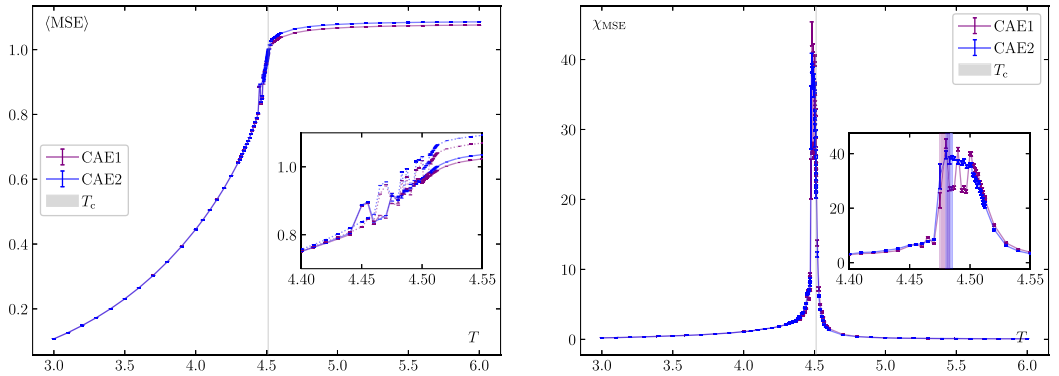


FIG. 1. MSE-derived quantities coming from the CAE. The vertical, gray band indicates the benchmark average $T_c = 4.511\,514(27)$. Left: Ensemble average of the MSE for our smallest ($L = 50$) and largest ($L = 130$) lattices from two runs of the CAE. Solid lines, drawn to guide the eye, indicate $L = 50$, while dashed lines indicate $L = 130$. Right: χ_{MSE} for $L = 50$ from two runs of the CAE. Vertical bands in the inset indicate the extracted $T_c(50)$ for each run.

is inherently quite sensitive to the medium-scale structure of configurations. Jaggedness near T_c could be a reflection of the confusion of the neural network near the boundary between two phases. Still, we find in all cases a drastic change in curvature near T_c, which is suggestive of the MSE feeling criticality.

Given the apparent sensitivity of $\langle \mathrm{MSE} \rangle$ to the phase change, we examine the susceptibility

$$\chi_{\mathrm{MSE}} = \frac{V}{T}(\langle \mathrm{MSE}^2 \rangle - \langle \mathrm{MSE} \rangle^2). \tag{8}$$

Results for χ_{MSE} at $L = 50$ are shown in Fig. 1 (right). We see a sharp peak near T_c for both CAE1 and CAE2, and in fact, we found such a peak structure for all lattice sizes. We use the maximum of this susceptibility, $\max(\chi_{\mathrm{MSE}})$, to define a pseudocritical temperature $T_c(L)$; namely, we locate $\max(\chi_{\mathrm{MSE}})$ and use the temperature immediately to its left, T_{lower}, and immediately to its right, T_{upper}, estimating

$$T_c(L) = \tfrac{1}{2}(T_{\mathrm{upper}} + T_{\mathrm{lower}}), \quad \sigma_{T_c(L)} = \tfrac{1}{2}(T_{\mathrm{upper}} - T_{\mathrm{lower}}). \tag{9}$$

Although $\langle \mathrm{MSE} \rangle$ and χ_{MSE} depend on the CAE run, we find that $T_c(L)$ estimated from CAE1 and CAE2 are compatible for all L. This is exemplified in the inset of Fig. 1 (right): The estimates from CAE1 and CAE2 overlap within uncertainty. While the peak location is found to be roughly independent of the CAE run, we found the peak height to depend on the run, a feature that can also be seen in the inset of Fig. 1.

We note that determining $T_c(L)$ in principle allows the $\langle \mathrm{MSE} \rangle$ data to be interpolated, yielding $\langle \mathrm{MSE}(T_c(L)) \rangle$, and from there a determination of the critical exponent β. In general, this exponent is not computed in ML studies of spin models [58]. For our data, the jaggedness of $\langle \mathrm{MSE} \rangle$ in the critical region prevents a trustworthy interpolation, and the dependence of $\langle \mathrm{MSE} \rangle$ on CAE run would lead to nonunique $\langle \mathrm{MSE}(T_c(L)) \rangle$ determinations. Therefore, our data cannot be used for a robust calculation of β. In contrast, as already noted, $T_c(L)$ itself is independent of the CAE run, and we are therefore able to use $T_c(L)$ to extract T_c and ν, which will be discussed shortly.

In Fig. 2 (left), we show χ_{MSE} coming from CAE1 for all lattice sizes. We see that as L increases, the peak shifts closer to T_c, another indication that the MSE can feel changes of phase, and is qualitatively suggestive that $T_c(L)$ will follow the finite-size scaling relation (4). The peak heights generically increase with increasing L, although this increase is not strictly monotonic, as would be expected from a true order parameter. One sees, for example, in Fig. 2 (left) that the peak height for $L = 110$ is greater than that of $L = 120$. The non-monotonicity even at large L indicates a violation

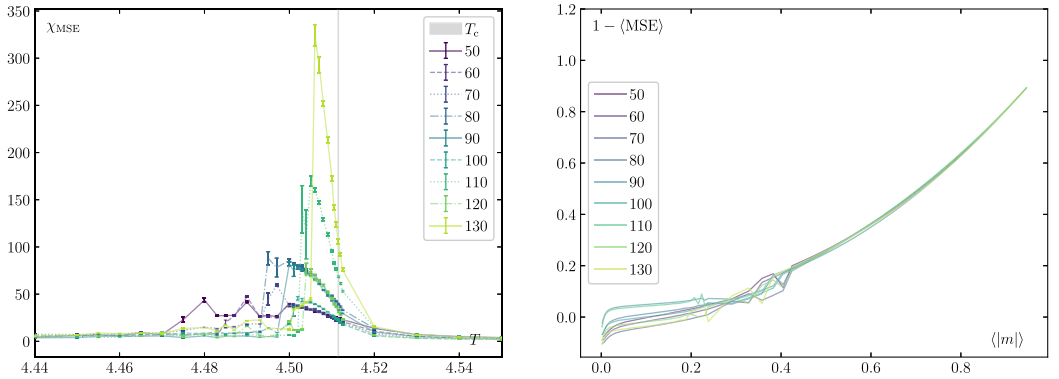


FIG. 2. Left: Dependence of χ_{MSE} vs T on L for CAE1. The vertical, gray band indicates T_c. Right: Correlation of $1-\langle\mathrm{MSE}\rangle$ with the order parameter for CAE1. Statistical uncertainties are suppressed for visibility.

of the scaling law expected for a true order parameter,

$$\chi_{\max} \sim L^{\gamma/\nu}, \tag{10}$$

and correspondingly, we found fits of $\max(\chi_{\mathrm{MSE}})$ to the above form to deliver unacceptable χ^2/d.o.f. This scaling law violation, along with the observed dependence of the peak height on the CAE starting seed mentioned earlier, precludes any reliable extraction of γ/ν using the MSE from our data.

In Fig. 2 (right), we plot $\langle\mathrm{MSE}\rangle$ against $\langle|m|\rangle$ for CAE1 for all lattice sizes. We see a clear correlation of $1-\langle\mathrm{MSE}\rangle$ with the order parameter, further reinforcing our interpretation that $\langle\mathrm{MSE}\rangle$ is sensitive to phase changes. We thus assume that $\langle\mathrm{MSE}\rangle$ can be viewed as an order parameter analog, at least to the extent that a meaningful $T_c(L)$ can be extracted from peaks in χ_{MSE}. Since $T_c(L)$ estimated from CAE1 and CAE2 are compatible, we focus the remainder of our analysis on CAE1.

Using the $T_c(L)$ extracted from the CAE, we extract T_c and ν using the finite-size scaling relation (4). In particular, we carry out a three-parameter fit of the form

$$T_c(L) = T_c + AL^{-1/\nu}. \tag{11}$$

The result of the fit is shown in Fig. 3. We find $T_c = 4.5128(58)$ and $\nu = 0.63(27)$ with χ^2/d.o.f. $= 1.16$. These results are in excellent agreement with the benchmark, albeit with substantially larger uncertainties. In particular, the result for ν has a relative uncertainty of 43%. Our results and comparison with benchmark values are given in Table I.

V. SUMMARY AND OUTLOOK

In this paper, we extended the application of a minimal training regimen from the 2D Ising to the 3D Ising model and from a supervised binary classifier to an unsupervised autoencoder. We find

TABLE I. Critical parameters of ML model compared against benchmark average over many literature results [26–53].

	Benchmark	CAE
T_c	4.511 514(27)	4.5128(58)
ν	0.630 36(39)	0.63(27)

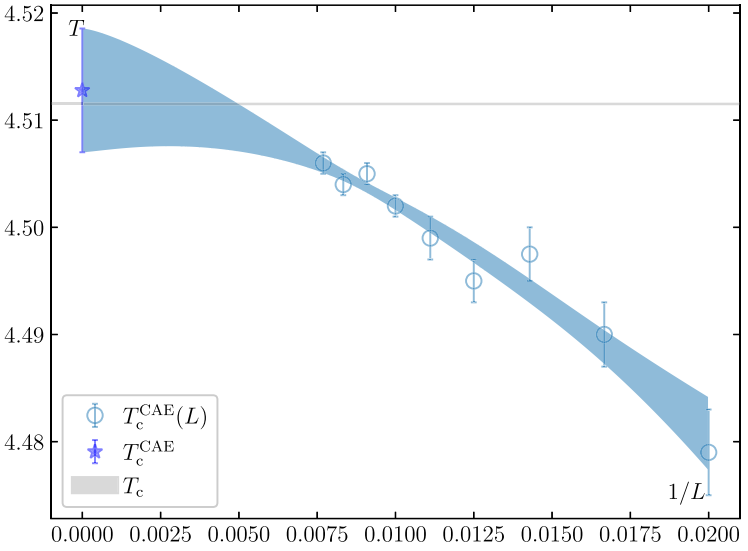


FIG. 3. Extraction of T_c and ν for the 3D Ising model using the MSE. The horizontal, gray band indicates our benchmark for T_c.

that the MSE is sensitive to a change of phase and correlates strongly with the magnetization. From its susceptibility, we are able to extract T_c and ν with good accuracy. Our results reinforce that one class is sufficient for finding phase transitions and demonstrate that a one-class autoencoder can recover nontrivial critical behavior directly from microscopic data.

For both the 2D and 3D Ising models, one- and two-class training regimens are capable of extracting T_c with a precision near the per-mille level. The critical exponent ν has a large relative uncertainty. The sensitivity of $\max(\chi_{\mathrm{MSE}})$ to the CAE along with the nonmonotonicity of $\max(\chi_{\mathrm{MSE}})$ with L prohibits an extraction of γ/ν; similarly, the sensitivity of the MSE to the CAE and its jaggedness in the critical region prohibits a determination of β. Taken with our previous work on two-class regimens [9], we conclude that one- and two-class regimens can be used to accurately and precisely extract critical temperatures in many statistical physics models, and they discover enough about the critical behavior of the system to provide at least rough estimates of critical exponents in some cases. The impressive ability of minimal training sets to find T_c, even when the system complexity is increased, increases our confidence that this approach will be useful for identifying phase boundaries for systems with crossovers or systems where the order parameter is not known *a priori*.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grants No. PHY-2208724, No. PHY-2116686, and No. PHY-2514763, and within the framework of the MUSES collaboration, under Grant No. OAC-2103680. This material is also based upon work supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics, under Award No. DE-SC0022023, as well as by the National Aeronautics and Space Agency (NASA) under Award No. 80NSSC24K0767. The authors acknowledge the use of the Carya Cluster and the advanced support from the Research Computing Data Core at the University of Houston to carry out the research presented here.

DATA AVAILABILITY

The data that support the findings of this article are openly available [59].

APPENDIX A: CAE ARCHITECTURE

The ML model applied in this study is a one-class convolutional autoencoder written in Python with Tensorflow [60] and Keras [61] libraries. Its architecture is illustrated in Fig. 4. The network consists of a 3D convolutional encoder followed by a decoder that progressively restores the original spatial resolution. The network operates directly on spin configurations of shape $L \times L \times L \times 1$. The encoder progressively reduces the spatial resolution of the input using strided 3D convolutions.

Two convolutional layers with $3 \times 3 \times 3$ kernels and 16 channels are first applied at full resolution, followed by a stride-2 convolution that downsamples the lattice by a factor of 2. A second pair of convolutional layers with 32 channels is then applied, followed by another stride-2 convolution that reduces the spatial resolution by an additional factor of 2. All convolutional layers use LeakyReLU activations [62]. A dropout layer with rate 0.2 is applied at the bottleneck, after which a final 64-channel convolution produces the latent feature representation used by the decoder. The decoder restores the original lattice resolution through a sequence of learned 3D transposed-convolution (Conv3DTranspose) layers. Starting from the latent representation, successive layers increase the lattice resolution by factors of 2 until the original lattice size is recovered. A final $1 \times 1 \times 1$ convolution with linear activation maps the reconstructed feature channels at each lattice site to a single spin value. The output therefore has spatial size $L \times L \times L$. To ensure consistency of the output size, the reconstruction is cropped to $L \times L \times L$ when necessary.

This fully convolutional architecture preserves spatial locality and translational structure, making it well suited for learning the characteristic patterns of the ordered phase directly from microscopic configurations.

APPENDIX B: GLOSSARY OF MACHINE-LEARNING TERMS

Autoencoder (AE). A neural network trained to reproduce its input, comprising two modules: an encoder module that learns how to map its input into a compressed representation, followed by a decoder module that learns how to recover the original input.

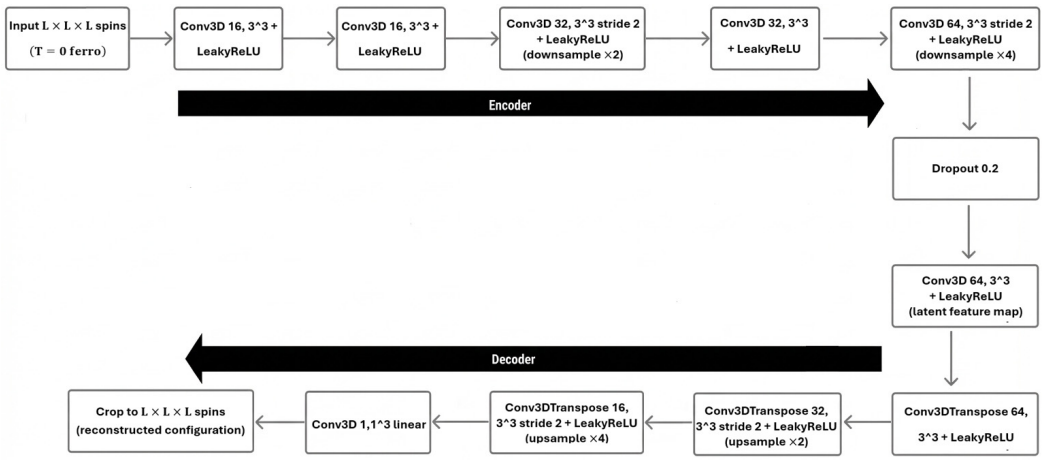


FIG. 4. Architecture of the 3D CAE used in this work.

Reconstruction error. A measure of the difference between an input configuration and the autoencoder's reconstruction of that configuration. In this work, the reconstruction error is quantified using the mean-square error (MSE).

Latent representation. The compressed representation produced by the encoder module of an autoencoder.

Convolutional autoencoder (CAE). An autoencoder whose encoder and decoder modules both comprise convolutional layers that progressively reduce (resp. increase) the spatial resolution of the input representation while preserving spatial locality and translational structure.

Convolutional layer. The core building block of a CAE, utilized sequentially within the encoder (resp. decoder) modules to map local, translation invariant spatial patterns into a representation at lower (resp. higher) resolution.

Kernel/filter. A local, translation invariant spatial pattern learned by a convolutional layer, whose spatial extent is referred to as its size.

Stride. The step size used to translate the learned local pattern (kernel/filter) across the full representation of a configuration. Step sizes greater than 1 are used to reduce spatial resolution (downsampling) in the encoder module and to increase spatial resolution (upsampling) in the decoder module.

Channel (number of). The number of kernels/filters learned by a convolutional layer.

Transpose convolution. A type of kernel/filter used in the convolutional layers of a decoder module to map local, translation-invariant spatial patterns to a higher-resolution representation.

Activation function. A nonlinear function applied to the output of a neural network's layer to facilitate learning of more complex patterns.

Dropout. A regularization technique in which a fraction of neurons is randomly omitted during training. This discourages the CAE from relying excessively on any single pattern and helps reduce overfitting.

[1] J. Carrasquilla and R. G. Melko, Machine learning phases of matter, Nat. Phys. **13**, 431 (2017).
[2] C.-D. Li, D.-R. Tan, and F.-J. Jiang, Applications of neural networks to the studies of phase transitions of two-dimensional Potts models, Ann. Phys. **391**, 312 (2018).
[3] G. Cossu, L. Del Debbio, T. Giani, A. Khamseh, and M. Wilson, Machine learning determination of dynamical parameters: The Ising model case, Phys. Rev. B **100**, 064304 (2019).
[4] S. S. Funai and D. Giataganas, Thermodynamics and feature extraction by machine learning, Phys. Rev. Res. **2**, 033415 (2020).
[5] D. R. Tan, C. D. Li, W. P. Zhu, and F. J. Jiang, A comprehensive neural networks study of the phase transitions of Potts model, New J. Phys. **22**, 063016 (2020).
[6] D. Bachtis, G. Aarts, and B. Lucini, Extending machine learning classification capabilities with histogram reweighting, Phys. Rev. E **102**, 033303 (2020).
[7] N. Walker, K.-M. Tam, and M. Jarrell, Deep learning on the 2-dimensional Ising model to extract the crossover region with a variational autoencoder, Sci. Rep. **10**, 13047 (2020).
[8] S.-W. Li, Y.-H. Tseng, M.-C. Hsieh, and F.-J. Jiang, Minimal training set for training a successful CNN: A case study of the frustrated J_1-J_2 Ising model on the square lattice, Prog. Theor. Exp. Phys. **2025**, 113A02 (2025).
[9] A. Abuali, D. A. Clarke, M. Hjorth-Jensen, I. Konstantinidis, C. Ratti, and J. Yang, Deep learning of phase transitions with minimal examples, Phys. Rev. E **112**, 035315 (2025).
[10] L. Wang, Discovering phase transitions with unsupervised learning, Phys. Rev. B **94**, 195105 (2016).
[11] 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).

[12] S. J. Wetzel, Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders, Phys. Rev. E **96**, 022140 (2017).
[13] J. F. Rodriguez-Nieva and M. S. Scheurer, Identifying topological order through unsupervised machine learning, Nat. Phys. **15**, 790 (2019).
[14] C. Alexandrou, A. Athenodorou, C. Chrysostomou, and S. Paul, The critical temperature of the 2D-Ising model through deep learning autoencoders, Eur. Phys. J. B **93**, 226 (2020).
[15] X.-Q. Han, S.-S. Xu, Z. Feng, R.-Q. He, and Z.-Y. Lu, Framework for contrastive learning phases of matter based on visual representations, Chin. Phys. Lett. **40**, 027501 (2023).
[16] B. Çivitcioğlu, R. A. Römer, and A. Honecker, Phase determination with and without deep learning, Phys. Rev. E **111**, 024131 (2025).
[17] I. Jang and A. Yethiraj, Unsupervised machine learning method for the phase behavior of the constant magnetization Ising model in two and three dimensions, J. Phys. Chem. B **129**, 532 (2025).
[18] G. Aarts *et al.*, Phase transitions in particle physics: Results and perspectives from lattice quantum chromo-dynamics, Prog. Part. Nucl. Phys. **133**, 104070 (2023).
[19] D. Kim and D.-H. Kim, Smallest neural network to learn the Ising criticality, Phys. Rev. E **98**, 022138 (2018).
[20] K. Ch'ng, N. Vazquez, and E. Khatami, Unsupervised machine learning account of magnetic transitions in the Hubbard model, Phys. Rev. E **97**, 013306 (2018).
[21] 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).
[22] X. Li, R. Guo, Y. Zhou, K. Liu, J. Zhao, F. Long, Y. Wu, and Z. Li, Machine learning phase transitions of the three-dimensional Ising universality class, Chin. Phys. C **47**, 034101 (2023).
[23] R. Guo, X. Li, R. Wang, S. Chen, Y. Wu, and Z. Li, Exploring percolation phase transition in the three-dimensional Ising model with machine learning, Chin. Phys. C **49**, 054103 (2025).
[24] R. K. Panda, R. Verdel, A. Rodriguez, H. Sun, G. Bianconi, and M. Dalmonte, Non-parametric learning critical behavior in Ising partition functions: PCA entropy and intrinsic dimension, SciPost Phys. Core **6**, 086 (2023).
[25] L. Onsager, Crystal statistics. 1. A two-dimensional model with an order disorder transition, Phys. Rev. **65**, 117 (1944).
[26] J. Adler, Critical temperatures of the $d = 3$, $s = 1/2$ Ising model; the effect of confluent corrections to scaling, J. Phys. A: Math. Gen. **16**, 3585 (1983).
[27] G. S. Pawley, R. H. Swendsen, D. J. Wallace, and K. G. Wilson, Monte Carlo renormalization group calculations of critical behavior in the simple cubic Ising model, Phys. Rev. B **29**, 4030 (1984).
[28] M. N. Barber, R. B. Pearson, D. Toussaint, and J. L. Richardson, Finite size scaling in three-dimensional Ising model, Phys. Rev. B **32**, 1720 (1985).
[29] F. Livet, The cluster updating Monte Carlo algorithm applied to the 3-d Ising problem, Europhys. Lett. **16**, 139 (1991).
[30] N. Ito and M. Suzuki, Monte Carlo study of the spontaneous magnetization of the three-dimensional Ising model, J. Phys. Soc. Jpn. **60**, 1978 (1991).
[31] H. Blöte and G. Kamieniarz, Finite-size calculations on the three-dimensional Ising model, Physica A **196**, 455 (1993).
[32] D. Landau, Computer simulation studies of critical phenomena, Physica A **205**, 41 (1994).
[33] H. W. J. Blote, E. Luijten, and J. R. Heringa, Ising universality in three dimensions: A Monte Carlo study, J. Phys. A: Math. Gen. **28**, 6289 (1995).
[34] A. L. Talapov and H. W. J. Blote, The magnetization of the 3-D Ising model, J. Phys. A: Math. Gen. **29**, 5727 (1996).
[35] R. Gupta and P. Tamayo, Critical exponents of the 3-D Ising model, Int. J. Mod. Phys. C **07**, 305 (1996).
[36] J.-K. Kim, A. J. F. de Souza, and D. P. Landau, Numerical computation of finite size scaling functions: An alternative approach to finite size scaling, Phys. Rev. E **54**, 2291 (1996).
[37] P. Butera and M. Comi, N vector spin models on the sc and the bcc lattices: A study of the critical behavior of the susceptibility and of the correlation length by high temperature series extended to order beta(21), Phys. Rev. B **56**, 8212 (1997).

[38] P. Butera and M. Comi, Renormalized couplings and scaling correction amplitudes in the N vector spin models on the sc and the bcc lattices, Phys. Rev. B **58**, 11552 (1998).
[39] Z. Salman and J. Adler, High and low temperature series estimates for the critical temperature of the 3-D Ising model, Int. J. Mod. Phys. C **09**, 195 (1998).
[40] H. W. J. Blote, L. N. Shchur, and A. L. Talapov, The cluster processor: New results, Int. J. Mod. Phys. C **10**, 1137 (1999).
[41] M. Hasenbusch, Finite size scaling study of lattice models in the three-dimensional Ising universality class, Phys. Rev. B **82**, 174433 (2010).
[42] B. Nickel, Hyperscaling and universality in 3 dimensions, Physica A **106**, 48 (1981).
[43] B. G. Nickel and J. J. Rehr, High-temperature series for scalar-field lattice models: Generation and analysis, J. Stat. Phys. **61**, 1 (1990).
[44] M. Hasenbusch and K. Pinn, Computing the roughening transition of Ising and solid-on-solid models by BCSOS model matching, J. Phys. A: Math. Gen. **30**, 63 (1997).
[45] M. Hasenbusch, K. Pinn, and S. Vinti, Critical exponents of the three-dimensional Ising universality class from finite-size scaling with standard and improved actions, Phys. Rev. B **59**, 11471 (1999).
[46] H. G. Ballesteros, L. A. Fernandez, V. Martin-Mayor, A. Munoz-Sudupe, G. Parisi, and J. J. Ruiz-Lorenzo, Scaling corrections: Site percolation and Ising model in three-dimensions, J. Phys. A: Math. Gen. **32**, 1 (1999).
[47] M. Campostrini, A. Pelissetto, P. Rossi, and E. Vicari, Improved high temperature expansion and critical equation of state of three-dimensional Ising-like systems, Phys. Rev. E **60**, 3526 (1999).
[48] M. Hasenbusch, A Monte Carlo study of leading order scaling corrections of ϕ^4 theory on a three-dimensional lattice, J. Phys. A: Math. Gen. **32**, 4851 (1999).
[49] M. Campostrini, A. Pelissetto, P. Rossi, and E. Vicari, 25th order high temperature expansion results for three-dimensional Ising like systems on the simple cubic lattice, Phys. Rev. E **65**, 066127 (2002).
[50] S. El-Showk, M. F. Paulos, D. Poland, S. Rychkov, D. Simmons-Duffin, and A. Vichi, Solving the 3d Ising model with the conformal bootstrap II. C-minimization and precise critical exponents, J. Stat. Phys. **157**, 869 (2014).
[51] 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.
[52] C. Bonanno, A. Cappelli, M. Kompaniets, S. Okuda, and K. J. Wiese, Benchmarking the Ising universality class in $3 \leqslant d < 4$ dimensions, SciPost Phys. **14**, 135 (2023).
[53] C.-H. Chang, V. Dommes, R. S. Erramilli, A. Homrich, P. Kravchuk, A. Liu, M. S. Mitchell, D. Poland, and D. Simmons-Duffin, Bootstrapping the 3d Ising stress tensor, J. High Energy Phys. 03 (2025) 136.
[54] These averages are not meant to be exhaustive: Our goal is to have high-precision results to benchmark against that incorporate recent papers.
[55] D. A. Clarke, L. Altenkort, J. Goswami, and H. Sandmeyer, Streamlined data analysis in Python, PoS **LATTICE2023**, 136 (2024).
[56] LatticeQCD, AnalysisToolbox, GitHub, 2026, https://github.com/LatticeQCD/AnalysisToolbox.
[57] J. Wojtkiewicz and K. Kalinowski, Monte Carlo simulations of the Ising model on GPU, CMST **21**, 69 (2015).
[58] In fact, depending on the ML observable used as pseudo-order parameter/order parameter analog, β may not be determinable even in principle. For example, observables that quantify the probability of belonging to a particular phase cannot be used to extract β [9].
[59] CAE Ising3D public code repository, https://github.com/clarkedavida/CAE_Ising3D.
[60] M. Abadi *et al.*, TensorFlow: Large-scale machine learning on heterogeneous systems, 2015, software available from https://www.tensorflow.org.
[61] F. Chollet *et al.*, Keras, 2015, https://keras.io.
[62] Compared to standard ReLU, LeakyReLU maintains a nonzero gradient for negative inputs, which improves training stability and preserves sensitivity to both signs of the spin field. This is particularly important for autoencoders operating on ± 1 spin configurations and for reconstruction-based anomaly detection.