An equation-of-state-meter of quantum chromodynamics transition from deep learning

Long-Gang Pang, Kai Zhou, Nan Su, Hannah Petersen, Horst Stöcker, & Xin-Nian Wang

A primordial state of matter consisting of free quarks and gluons that existed in the early universe a few microseconds after the Big Bang is also expected to form in high-energy heavy-ion collisions. Determining the equation of state (EoS) of such a primordial matter is the ultimate goal of high-energy heavy-ion experiments. Here we use supervised learning with a deep convolutional neural network to identify the EoS employed in the relativistic hydrodynamic simulations of heavy ion collisions. High-level correlations of particle spectra in transverse momentum and azimuthal angle learned by the network act as an effective EoS-meter in deciphering the nature of the phase transition in quantum chromodynamics. Such EoS-meter is model-independent and insensitive to other simulation inputs including the initial conditions for hydrodynamic simulations.
Deep learning (DL) is a branch of machine learning that learns multiple levels of representations from data\textsuperscript{15}. DL has been successfully applied in pattern recognition and classification tasks, such as image recognition and language processing. Recently, the application of DL to physics research is rapidly growing, such as in particle physics\textsuperscript{17-23}, nuclear physics\textsuperscript{24}, and condensed matter physics\textsuperscript{9-14}. DL is shown to be very powerful in extracting pertinent features especially for complex non-linear systems with high-order correlations that conventional techniques are unable to tackle. This suggests that it could be utilized to unveil hidden information from the highly implicit data of heavy-ion experiments.

Strong interaction in nuclear matter is governed by the theory of quantum chromodynamics (QCD). It predicts a transition from the normal nuclear matter, in which the more fundamental constituents, quarks and gluons, are confined within the domains of nucleons, to a new form of matter with freely roaming quarks and gluons as one increases the temperature or density. The QCD transition is conjectured to be a crossover at small density (and lower temperature), with a critical point separating the two, see Fig. 1 for a schematic QCD phase diagram and\textsuperscript{15-17} for some reviews. One primary goal of ultra-relativistic heavy-ion collisions is to study the QCD transition.

Though it is believed that strongly coupled QCD matter can be formed in heavy-ion collisions at the Relativistic Heavy Ion Collider (RHIC, Brookhaven National Laboratory, USA)\textsuperscript{18}, Large Hadron Collider (LHC, European Organization for Nuclear Research, Switzerland)\textsuperscript{19}, and at the forthcoming Facility for Anti-proton and Ion Research (FAIR, GSI Helmholtz Centre for Heavy Ion Research, Germany)\textsuperscript{20,21}, a direct access to the bulk properties of the matter such as the equation of state (EoS) and transport coefficients is impossible due to the highly dynamical nature of the collisions. In heavy-ion collisions where two high-energy nuclei collide along the longitudinal ($z$) direction, what experiments measure directly are the final-state particle distributions in longitudinal momentum (rapidity), transverse momentum $p_{\text{T}}$ and azimuthal angle $\phi$.

Current efforts to extract physical properties of the QCD matter from experimental data are through direct comparisons with model calculations of event-averaged and predefined observables, such as anisotropic flow\textsuperscript{22} or global fitting of a set of observables with Bayesian method\textsuperscript{23,24}. However, event-by-event raw data on $\rho(p_{\text{T}}, \phi)$ at different rapidities provide much more information that contains hidden correlations. These hidden correlations can be sensitive to physical properties of the system but independent of other model parameters.

The aim of the present exploratory study is a first step in directly connecting QCD bulk properties and raw data of heavy-ion collisions using state-of-the-art deep-learning techniques. We use the relativistic hydrodynamic model which has been very successful in simulating heavy-ion collisions and connecting experiments with theory\textsuperscript{25-29}. We find unique encoders of bulk properties (here we focus on the EoS) inside $\rho(p_{\text{T}}, \phi)$ in terms of high-level representations using deep-learning techniques, which are not captured by conventional observables. This is achieved by constructing a convolutional neural network (CNN) and training it with labeled $\rho(p_{\text{T}}, \phi)$ of charged pions generated from the relativistic hydrodynamic program CLVisc\textsuperscript{30,31} with two different EoSs as input: crossover\textsuperscript{32} and first order\textsuperscript{33}. The CNN is then trained with supervision in identifying different EoSs. The performance is surprisingly robust against other simulation parameters such as the initial conditions, equilibrium time $\tau_0$, transport coefficients and freeze-out temperature. The supervised learning with deep CNN identifies the hydrodynamic response which is much more tolerant to uncertainties in the initial conditions, $\rho(p_{\text{T}}, \phi)$ as generated by independent simulations (CLVisc with different setup parameters and another hydrodynamic package iEBE-VISHNU\textsuperscript{34} which implements a different numerical solver for partial differential equations) are used for testing—on average a larger than 95% testing accuracy is obtained. It has been recently pointed out that model-dependent features (features in the training data that depends on the simulation model and parameters) may generate large uncertainties in the network performance\textsuperscript{6}. The network we develop below is, however, not sensitive to these model-dependent features.

### Results

#### Training and testing data sets

The evolution of strongly coupled QCD matter can be well described by second-order dissipative hydrodynamics governed by $\partial_\mu T^\mu = 0$, with $T^\mu_\nu$ the energy–momentum tensor containing viscous corrections governed by the Israel–Stewart equations\textsuperscript{25,26}. In order to close the hydrodynamic equations, one must supply the EoS of the medium as one crucial input. The nature of the QCD transition in the EoS strongly affects the hydrodynamic evolution\textsuperscript{35}, since different transitions are associated with different pressure gradients which consequently induce different expansion rates, see the small chart in Fig. 1. Final $\rho(p_{\text{T}}, \phi)$ are obtained from the Cooper–Frye formula for particle $i$ at mid-rapidity

$$
\rho(p_{\text{T}}, \phi) = \frac{dN_i}{d^3p_\text{T}d\phi} = g_i \int P_{\text{T}} d\sigma_{\rho_i} f_i, \tag{1}
$$

Here $N_i$ is the particle number density, $Y$ is the rapidity, $g_i$ is the degeneracy, $\sigma_{\rho_i}$ is the freeze-out hypersurface element, $f_i$ is the

| Table 1: The training data set |
|-----------------------------|-----------------------------|
| Training data set           | $\eta/s = 0$ | $\eta/s = 0.08$ |
|                            | EOSL | EOSQ | EOSL | EOSQ |
| Au–Au $\sqrt{s_{\text{NN}}} = 200$ GeV | 7435 | 5328 | 500 | 500 |
| Pb–Pb $\sqrt{s_{\text{NN}}} = 2.76$ TeV | 4967 | 2827 | 500 | 500 |

Numbers of $\rho(p_{\text{T}}, \phi)$ generated by the CLVisc hydrodynamic package with the AMPT initial conditions in the centrality range 0–60%. $\eta/s$ is ratio of shear viscosity to entropy density. $\tau_0 = 0.4$ fm for the Au–Au collisions and $\tau_0 = 0.2$ fm for the Pb–Pb collisions. The freeze-out temperature is set to 137 MeV.
The existence of physical encoders and neural-network decoder.

After training and validating the network, it is tested on the testing data set of \( \rho(p_T, \phi) \) events (see Sec. 4 for the details of our neural-network model). As shown in Table 3, high prediction accuracies—on average larger than 95% with small model uncertainties given by a 10-fold cross validation tests—are achieved for these three groups of testing data sets, which indicates that our method is highly independent of initial conditions. The network is robust against shear viscosity and \( \tau_0 \) due to the inclusion of events with different \( n/s \) and \( \tau_0 \) in the training. In the testing stage the neural network identifies the type of the QCD transition solely from the spectra of each single event. Furthermore, in the training only one freeze-out temperature is used, while the network is tolerant to a wide range of freeze-out temperatures during the testing. For simplicity, the exploratory study has not included pions from resonance decays (the hadronic transport module UrQMD is switched off in iEBE-VISHNU to exclude contributions from resonance decays in testing data).

For complex and dynamically evolving systems, the final states may not contain enough information to retrieve the physical properties of initial and intermediate states due to entropy production (information loss) during the evolution. The mean prediction accuracy decreases from 97.1% (for \( n/s = 0.0 \)) to 96.6% (for \( n/s = 0.08 \)) and 87% (for \( n/s = 0.16 \)) in the 10-fold cross validation for testing GROUP 1. Besides, the construction of conventional observables may introduce further information loss due to projection of raw data to lower dimensions, as well as information interference due to its sensitivity to multiple factors. These make it yet unclear how to reliably extract physical properties from raw data. Our study firmly demonstrates how to detect the existence of physical encoders in final states with deep CNN decoders, and sets the stage for further applications, such as identifying all relevant physical properties of the systems.

**Observation from the neural-network decoder.** In order to get physical insights from the neural-network model, it is instructive to visualize the complex dependences learned by the network. For this purpose, we employ the recently developed Prediction Difference Analysis method. This method uses the observation that replacing one feature in the input image can induce a sizable prediction difference if that feature is important for classification. The prediction differences can be visualized as the importance maps of all the input features for the classification network.

Shown in Fig. 2 are importance maps which illustrate the \( (p_T, \phi) \) dependence of the mean prediction difference averaged over 800 events for different model setups (initial conditions, PDE solver and model parameters), EoSs and values of the shear viscosity. For a given event, the mean prediction difference in each \( (p_T, \phi) \) bin is computed against ten random reference events from the same data set. Comparing different columns in the same row in Fig. 2, we can see that important features differ slightly for different values of viscosity and model setups (Run 1: IEBE-VISHNU + MC-Glauber, Run 2: CLVisc + IP-Glasma) for the same EoS. However, importance maps for EOS in general have a distinctly narrower width in the \( p_T \) range than that for EOSQ, independently of the model setup and the value of viscosity.

This might be the important region of hidden features the network recognizes in classifying the EoS under each event.

**Discussion**

Besides the deep CNN method employed in the present paper, there are also some other machine learning classifiers. In Supplementary Note 2 we attached the results from several traditional

---

### Table 2 The testing data set

| Centrality: 10–60% | \( n/s \in [0, 0.05] \) | \( n/s \in (0.05, 0.10) \) | \( n/s \in (0.10, 0.16) \) |
|-------------------|-------------------|-------------------|-------------------|
| Au-Au \( \sqrt{S_{NN}} = 200 \text{ GeV} \) | EOSL 650 EOSQ 850 | EOSL 900 EOSQ 750 | EOSL 660 EOSQ 644 |
| Pb-Pb \( \sqrt{S_{NN}} = 2.76 \text{ TeV} \) | 500 650 | 600 644 | 499 150 |
| Testing data set group 2: CLVisc + IP-Glasma Au-Au \( \sqrt{S_{NN}} = 200 \text{ GeV} \), \( b \leq 8 \text{ fm} \) | EOSL 4164 EOSQ 4752 | EOSL 1173 EOSQ 864 |
| \( n/s = 0 \) | | | |
| \( n/s = 0.08 \) | | | |

Numbers of \( \rho(p_T, \phi) \) generated by the CLVisc and (IEBE-VISHNU) hydrodynamic packages with different initial conditions. \( n/s \) is ratio of shear viscosity and entropy density, \( b \) is the impact parameter, \( \tau_0 = 0.06 \text{ fm} \) for all the collisions. In iEBE-VISHNU simulations, the freeze-out temperature is varied in the range \([115, 142]\text{MeV}\). In CLVisc simulations, the freeze-out temperature is set to be \(137 \text{MeV}\).

### Table 3 Testing accuracies

| Testing data | Group 0 | Group 1 | Group 2 |
|--------------|---------|---------|---------|
| Number of events | 4000 | 7343 | 10,953 |
| Accuracy | 99.88 ± 0.04% | 93.46 ± 1.35% | 93.91 ± 3.92% |

The mean prediction accuracies and the standard deviations given by ten trained models in cross validation method, from three groups of testing data sets, (GROUP 0) CLVisc with AMPT initial condition, (GROUP 1) IEBE-VISHNU and (GROUP 2) CLVisc with the IP-Glasma-like initial condition.
CNN provides a powerful and efficient decoder from which the EoS information can be extracted directly from the $\rho(p_T, \phi)$. It is in this sense that the high-level representations, which help decoding the EoS information in the present method, act as an EoS-meter for the QCD matter created in heavy-ion collisions. The Prediction Difference Analysis method is employed to extract the most relevant features for the classification task, which may inspire phenomenological and experimental studies. Our study might provide a key to the success of the experimental determination of QCD EoS and search for the critical end point. Another intriguing application of our framework is to extract the QGP transport coefficients from heavy-ion collisions. The present method can be further improved by including hadronic rescattering and detector efficiency corrections.

**Methods**

**Network architecture.** The decisive ingredients for the success of hydrodynamic modeling of relativistic heavy-ion collisions are the bulk-matter EoS and the viscosity. In the study of the QCD transition in heavy-ion collisions, one of the holy-grail questions is how to reliably extract EoS and the nature of the QCD transition from the experimental data? The CNN\(^{11,12}\) is a powerful technique in tasks such as image and video recognition, natural language processing, Supervised training of the CNN with labeled data $\rho(p_T, \phi)$ generated by CLVisc is tested batch $\rho(p_T, \phi)$ generated by iEBE-VISHNU. The training and testing $\rho(p_T, \phi)$ can be regarded as numerical experimental data. Hence, analyzing real experimental data is possible with straightforward generalizations of the current prototype setup.

Our CNN architecture is shown in Fig. 3. The input $\rho(p_T, \phi)$ consists of 15 $p_T$ bins and 48 $\phi$ bins. We use two convolutional layers each followed by batch normalization\(^{43}\), dropout\(^{44,45}\) with a rate 0.2 and PReLU activation\(^{46}\). These technical terms are briefly explained in Supplementary Note 1. In the first convolutional layer, there are 16 filters of size $8 \times 8$ scanning through the input $\rho(p_T, \phi)$ and creating 16 features of size $15 \times 48$. These features are further convoluted in the second convolutional layer that has 32 filters of size $7 \times 7 \times 16$. The weight matrix of both convolutional layers is initialized with normal distribution and constrained with $L_2$ regularization\(^{47}\). In a convolutional layer, each neuron only locally connects to a small chunk of neurons in the previous layer by a convolution operation—this is a key reason for the success of the CNN architecture. Dropout, batch normalization, PReLU, and $L_2$ regularization work together to prevent overfitting that may generate model-dependent features from the training data set and thus hinder the generalizability of the method. The resulting 32 features of size $8 \times 24$ from the second convolutional layer are flattened and connected to a 128-neuron fully connected layer with batch normalization, dropout with rate 0.5 and sigmoid activation. The output layer is another fully connected layer with softmax activation and two neurons to indicate the type of the EoS. For multi-class classification, one may use more neurons in the output layer.

There are several non-trainable parameters in the neural network, such as the number of hidden layers, the size of the convolution kernels, the size of the convolution kernels, the size of the hidden layer and the dropout rate. The neural network in the present work can be easily rebuilt with these hyper-parameters in Keras\(^{48}\) (the source code is also available as requested). These parameters are adjusted heuristically to maximize the

![Figure 2](image1.png)  
**Fig. 2** Importance maps of the particle momentum distribution. The values in the scale bar represent the relative importance of each bin for classification computed using the Prediction Difference Analysis method by averaging over about 800 events for each category. EoSQ in the first row represents the equation of state with a smooth crossover, EOSL in the second row represents a first-order phase transition equation of state. The G1 in a–d represents the testing data set Group 1 from iEBE-VISHNU model while G2 in e–h represents the testing data set from CLVisc + IP-Glasma model. The $\eta/s = 0$ in a, b, e, f represents ideal hydrodynamics while $\eta/s = 0.08$ in c, d, g, h represents viscous hydrodynamics with shear viscosity over entropy density ratio 0.08

![Figure 3](image2.png)  
**Fig. 3** The convolutional neural network architecture. The architecture is designed to identify the quantum chromodynamics transition by using particle spectra with 15 transverse momentum $p_T$ bins and 48 azimuthal angle $\phi$ bins

machine learning methods, such as support vector machine classifier (SVC), decision trees, random forests and gradient boosting trees. The best classifier (linear SVC) that generalizes well on two testing data sets achieves on average ~80% prediction accuracy. The important features from different classifiers differ from each other, however, those with good generalization capability have similar importance regions as given by the deep CNN. The deep CNN with on average ~95% prediction accuracy works much better to answer the core questions—is there a traceable encoder of the dynamical information from phase structure (EoS) that survives the evolution and exists in the final snapshot? If yes, then how to exclusively and effectively decode these information from the highly complex final output? These questions are crucial but unclear for decades in high-energy heavy-ion physics (and also in physical cosmology) due to the complexity and highly-dynamical characteristics in the collision evolution. The deep CNN demonstrates the revolution that big data analysis and machine learning might bring to the high energy physics and astrophysics.

The present method yields a perspective on identifying the nature of the QCD transition in heavy-ion collisions. With the help of deep CNNs and its well generalization performance, we firmly demonstrate that discriminative and traceable projections—encoders—from the QCD transition onto the final-state $\rho(p_T, \phi)$ do exist in the complex and highly dynamical heavy-ion collisions, although these encoders may not be intuitive. The deep
training accuracy and validation accuracy but not the testing accuracy. The first step is to choose the number of hidden layers, the size of the convolution kernels and the size of the final hidden layer such that the model has enough capacity to describe the training data. At this step, we use a small portion of the training data, tune the widely used values of parameters and observe big training accuracy but small validation accuracy. It is found that the widely used convolution kernel sizes $5 \times 5$ and $3 \times 3$ do not work well at this step and increasing the number of the convolution layers from 2 to 3 does not improve the training accuracy and the validation accuracy. The next step is to increase the validation accuracy, in addition to the batch normalization and L2 regularization, it is found that dropout with a proper rate and tuning the size of the final hidden layer help to increase the validation accuracy. With this minimal working neural network, the validation accuracy increases rapidly with more training data. What is interesting is that when there are big training data, the previously not functioning architectures (with smaller convolution kernels and more hidden layers) also start to work and produces similar testing accuracy. The optimal neural network architecture and the values of the non-trainable parameters with big training data may desire future investigation.

**Training and validation.** We use supervised learning to tackle this binary classification problem with the crossover case labeled by $(1, 0)$ and the first-order case labeled by $(0, 1)$. The difference between the true label and the predicted label from $f_\theta$ produces similar testing accuracy. The optimal neural network architecture and the size of the training data are big enough; we experimented with fully connected neural networks. However, the network always overfits the training data and fails to work with the validating data. We noticed that CNN has much better generalizability than fully connected neural network with small set of data. With 22,000 events, the best performance of fully connected neural networks, with 2–5 hidden layers, gave on average 90% recognition rate on the testing data. Data augmentation in fully connected neural networks bring negligible improvement (less than 1%) on the testing data. The fully connected neural networks neglect the translation invariance of the local correlations of particles that are close to each other in momentum space.

**Data availability.** The data sets generated and analyzed during the current study are available in the public repository\cite{figshare}, https://doi.org/10.6084/m9.figshare.5457220.v1.

Received: 20 March 2017 Accepted: 20 December 2017 Published online: 15 January 2018

**References**

1. Schmidhuber, J. Deep learning in neural networks: an overview. Neural Netw. 61, 55–117 (2015).
2. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature 521, 436–444 (2015).
3. Baldi, P., Sadowski, P. & Whitson, D. Searching for exotic particles in high-energy physics with deep learning. Nat. Commun. 5, 4308 (2014).
4. Baldi, P., Sadowski, P. & Whitson, D. Enhanced Higgs Boson to $\tau^+\tau^-$ search with deep learning. Phys. Rev. Lett. 114, 111801 (2015).
5. Searcy, J., Huang, L., Pleier, M. A. & Zhu, J. Determination of the WW polarization fractions in pp $\rightarrow W^+W^-$ using a deep machine learning technique. Phys. Rev. D 93, 094033 (2016).
6. Barnard, J., Dawe, E. N., Dolan, M. J. & Rajic, N. Parton shower uncertainties in jet substructure analyses with deep neural networks. Phys. Rev. D 95, 014018 (2017).
7. Mousli, L. Necib, L. & Talier, J. New angles on energy correlation functions. J. High Energy Phys. 12, 153 (2016).
8. Utama, R., Chen, W. C. & Piekarewicz, J. Nuclear charge radii: density functional theory meets Bayesian neural networks. J. Phys. G 43, 114002 (2016).
9. Mehta, P. & Schwab, D. J. An exact mapping between the Variational Renormalization Group and deep learning. Preprint at https://arxiv.org/abs/1410.3831 (2014).
10. Carrasquilla, J. & Melko, R. G. Machine learning phases of matter. Nat. Phys. 13, 431–434 (2017).
11. Carleo, G. & Troyer, M. Solving the quantum many-body problem with artificial neural networks. Science 355, 602–606 (2017).
12. Torlai, G. & Melko, R. G. Learning thermodynamics with boltzmann machines. Phys. Rev. B 94, 165134 (2016).
13. Broecker, P., Carrasquilla, J., Melko, R. G. & Trebst, S. Machine learning quantum phases of matter beyond the fermion sign problem. Sci. Rep. 7, 8823 (2017).
14. Ch’ng, K., Carrasquilla, J., Melko, R. G. & Khatami, E. Machine learning phases of strongly correlated fermions. Phys. Rev. X 7, 031038 (2017).
15. Stöcker, H. & Greiner, W. High-energy heavy ion collisions: probing the equation of state of highly excited hadronic matter. Phys. Rep. 137, 277–392 (1986).
16. Stephanov, M. A. QCD phase diagram: an overview. PoS LAT 2006, 024 (2006).
17. Fukushima, K. & Hadsuda, T. The phase diagram of dense QCD. Rept. Prog. Phys. 74, 014001 (2011).
18. Adams, J. et al. Experimental and theoretical challenges in the search for the quark gluon plasma: the STAR Collaboration’s critical assessment of the evidence from RHIC collisions. Nucl. Phys. A 757, 102–183 (2005).
19. Muller, B., Schardel, J. & Wysocki, B. First Results from Pb + Pb collisions at the LHC. Ann. Rev. Nucl. Part. Sci. 62, 361–386 (2012).
20. Friman, B. et al. The CRM physics book: compressed baryonic matter in laboratory experiments. Lect. Notes Phys. 814, 1–980 (2011).
21. Ablyazimov, T. et al. Challenges in QCD matter physics—the scientific programme of the Compressed Baryonic Matter experiment at FAIR. Eur. Phys. J. A 53, 60 (2017).
22. Luzum, M. & Romatschke, P. Conformal relativistic viscous hydrodynamics: applications to RHIC results at s(NN)**(1/2) = 200-GeV. Phys. Rev. C 78, 034915 (2008).
23. Pratt, S., Sangaline, E., Sorensen, P. & Wang, H. Constraining the Eq. of state of super-dense hadronic matter from heavy-ion collisions. Phys. Rev. Lett. 114, 202301 (2015).
24. Bernhard, J. E., Moreland, I. S., Bass, S. A., Liu, J. & Heinz, U. Applying Bayesian parameter estimation to relativistic heavy-ion collisions: simultaneous characterization of the initial state and quark-gluon plasma medium. Phys. Rev. C 94, 024907 (2016).
25. Heinz, U. W. Early collective expansion: relativistic hydrodynamics and the transport properties of QCD matter. Landolt-Börnstein 23, 240–292 (2010).
26. Romatschke, P. New Developments in Relativistic Viscous Hydrodynamics. Int. J. Mod. Phys. E 19, 1–53 (2010).
27. Teaney, D. A. Viscous hydrodynamics and the quark gluon plasma. Preprint at https://arxiv.org/abs/0905.2423 (2009).
28. Gale, C., Jeon, S. & Schenke, B. Hydrodynamical modeling of heavy-ion collisions. Int. J. Mod. Phys. A. 28, 1340011 (2013).
29. Strickland, M. Anisotropic hydrodynamics: three lectures. Acta Phys. Pol. B 45, 2355 (2014).
30. Pang, L. G., Wang, Q. & Wang, X. N. Effects of initial flow velocity fluctuation in event-by-event (3 + 1D) hydrodynamics. Phys. Rev. C 86, 024911 (2012).
31. Pang, L. G., Hatta, Y., Wang, X. N. & Xiao, B. W. Analytical and numerical Guoser solutions of the second-order hydrodynamics. Phys. Rev. D 91, 074027 (2015).
32. Huovinen, P. & Petreczky, P. QCD equation of state and hadron resonance gas. Nucl. Phys. A 837, 26–53 (2010).
33. Sollfrank, J. et al. Hydrodynamical description of 200-Au+Au collisions: Hadron and electromagnetic spectra. Phys. Rev. C 55, 392 (1997).
34. Shen, C. et al. The iEBE-VISHNU code package for relativistic heavy-ion collisions. Comput. Phys. Commun. 199, 61–85 (2016).
35. Stöcker, H. Collective flow signals the quark gluon plasma. Nucl. Phys. A 750, 121–147 (2005).
36. Lin, Z. W., Koc, M. L., Li, B. A., Zhang, R. & Pal, S. A Multi-phase transport model for relativistic heavy ion collisions. Phys. Rev. C 72, 064901 (2005).
37. Gale, C., Jeon, S., Schenke, B., Triadelphia, P. & Venugopalan, R. Event-by-event anisotropic flow in heavy-ion collisions from combined Yang-Mills and viscous fluid dynamics. Phys. Rev. Lett. 110, 012302 (2013).
38. Robnik-Sikonja, M. & Kononenko, I. Explaining classifications for individual instances. Knovel Data. Eng. IEEE Trans. 20, 589–600 (2008).
39. Zintgraf, I. M., Cohen, T. S., Adel, T. & Welling, M. Visualizing deep neural network decisions: prediction difference analysis. Preprint at https://arxiv.org/abs/1702.04959 (2017).
40. Chaudhuri, A. K. & Heinz, U. W. Hydrodynamical evolution of dissipative QGP fluid. J. Phys. Conf. Ser. 50, 251–258 (2006).
41. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. In: Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q. (eds) Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS, 2012).
42. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. Preprint at https://arxiv.org/abs/1409.1556 (2015).
43. Ioffe, S. & Szegedy, C. Batch normalization: accelerating deep network training by reducing internal covariate shift. Preprint at https://arxiv.org/abs/1502.03167 (2015).
44. Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Improving neural networks by preventing co-adaptation of feature detectors. Preprint at https://arxiv.org/abs/1207.0580 (2012).
45. Srivastava, N. et al. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929 (2014).
46. He, K., Zhang, X., Ren, S. & Sun, J. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. https://arxiv.org/abs/1502.01852 (2015).
47. Ng, A. Y. Feature selection. In Proc 21st International Conference on Machine Learning. (Banff, Canada, 2004).
48. Chollet, F. Keras: The Python Deep Learning library https://github.com/fchollet/keras (2015).
49. Kullback, S. & Leibler, R. A. On information and sufficiency. Ann. Math. Stat. 22, 79–86 (1951).
50. Kingma, D. & Ba, J. Adam: a method for stochastic optimization. Preprint at https://arxiv.org/abs/1412.6980 (2015).
51. Abadi, M. et al. TensorFlow: Large-scale machine learning on heterogeneous systems. Preprint at https://arxiv.org/abs/1603.04467 (2016).
52. Pang, L. G. et al. Training and testing data used in the paper “An equation-of-state-meter of QCD transition from deep learning”, figshare. https://doi.org/10.6084/m9.figshare.5457220.v1 (2017).

Acknowledgements
L.-G.P. and H.P. acknowledge funding of a Helmholtz Young Investigator Group VH-NG-822 from the Helmholtz Association and the GSI Helmholtzentrum für Schwerionenforschung (GSI). N.S. and K.Z. acknowledge the generous support of their DL-research at FIAS by SAMSON AG, Frankfurt and the support from GSI. H.St. acknowledges the support through the Judah M. Eisenberg Laureatus Chair at Goethe University. L.-G.P. and X.-N.W. are supported in part by the National Science Foundation (NSF) within the framework of the JETSCAPE collaboration, under grant number ACI-1505028. X.-N.W. was supported in part by NSFC under the Grant No. 11521064, by MOST of China under Grant No. 2014DFG02050, by the Major State Basic Research Development Program (MSBRD) in China under the Grant No. 2015CB856902 and by U. S. DOE under Contract No. DE-AC02-05CH11231. This work was supported in part by the Helmholtz International Center for the Facility for Antiproton and Ion Research (HIC for FAIR) within the framework of the Landes-Offensive zur Entwicklung Wissenschaftlich-Oekonomischer Exzellenz (LOEWE) program launched by the State of Hesse. The computations were done in the Green-Cube GPU cluster LCSC at GSI, the Loewe-CSC at Goethe University, NERSC at LRNL and the GPU cluster at Central China Normal University.

Author contributions
L.-G.P. contributed to the idea, the training and the second testing data set, the neural network construction for training/testing and the manuscript preparation; K.Z. contributed to the idea, the first testing data set, intensive discussions on neural network structures, physical explanations of the results and the manuscript edition; N.S. contributed to intensive discussions on neural network structures, physical explanations of the results and the manuscript edition; H.P. and H.S. and X.-N.W. contributed to the computing resources, physical insights and manuscript editions.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-017-02726-3.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.