Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **498,** 5620–5628 (2020)
Advance Access publication 2020 September 10

doi:10.1093/mnras/staa2690

A deep learning view of the census of galaxy clusters in IllustrisTNG

Y. Su,[1★] Y. Zhang,[1,2] G. Liang,[1,2] J. A. ZuHone,[3] D. J. Barnes,[4] N. B. Jacobs,[2] M. Ntampaka,[3,5] W. R. Forman,[3] P. E. J. Nulsen,[3] R. P. Kraft[3] and C. Jones[3]

[1] *Department of Physics and Astronomy, University of Kentucky, 505 Rose Street, Lexington, KY 40506, USA*
[2] *Department of Computer Science, University of Kentucky, 329 Rose Street, Lexington, KY 40506, USA*
[3] *Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA 02138, USA*
[4] *Department of Physics, Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[5] *Harvard Data Science Initiative, Harvard University, Cambridge, MA 02138, USA*

Accepted 2020 August 25. Received 2020 August 20; in original form 2020 July 6

ABSTRACT

The origin of the diverse population of galaxy clusters remains an unexplained aspect of large-scale structure formation and cluster evolution. We present a novel method of using X-ray images to identify cool core (CC), weak cool core (WCC), and non-cool core (NCC) clusters of galaxies that are defined by their central cooling times. We employ a convolutional neural network, ResNet-18, which is commonly used for image analysis, to classify clusters. We produce mock *Chandra* X-ray observations for a sample of 318 massive clusters drawn from the *IllustrisTNG* simulations. The network is trained and tested with low-resolution mock *Chandra* images covering a central 1 Mpc square for the clusters in our sample. Without any spectral information, the deep learning algorithm is able to identify CC, WCC, and NCC clusters, achieving balanced accuracies (BAcc) of 92 per cent, 81 per cent, and 83 per cent, respectively. The performance is superior to classification by conventional methods using central gas densities, with an average BAcc = 81 per cent, or surface brightness concentrations, giving BAcc = 73 per cent. We use class activation mapping to localize discriminative regions for the classification decision. From this analysis, we observe that the network has utilized regions from cluster centres out to $r \approx 300$ kpc and $r \approx 500$ kpc to identify CC and NCC clusters, respectively. It may have recognized features in the intracluster medium that are associated with AGN feedback and disruptive major mergers.

Key words: methods: data analysis – galaxies: clusters: intracluster medium – X-rays: galaxies: clusters.

1 INTRODUCTION

As the product of hierarchical structure formation, clusters of galaxies are the largest gravitationally collapsed objects in the Universe, carrying valuable information on the nature of dark matter and dark energy. Clusters of galaxies contain vast reservoirs of intracluster medium (ICM), radiating vigorously in X-rays, providing unique laboratories to study the cooling and heating of the hot baryons and the astrophysical processes that shape their thermodynamical properties.

Galaxy clusters are conventionally divided into three categories: cool core (CC), weak cool core (WCC), and non-cool core (NCC) based on their core properties. CC clusters feature a sharp X-ray emission peak associated with a dense, cool, and enriched core (Sanders et al. 2004). The gas cooling time at centres of CC clusters is much shorter than the Hubble time. High sensitivity X-ray observations provided by *Chandra* and *XMM–Newton* reveal interactions between the active galactic nuclei (AGNs) at the centres of the brightest cluster galaxies (BCG) and the ambient ICM manifested by X-ray cavities, jets, and shocks (e.g. Fabian 2012; Randall et al. 2015; Su et al. 2017a), which could pump additional energy into the ICM and compensate for the radiative losses. In contrast, the gaseous, thermal, and chemical distributions of NCC clusters are relatively homogeneous over the inner region of a cluster. WCC clusters, often featuring a remnant CC, appear to be an intermediate class (and possibly a transitional phase) between CC and NCC clusters (Markevitch et al. 2003; Su et al. 2016).

The origin of different populations of galaxy clusters has been a subject of debate for decades. In the prevailing model, a CC is considered to be the natural state resulting from radiative cooling. Major mergers may have disrupted cluster CCs and created NCC clusters, while CC clusters have only experienced minor or off-axis mergers. This interpretation is supported by X-ray observations showing that CC clusters appear to have a more symmetric morphology than NCC clusters (Buote & Tsai 1996; Lovisari et al. 2017). Radio observations also reveal that clusters that host large-scale diffuse synchrotron emissions, suggesting that they have undergone a recent merger, are predominantly NCC clusters (Rossetti et al. 2011). However, CC and NCC clusters do not appear to have different gas properties at large radii (Ghirardini et al. 2019; Ghizzardi et al. 2020). Conflicting results have also emerged in numerical simulations as to whether mergers are capable of transforming CC clusters into NCC clusters (Poole et al. 2008; Rasia et al. 2015; Barnes et al. 2018). In an alternative scenario, the presence (or absence) of a CC is determined by the physical conditions and mechanisms at cluster centres, e.g. the level of thermal conduction (Cavagnolo et al. 2008; Voit et al. 2008) and precipitation (Voit et al. 2015), the power of AGN outburts (Guo & Mathews 2010), or the combined effect of mergers and AGN activity (Chadayammuri et al. 2020). X-ray observations indicate that gas properties of cluster cores display little evolution over the last 10 Gyr, suggesting that thermal equilibrium and feedback processes in cluster cores have been in place since

[★] E-mail: ysu262@g.uky.edu

the early Universe (Hlavacek-Larrondo et al. 2015; McDonald et al. 2017; Su et al. 2019b; Ghirardini et al. 2020).

It is desirable to obtain a complete and unbiased picture of galaxy clusters to understand the origin of their diversity, the interplay between the ICM and AGN feedback, and the formation and evolution of large-scale structure. Flux-limited X-ray-selected samples are biased towards CC clusters as their centres are X-ray brighter than NCC clusters at a given cluster mass (Hudson et al. 2010; Eckert, Molendi & Paltani 2011). Recent Sunyaev–Zel'dovich (SZ) surveys provide nearly unbiased mass-limited samples of galaxy clusters. It was found that two-thirds of the Planck clusters are NCC or WCC (Andrade-Santos et al. 2017; Rossetti et al. 2017).

Ongoing and future extragalactic surveys such as eROSITA, SPT-3G, and LSST are designed to detect ∼100 000 clusters, allowing the model-independent determination of cosmological parameters (Haiman, Mohr & Holder 2001). Modern data analysis techniques can be utilized to efficiently characterize the cluster properties across the electromagnetic spectrum. Machine learning tools have been applied to reduce errors in galaxy cluster X-ray masses (Green et al. 2019; Ntampaka et al. 2019), dynamical masses (Ntampaka et al. 2015, 2016; Ho et al. 2019; Kodi Ramanah et al. 2020), SZ masses (Gupta & Reichardt 2020a), lensing analyses (Gupta & Reichardt 2020b; Springer et al. 2020), and to model micro-calorimeter X-ray spectra (Ichinohe et al. 2018). These techniques offer flexibility to take advantage of complicated correlations, well suited for mining large data sets and extracting information in the observational data that is inaccessible by conventional methods.

We present a deep learning approach to characterizing the thermodynamic structures of clusters of galaxies. The paper is structured as follows. In Section 2, we describe the IllustrisTNG simulations, the mock *Chandra* observations, and the network architecture. We present the predicted cluster type classifications in Section 3. We discuss the implication of this work in Section 4, and conclude in Section 5.

2 METHODS

2.1 IllustrisTNG clusters

The IllustrisTNG project includes a series of state-of-the-art cosmological magnetohydrodynamical simulations of galaxy formation (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018, 2017) . It is a successor to the original Illustris simulation (Vogelsberger et al. 2014). IllustrisTNG utilizes both large volumes and high resolutions, which reproduces relations between black hole masses and the properties of their host galaxies (Li et al. 2019a), the metal abundance of the ICM (Vogelsberger et al. 2018), and the cosmic large-scale structures (Springel et al. 2018). TNG300 is the largest simulation volume in IllustrisTNG, containing a simulated cubic volume of (300 Mpc)^3 with a baryonic mass resolution of 7.6×10^6 M$_\odot$ (Nelson et al. 2019), providing a rich and diverse collection of collapsed haloes (Pillepich et al. 2018). The simulations use a cosmological model based on the constraints of Planck Collaboration XXIV (2016) with $\Omega_{\rm m} = 0.3089$, $\Omega_\Lambda = 0.6911$, and $H_0 = 67.74\,{\rm km\,s^{-1}\,Mpc^{-1}}$.

We select galaxy clusters with a total mass within R_{500}[1] above $M_{500} = 10^{13.75}$ M$_\odot$ using the Friends-of-Friends algorithm (Davis et al. 1985) from the $z = 0$ snapshot in the TNG300 simulation, which forms an unbiased mass-limited sample of 318 massive clusters. A detailed analysis of the cluster populations in TNG300 is presented in Barnes et al. (2018). The radiative cooling time is defined as

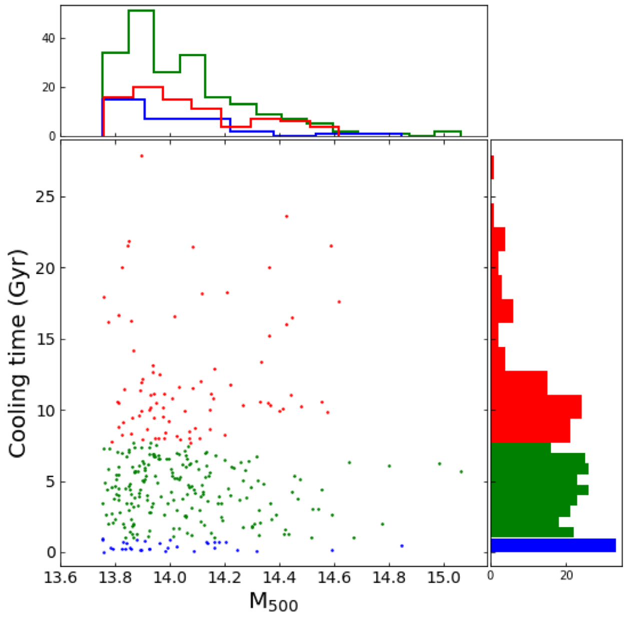

Figure 1. Distributions of central cooling times and log$M_{500}/M_\odot$ of TNG300 clusters in our sample. Their central cooling times are in the range of 0.012–27.85 Gyr. We define CC and NCC clusters as those with cooling times shorter than 1 Gyr and longer than 7.7 Gyr, respectively. Clusters with $1 < t_{\rm cool} <$ 7.7 Gyr are defined as WCC clusters. Clusters in our sample have M_{500} in the range of $10^{13.75–15.06}$ M$_\odot$.

$$t_{\rm cool} = \frac{3}{2}\frac{(n_e + n_i)k_B T}{n_e n_i \Lambda(T, Z)} \qquad (1)$$

where n_e and n_i are the number densities of electrons and ions, respectively; k_B is Boltzmann constant, and T is the gas temperature; Λ, the cooling function, is determined by the plasma temperature and metallicity. Following Barnes et al. (2018), we calculate the average $t_{\rm cool}$ from a 3D volume within 0.012 R_{500}. CC clusters are defined as those with $t_{\rm cool} < 1$ Gyr, an observation-based threshold for the presence of multiphase gas likely due to the thermally unstable cooling. NCC clusters are those with $t_{\rm cool} > 7.7$ Gyr, corresponding to a lookback time to $z \approx 1$ and representing the period since the last major merger. Clusters with $t_{\rm cool}$ between 1 and 7.7 Gyr are classified as WCC clusters. Such divisions for CC, WCC, and NCC clusters are commonly adopted in practice (e.g. Hudson et al. 2010; McDonald et al. 2013; Hogan et al. 2017; Barnes et al. 2018). 10 per cent, 61 per cent, 29 per cent of clusters in our sample are CC, WCC, and NCC, respectively. Distributions of the masses and cooling times of clusters in our sample are shown in Fig. 1.

2.2 Mock *Chandra* observations

Mock *Chandra* X-ray observations of the TNG300 clusters are produced in an end-to-end fashion using PYXSIM v2.2.0[2], an implementation of the PHOX algorithm (Biffi, Dolag & Böhringer 2013; ZuHone et al. 2014), and the SOXS v2.2.0[3] software suite for simulating X-ray events and producing mock observations. A large number of photons in the energy band of 0.5–7.0 keV are

[1] R_Δ is the radius within which the overdensity of the galaxy cluster is Δ times the critical density of the Universe.

[2] http://hea-www.cfa.harvard.edu/∼jzuhone/pyxsim

[3] http://hea-www.cfa.harvard.edu/soxs

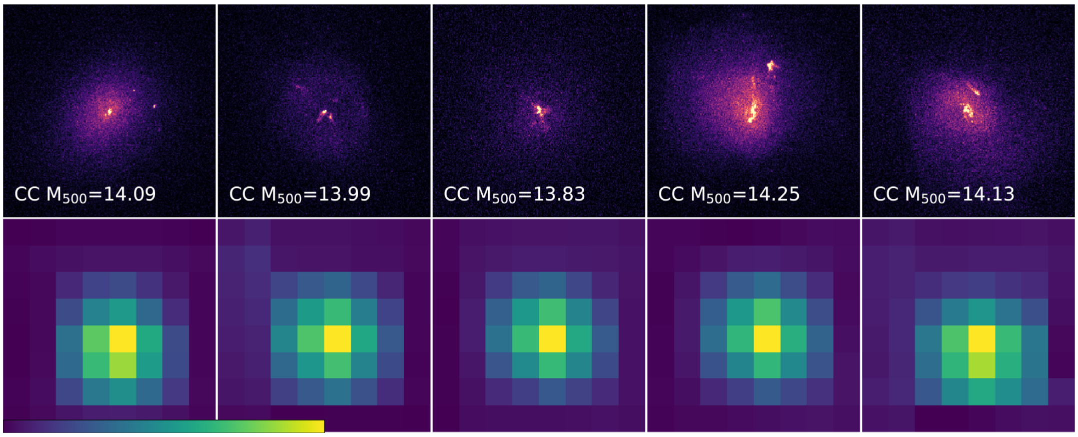

Figure 2. **Top**: Example images of mock *Chandra* observations of CC clusters. Each image covers a $D = 1$ Mpc square region. **Bottom**: Class activation maps highlight the discriminative regions in an image for the CNN to classify that image into a category. Each map corresponds to the above input image. All these clusters are predicted correctly with a probability above 0.9. The network has utilized radial ranges more extended than $r < 0.012\,R_{500}$ (where the central cooling time and density are measured) to identify CC clusters.

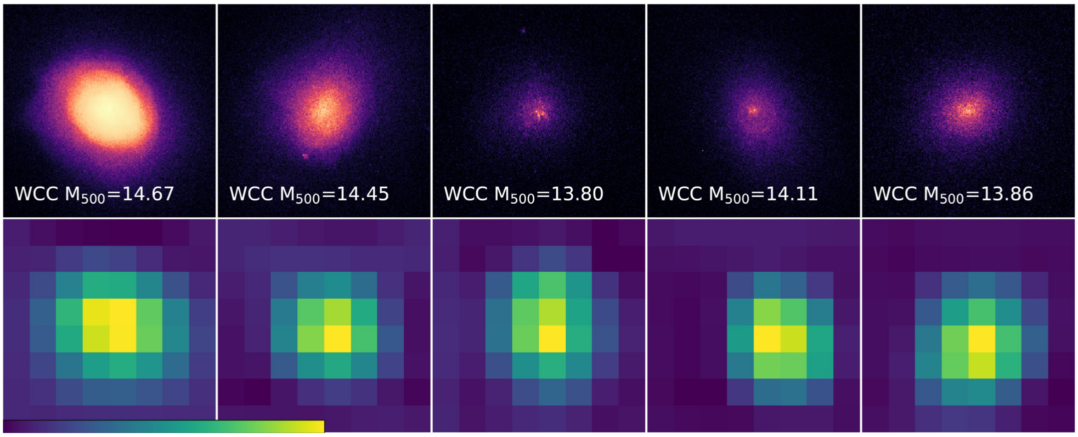

Figure 3. Same as Fig. 2 but for WCC clusters.

generated with PYXSIM for each cluster over a spherical volume with a radius of 2 Mpc, based on their 3D distributions of density, temperature, and metallicity in TNG300. We adopt a *wabs* × *apec* model, where the *apec* thermal emission model (Foster et al. 2012) represents the ICM component and the *wabs* model (Morrison & McCammon 1983) characterizes the foreground Galactic absorption assuming a hydrogen column density of 4×10^{20} cm^{-2}. We assume all the clusters reside at a redshift of $z = 0.05$, such that 1 arcsec = 1.01 kpc for the assumed cosmological parameters in IllustrisTNG. Each data set is then projected along three orthogonal directions x, y, z. Mock *Chandra* ACIS-I event files are produced by convolving each photon list with an instrument model for the ACIS-I detector of *Chandra*. The effective area and spectral response are based on the Cycle 0 response files. The ACIS-I particle background, the galactic foreground, and the Cosmic X-ray background are also included. Each mock observation is integrated for an exposure time of 100 ks. We extract images of the central 16.8 arcmin square region in the 0.5–7.0 keV energy band from the simulated event files. The field of view corresponds to a 1 Mpc square at the assumed redshift. Each 8 × 8 pixel^2 is binned up into a single pixel such that the final mock ACIS-I images have a dimension of 256 × 256. Example mock *Chandra* images of CC, WCC, and NCC clusters are shown in Figs 2, 3, and 4, respectively.

2.3 Neural network architecture

Convolutional neural networks (CNNs; Fukushima & Miyake 1982; LeCun et al. 1999; Krizhevsky, Sutskever & Hinton 2012; Simonyan & Zisserman 2014) are a class of deep machine learning algorithms that are commonly used for image analysis. Unlike traditional (shallow) image understanding methods, CNNs extract meaningful patterns from the input imagery using sets of convolutional layers (Conv-layer) with weights that are optimized for a given loss function. The output of each convolutional layer is a feature map, which is a vector-valued spatial function defined over a grid of image locations. Network architectures typically consist of a linear sequence of Conv-layers followed by a set of fully connected layers. The Conv-layers extract spatial features, often with a reduction in spatial resolution later in the sequence. After the Conv-layers, the spatial feature map is converted into vector, by either averaging the features across the image or just reshaping the feature map into a vector (‘flattening’). The subsequent fully connected layers label the data with discrete labels for classification tasks or continuous labels for regression tasks. Increasing the number of layers in a CNN will tend to improve results, but at some point, very deep models become too difficult to train. Residual neural networks (ResNets He et al. 2016a; He et al. 2016b) are a type of CNNs that use skip connections,

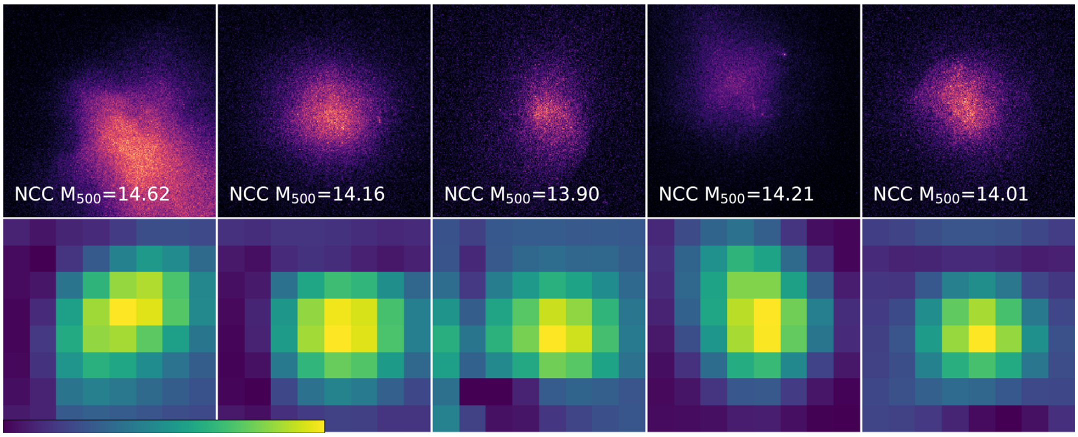

Figure 4. Same as Fig. 2 but for NCC clusters. The network has utilized regions out to the edge of the input image to identify NCC clusters.

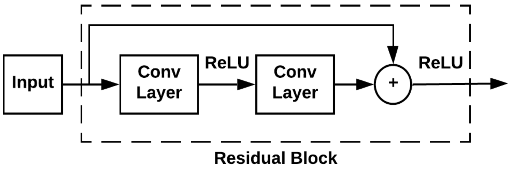

Figure 5. A residual block is a two Conv-layers shallow network with each Conv-layer followed by a ReLU. A skip connection passes the input of the residual block to be added to the output of the second Conv-layer and fed to the second ReLU. One advantage of building a model with residual blocks is that it allows for deeper and more flexible models that can be trained efficiently.

which has been shown to reduce the difficulty in training CNNs with many layers. ResNets have been used in astronomical applications including finding strong gravitational lenses (Lanusse et al. 2018), galaxy morphology classification (Zhu et al. 2019), and identifying candidate Lyman α emitting galaxies (Li et al. 2019b).

The ResNet-18 network, employed in this study, contains one Conv-layer, eight residual blocks, and one fully connected layer. A residual block is a shallow network of two Conv-layers (Fig. 5). Each Conv-layer is followed by a rectified linear unit (ReLU) (Zeiler et al. 2013). A skip connection is added to the data passing flow to directly pass the input of the residual block to the end of the second Conv-layer. The input of the residual block and the output of the second Conv-layer are then added together to be fed to the second ReLU. The output of the second ReLU is the output of the residual block. A 3 × 3 max-pooling layer follows the first Conv-layer, and a global average pooling (GAP) layer follows the last residual block. The ResNet-18 network contains a total of 18 hidden layers. Its basic architecture is shown in Fig. 6.

Our network is implemented in PyTorch (Paszke et al. 2019). A learning rate of lr = 0.001, a batch size of 64, and Adam optimizer (Kingma & Ba 2014) are used during training. A ResNet-18 model is pre-trained on the ImageNet Dataset which contains over one million images for a 1000-class classification (Deng et al. 2009). The pre-trained network is fine tuned with our data set for predicting cluster types. Weighted cross-entropy (LeCun, Bengio & Hinton 2015) is used as our loss function. Weights that are inversely proportional to the number of data in each class are included in the loss function to mitigate the impacts of the imbalanced data set. The mock *Chandra* images have a dimension of 256 × 256. Since the ResNet-18 network expects a 3-channel input image, each image is replicated three times to form a 256 × 256 × 3 image. All the input images are randomly split into 10 folds (groups) of roughly equal size. No image from the same cluster appears in more than one fold. We use 8 folds for training, 1 fold for validation, and 1 for testing. Input images are augmented by a random combination of horizontal/vertical flip and 0/90/180/270 degrees rotation during training. Each model is trained for 200 epochs. The model that gives the highest F_1-score (equation 5) on the validation set is chosen and used for testing. A 10-fold cross-validation has been applied to cycle through all the data.

We apply a class activation mapping (Zhou et al. 2016) technique to highlight regions that are discriminative for the CNN. We compute

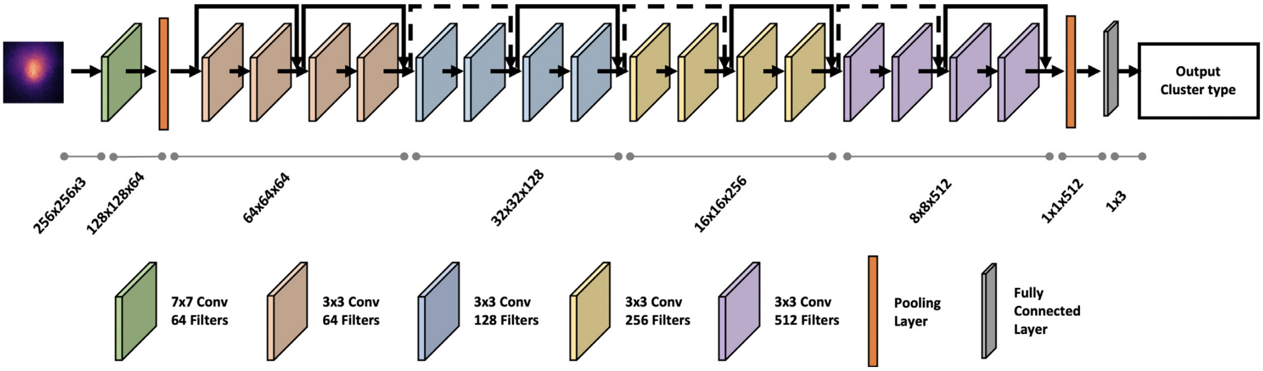

Figure 6. Architecture of a ResNet-18 neural network. The input and output shapes of each layer are labelled. After the first pooling layer, every two Conv-layers form a residual block as illustrated in Fig. 5. The dashed shortcuts involve dimension changes.

a weighted sum of the feature maps of the last Conv-layer to obtain a class activation map (CAM) for each image. The activation of unit k in the last Conv-layer at a 2D coordinate of (x, y) is $f_k(x, y)$. The result of GAP for that unit is $F_k = \sum_{x,y} f_k(x, y)$. The input to the softmax for class c is $S_c = \sum_k w_k^c F_k$, where w_k^c is the weight for class c and unit k. We obtain

$$S_c = \sum_{x,y}\sum_{k} w_k^c f_k(x, y) = \sum_{x,y} M_c(x, y), \tag{2}$$

where $M_c(x, y)$ is the value on the CAM for position (x, y). The probability for each class, $P_c = \exp(S_c)/\sum_c \exp(S_c)$, is used to make the final decision. CAM therefore reveals the importance of each part in an image that leads to the classification of an image to a class. The resulting CAM has the same dimension as the output of the last Conv-layer and the input of the GAP layer of 8 × 8.

3 RESULTS

We use our CNN algorithm to predict whether a cluster is CC, WCC, or NCC from the mock *Chandra* X-ray images. The cluster types are defined by their actual central cooling times in TNG300. We compare the performances with the estimates given by two traditional methods of using central gas densities and surface brightness concentrations.

We use the following criteria to evaluate the performance of each experiment. Hereafter, *tp*, *fp*, *tn*, and *fn* are the numbers of true positive, false positive, true negative, and false negative predictions, respectively. Precision, also called positive predictive value, is the number of true positives, divided by the number of all positive calls:

$$\text{Precision} = \frac{tp}{tp + fp}. \tag{3}$$

Recall, also called true positive rate, is the number of true positives divided by the number of positive samples:

$$\text{Recall} = \frac{tp}{tp + fn}. \tag{4}$$

F_1-score is the harmonic mean of precision and recall, defined as

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{5}$$

It conveys the balance between precision and recall and provides a more comprehensive evaluation. We base our main conclusions on F_1-score. Balanced accuracy (BAcc) is the average of true positive predictions divided by the number of positive samples and true negative predictions divided by the number of negative samples. It is related to *tp*, *fp*, *tn*, and *fn*:

$$\text{BAcc} = \frac{1}{2}\left(\frac{tp}{tp + fn} + \frac{tn}{tn + fp}\right). \tag{6}$$

BAcc is a measurement of accuracy that does not suffer from imbalanced data sets.

We train and test our deep learning classification algorithm with mock *Chandra* ACIS-I images as shown in Figs 2–4 and described in Section 2.2. Each ACIS-I field covers a 1 Mpc square, whereas clusters in our sample have a median R_{500} of 710 kpc. The spatial resolution is degraded to 3.9″/pixel which is 8 times worse than the half arcsec resolution of *Chandra* ACIS. Without any spectral information, the network is able to distinguish CC, WCC, and NCC clusters with F_1-scores of 0.83, 0.82, and 0.73, respectively. Details of the results are shown in Fig. 7 and values of performance measures are listed in Table 1. Predictions and the ground truths are compared in the normalized confusion matrix as shown in Fig. 8. Diagonal elements represent the fraction of data for which the predicted class is the same as the true class, while off-diagonal elements are those that are misclassified. The deep learning algorithm gives a confusion matrix with high diagonal values, indicating good predictions.

Here, we compare the deep learning method to more conventional methods for cluster classification. Although these approaches are not directly comparable, the comparisons are instructive. A rapidly cooling core implies a high central gas density as the ICM gradually loses its pressure support and falls to smaller radii. Central gas densities have been widely used to determine whether a cluster contains a CC, which requires far fewer counts than measuring the temperatures and metallicities (Lovisari, Reiprich & Schellenberger 2015; Su et al. 2019a). We calculate the central electron number density n_e as the average n_e of a 3D volume within 0.012 R_{500} as shown in Barnes et al. (2018). Following Barnes et al. (2018) and Hudson et al. (2010), clusters with a central $n_{\rm e} > 1.5 \times 10^{-2}\,\rm cm^{-3}$, $1.5 \times 10^{-2} > n_{\rm e} > 0.5 \times 10^{-2}\,\rm cm^{-3}$, and $n_{\rm e} < 0.5 \times 10^{-2}\,\rm cm^{-3}$ are classified as CC, WCC, and NCC, respectively. For clusters in our sample, this method achieves $F_1 = 0.69$, averaged over the three cluster types (see Figs 7 and 8 and Table 1), which is not as accurate as the predictions given by our ResNet-18 classifier.

In X-ray observations, it is challenging to directly measure gas properties within $r \lesssim 10$ kpc for a modest exposure time. The elevated ICM metallicity and density at the centre of a CC cluster produce a central peak in X-ray surface brightness. The ratio of this peak emission to the ambient emission is therefore sensitive to the CC strength. The X-ray concentration parameter was originally introduced by Santos et al. (2008) to infer whether a cluster contains a CC:

$$C_{\rm SB} = \frac{\sum(<40\,\rm kpc)}{\sum(<400\,\rm kpc)}, \tag{7}$$

where $\sum(<r)$ is the accumulated projected ICM emission in 0.5–5.0 keV from a circular region with a radius of r. We extract images in the 0.5–5.0 keV energy band from mock *Chandra* observations. Following Barnes et al. (2018) and Andrade-Santos et al. (2017), clusters with $C_{\rm SB} > 0.155$, $0.075 < C_{\rm SB} < 0.155$, and $C_{\rm SB} < 0.075$ are classified as CC, WCC, and NCC, respectively. Using this method, we obtain an average F_1-score of 0.33 for clusters in our sample. Barnes et al. (2018) also note that this criterion overpredicts NCC clusters and fails to identify CC clusters. We further sort all the images with a decreasing $C_{\rm SB}$ and divide them into the three categories based on the fractions of CC, WCC, and NCC in our sample. We obtain an F_1-score of 0.64 (see Figs 7 and 8 and Table 1). Our ResNet-18 classifier which utilizes the 2D ICM distribution outperforms the 1D concentration measurement.

4 INTERPRETATION AND DISCUSSION

Using mock *Chandra* X-ray images of a 1 Mpc square centred on each cluster, our network is able to predict whether a cluster is CC, WCC, or NCC with an average F_1-score of 0.79. The cluster types are defined by their actual central cooling times, which depend on temperature, density, and metallicity as shown in equation (1). Our deep learning method is superior to the estimate using the actual central gas densities of these clusters with an average $F_1 = 0.69$. X-ray images may contain information that is more directly related to the cooling time than gas density. To localize features that are most useful for the network to make classification decisions, we generate a CAM for each input image as described in Section 2.3. Example CAM images and their original images are compared in Figs 2–4. Regions that are brighter in CAM are more informative for the network. We stack and normalize all the CAM images associated with correct predictions with a probability above 0.9 for CC, WCC, and NCC clusters, respectively, as shown in Fig. 9. The radial profiles of the values in the activation maps are shown in the right-hand panel

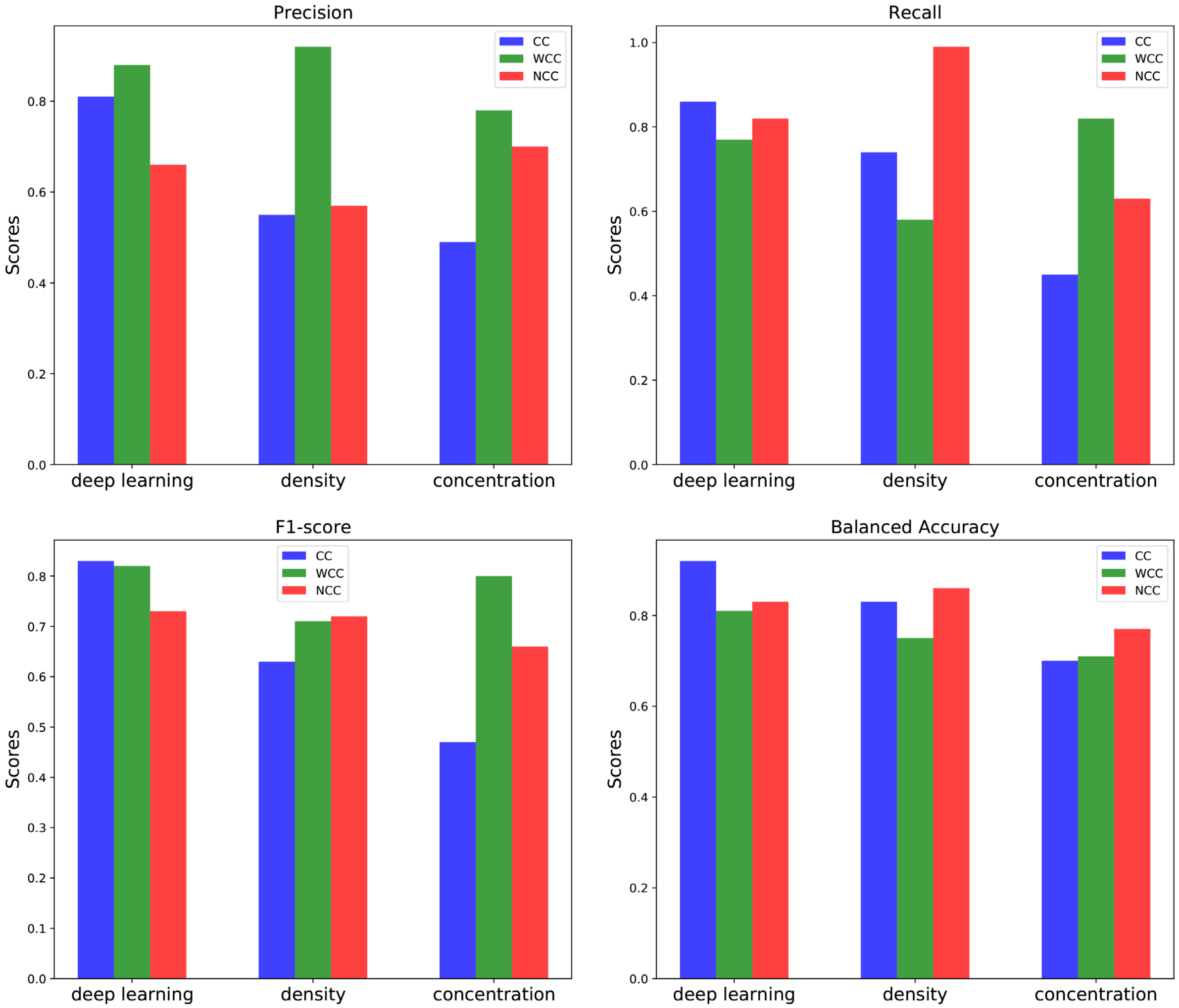

Figure 7. Scores of precision (top-left), recall (top-right), F_1-score (bottom-left), and balanced accuracy (bottom-right) for CC, WCC, and NCC classifications obtained using different methods (x-axis): 1. Deep learning using mock *Chandra* images. 2. Central gas density. 3. X-ray concentration parameter derived from mock *Chandra* images. Details of the performance measures are listed in Table 1.

Table 1. Values of performance measures for different experiments. Precision, recall, F_1-score, and balanced accuracy are defined in Section 3.

Method	Ave. F_1	Ave. BAcc	Class	Precision	Recall	F_1	BAcc
			CC	0.81	0.86	0.83	0.92
Deep learning	0.79	0.85	WCC	0.88	0.77	0.82	0.81
			NCC	0.66	0.82	0.73	0.83
			CC	0.55	0.74	0.63	0.83
Density	0.69	0.81	WCC	0.92	0.58	0.71	0.75
			NCC	0.57	0.99	0.72	0.86
			CC	0.49	0.45	0.47	0.70
Concentration	0.64	0.73	WCC	0.78	0.82	0.8	0.71
			NCC	0.70	0.63	0.66	0.77

of Fig. 9. To identify CC clusters, the network uses 2D information out to $r \approx 300\,\mathrm{kpc}$ which is broader than $r < 0.012\,R_{500}$ (10 kpc) where the central gas density and cooling time are measured. Patterns in X-ray images could provide important clues about the central cooling time. For example, sizes of X-ray cavities and their distances to the cluster centre are determined by the outburst of the AGN, which is related to the radiative cooling rate (Bîrzan et al. 2012; Li, Su & Jones 2018). Mechanical energies released by AGN could be dissipated by heating the ICM via turbulent cascades, which can be probed through the power spectrum of X-ray surface brightness fluctuations (Zhuravleva et al. 2014). These informative 2D features in X-ray images may have allowed the network to obtain stronger constraints on t_{cool} than the methods of using central gas densities and surface brightness concentrations.

While the discriminating power of each region declines quickly as a function of radius for CC clusters, it is relatively uniform for

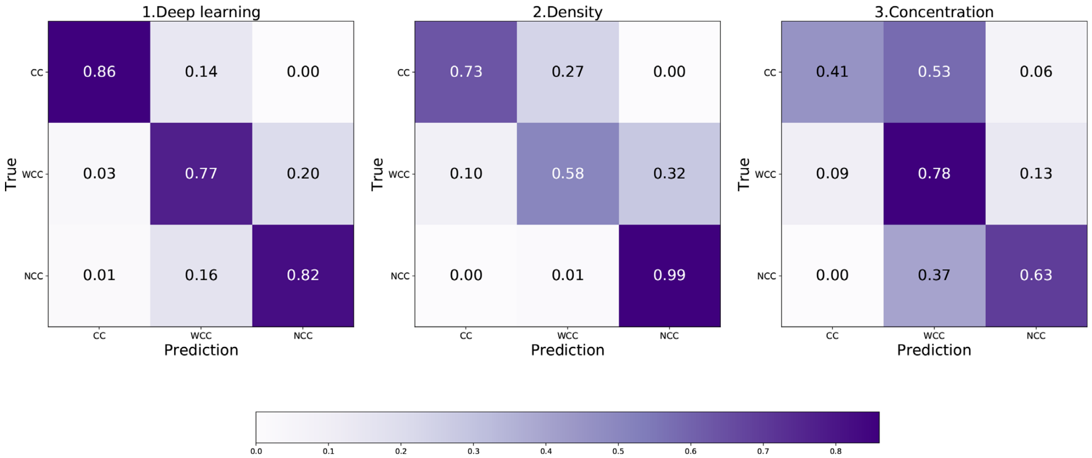

Figure 8. Normalized confusion matrix of CC, WCC, and NCC classification which compares the predicted class and the true class for each experiment. 1. Deep learning using mock *Chandra* images. 2. Central gas density. 3. X-ray concentration parameter derived from mock *Chandra* images. Details of the performance measures are listed in Table 1.

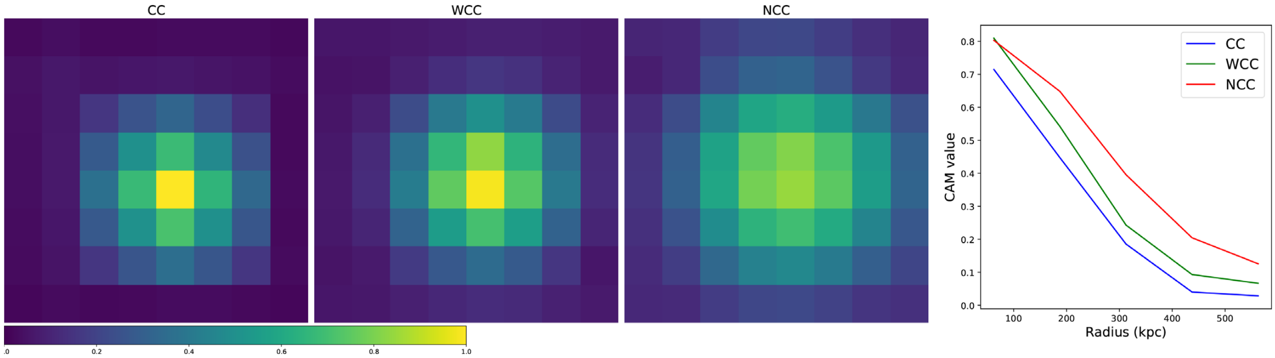

Figure 9. Class activation maps of the central $D = 1$ Mpc averaged over CC, WCC, and NCC clusters, respectively. All these clusters are predicted correctly with a probability above 0.9. The right-hand panel shows the radial profiles of the three CAM maps. The network utilizes relatively more information from the cluster centres to identify CC clusters but relies on the morphology over a wider radial range to identify NCC clusters. The radial dependance of the discriminating power of regions in WCC clusters is between those of CC and NCC clusters.

NCC cluster. The network uses extended regions out to the edge of the input images of $r \sim 500$ kpc to identify NCC clusters as shown in Fig. 9. Mergers may have disrupted the thermal structures of the ICM and contributed to the formation of NCC clusters. Interestingly, Barnes et al. (2018) find that NCC clusters do not have significantly higher kinetic energies than CC clusters. We speculate that head-on mergers and certain off-axis mergers may have similar impacts on the global kinetics of the clusters. Cluster CCs are resilient to off-axis mergers as indicated by the ubiquitous presence of sloshing cold fronts in CC clusters (Markevitch & Vikhlinin 2007; Su et al. 2017b). The impact parameter and angular momentum of a merger may be critical in determining the fate of cluster CCs, as seen in numerical simulations (Hahn et al. 2017).

We note that the misclassified cases are either associated with cooling times that are close to the class boundaries or outliers for their class. Some of the failures in our model are shown and discussed in detail in Appendix A. We only considered clusters at a single redshift in this work. Our algorithm may have some resilience to distance since the data set consists of clusters with a wide range of masses and physical sizes. The application of deep learning techniques to systems at different redshifts will be an important aspect of future studies. In overall, this work demonstrates that CNNs are able to take advantage of X-ray images and provide a unique approach to the thermodynamics of the ICM. The neural network can, in principle, be trained to predict the specific values of $t_{\rm cool}$ for CC clusters as a regression task, and fetch features associated with the life cycles of AGN feedback. Potentially, the deep learning algorithm can also be used to determine the merger history of a cluster from multiwavelength images – X-ray, radio, and optical, which would greatly enhance our understanding of cluster formation, thermalization, and particle acceleration.

5 CONCLUSIONS

ResNet-18 is a subclass of CNNs that is well suited for image classification. We employ a ResNet-18 network to assess whether a cluster is CC, WCC, or NCC from their X-ray images. The cluster type is defined purely by its central cooling time, which is related to the gas density, temperature, and metallicity. We produce mock *Chandra* observations for 318 clusters of galaxies in TNG300 with particle background, contaminating point sources, galactic foreground, etc. included. We train and test the network with low-resolution mock *Chandra* ACIS-I images. It achieves an average precision, recall, F_1-score, and balanced accuracy of 0.78, 0.82, 0.79, and 0.85, respectively, well above a random prediction of 0.33. Our deep learning algorithm outperforms the estimates given by the central gas densities and surface brightness concentration parameters. We use the class activation mapping to probe the contribution of each region to the classification decisions. The network may have utilized 2D features in X-ray images that are related to the cooling and heating mechanisms in the ICM. Features at larger radii are more important for identifying NCC clusters than CC clusters, possibly due to the role of head-on major mergers in disrupting cluster CCs.

Unlike traditional methods of using one-dimensional information to estimate the cluster type, the neural network is able to identify features on different scales and at various radii, making it a potentially powerful tool to probe the thermodynamic state of a cluster. CNNs can be utilized to exploit cluster images in *Chandra* and *XMM–Newton* archives, large cluster samples from the ongoing eROSITA all-sky survey, and the exquisite data promised by next-generation X-ray observatories, such as *Lynx*.

ACKNOWLEDGEMENTS

The authors thank the anonymous referee for his/her helpful comments. We acknowledge the use of the Lipscomb High-Performance Computing Cluster at the University of Kentucky for conducting this research. YS was partially supported by Smithsonian Astrophysical Observatory grants AR8-19020A and GO6-17125A.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Andrade-Santos F. et al., 2017, ApJ, 843, 76
Barnes D. J. et al., 2018, MNRAS, 481, 1809
Biffi V., Dolag K., Böhringer H., 2013, MNRAS, 428, 1395
Bîrzan L., Rafferty D. A., Nulsen P. E. J., McNamara B. R., Röttgering H. J. A., Wise M. W., Mittal R., 2012, MNRAS, 427, 3468
Buote D. A., Tsai J. C., 1996, ApJ, 458, 27
Cavagnolo K. W., Donahue M., Voit G. M., Sun M., 2008, ApJ, 683, L107
Chadayammuri U., Tremmel M., Nagai D., Babul A., Quinn T., 2020, preprint (arXiv:2001.06532)
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Deng J., Dong W., Socher R., Li L.-J., Li K., Fei-Fei L., 2009, ImageNet: A large-scale hierarchical image database, IEEE Conference on Computer Vision and Pattern Recognition. p. 248
Eckert D., Molendi S., Paltani S., 2011, A&A, 526, A79
Fabian A. C., 2012, ARA&A, 50, 455
Foster A. R., Ji L., Smith R. K., Brickhouse N. S., 2012, ApJ, 756, 128
Fukushima K., Miyake S., 1982, Pattern Recognit., 15, 455
Ghirardini V., Ettori S., Eckert D., Molendi S., 2019, A&A, 627, A19
Ghirardini V. et al., 2020, preprint (arXiv:2004.04747)
Ghizzardi S. et al., 2020, preprint (arXiv:2007.01084)
Green S. B., Ntampaka M., Nagai D., Lovisari L., Dolag K., Eckert D., ZuHone J. A., 2019, ApJ, 884, 33
Guo F., Mathews W. G., 2010, ApJ, 717, 937
Gupta N., Reichardt C. L., 2020a, preprint (arXiv:2003.06135)
Gupta N., Reichardt C. L., 2020b, preprint (arXiv:2005.13985)
Hahn O., Martizzi D., Wu H.-Y., Evrard A. E., Teyssier R., Wechsler R. H., 2017, MNRAS, 470, 166
Haiman Z., Mohr J. J., Holder G. P., 2001, ApJ, 553, 545
He K., Zhang X., Ren S., Sun J., 2016a , Proc. IEEE Conf. Computer Vision and Pattern Recognition, Deep Residual Learning for Image Recognition (CVPR), IEEE, Las Vegas, NV. p. 770
He K., Zhang X., Ren S., Sun J., 2016b , preprint (arXiv:1603.05027)
Hlavacek-Larrondo J. et al., 2015, ApJ, 805, 35
Ho M., Rau M. M., Ntampaka M., Farahi A., Trac H., Póczos B., 2019, ApJ, 887, 25
Hogan M. T. et al., 2017, ApJ, 851, 66
Hudson D. S., Mittal R., Reiprich T. H., Nulsen P. E. J., Andernach H., Sarazin C. L., 2010, A&A, 513, A37
Ichinohe Y., Yamada S., Miyazaki N., Saito S., 2018, MNRAS, 475, 4739
Kingma D. P., Ba J., 2014, preprint (arXiv:1412.6980)
Kodi Ramanah D., Wojtak R., Ansari Z., Gall C., Hjorth J., 2020, preprint (arXiv:2003.05951)
Krizhevsky A., Sutskever I., Hinton G. E., 2012, in Pereira F., Burges C. J. C., Bottou L., Weinberger K. Q., eds, Advances in Neural Information Processing Systems 25. Curran Associates, Inc., Lake Tahoe, NV, p. 1097
Lanusse F., Ma Q., Li N., Collett T. E., Li C.-L., Ravanbakhsh S., Mandelbaum R., Póczos B., 2018, MNRAS, 473, 3895
LeCun Y., Haffner P., Bottou L., Bengio Y., 1999, in Mundy J., Cipolla R., Forsyth D., di Gesu V., eds, Shape, Contour and Grouping in Computer Vision. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer-Verlag, Berlin, p. 319
LeCun Y., Bengio Y., Hinton G., 2015, Nature, 521, 436
Li Y., Su Y., Jones C., 2018, MNRAS, 480, 4279
Li Y. et al., 2019a, preprint (arXiv:1910.00017)
Li R., Shu Y., Su J., Feng H., Zhang G., Wang J., Liu H., 2019b, MNRAS, 482, 313
Lovisari L., Reiprich T. H., Schellenberger G., 2015, A&A, 573, A118
Lovisari L. et al., 2017, ApJ, 846, 51
McDonald M. et al., 2013, ApJ, 774, 23
McDonald M. et al., 2017, ApJ, 843, 28
Marinacci F. et al., 2018, MNRAS, 480, 5113
Markevitch M., Vikhlinin A., 2007, Phys. Rep., 443, 1
Markevitch M. et al., 2003, ApJ, 586, L19
Morrison R., McCammon D., 1983, ApJ, 270, 119
Naiman J. P. et al., 2018, MNRAS, 477, 1206
Nelson D. et al., 2017, MNRAS, 475, 624
Nelson D. et al., 2019, Computational Astrophysics and Cosmology, 6, 2
Nelson D. et al., 2019, MNRAS, 490, 3234
Ntampaka M., Trac H., Sutherland D. J., Battaglia N., Póczos B., Schneider J., 2015, ApJ, 803, 50
Ntampaka M., Trac H., Sutherland D. J., Fromenteau S., Póczos B., Schneider J., 2016, ApJ, 831, 135
Ntampaka M. et al., 2019, ApJ, 876, 82
Paszke A. et al., 2019, Advances in Neural Information Processing Systems, Curran Associates, Inc., Red Hook, NY.p. 8026
Pillepich A. et al., 2018, MNRAS, 475, 648
Planck Collaboration XXIV, 2016, A&A, 594, A24
Poole G. B., Babul A., McCarthy I. G., Sand erson A. J. R., Fardal M. A., 2008, MNRAS, 391, 1163
Randall S. W. et al., 2015, ApJ, 805, 112
Rasia E. et al., 2015, ApJ, 813, L17
Rossetti M., Eckert D., Cavalleri B. M., Molendi S., Gastaldello F., Ghizzardi S., 2011, A&A, 532, A123
Rossetti M., Gastaldello F., Eckert D., Della Torre M., Pantiri G., Cazzoletti P., Molendi S., 2017, MNRAS, 468, 1917
Sanders J. S., Fabian A. C., Allen S. W., Schmidt R. W., 2004, MNRAS, 349, 952
Santos J. S., Rosati P., Tozzi P., Böhringer H., Ettori S., Bignamini A., 2008, A&A, 483, 35
Simonyan K., Zisserman A., 2015, Very Deep Convolutional Networks for Large-Scale Image Recognition, Computing Research Repository
Springel V. et al., 2018, MNRAS, 475, 676
Springer O. M., Ofek E. O., Weiss Y., Merten J., 2020, MNRAS, 491, 5301
Su Y., Buote D. A., Gastaldello F., van Weeren R., 2016, ApJ, 821, 40
Su Y., Nulsen P. E. J., Kraft R. P., Forman W. R., Jones C., Irwin J. A., Randall S. W., Churazov E., 2017a, ApJ, 847, 94
Su Y. et al., 2017b, ApJ, 851, 69
Su Y. et al., 2019a, AJ, 158, 6
Su Y. et al., 2019b, ApJ, 881, 98
Vogelsberger M. et al., 2014, MNRAS, 444, 1518
Vogelsberger M. et al., 2018, MNRAS, 474, 2073
Voit G. M., Cavagnolo K. W., Donahue M., Rafferty D. A., McNamara B. R., Nulsen P. E. J., 2008, ApJ, 681, L5

Voit G. M., Donahue M., Bryan G. L., McDonald M., 2015, Nature, 519, 203
Zeiler M. D. et al., 2013, 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, Vancouver, BC, Canada. p. 3517
Zhou B., Khosla A., Lapedriza g., Oliva A., Torralba A., 2016, IEEE Conf. Computer Vision and Pattern Recognition, IEEE, Las Vegas, NV
Zhu X.-P., Dai J.-M., Bian C.-J., Chen Y., Chen S., Hu C., 2019, Ap&SS, 364, 55
Zhuravleva I. et al., 2014, Nature, 515, 85
ZuHone J. A., Biffi V., Hallman E. J., Randall S. W., Foster A. R., Schmid C., 2014, preprint (arXiv:1407.1783)

APPENDIX A: MISCLASSIFIED CLUSTERS

To better understand the failures of our model, we inspect the cases that are classified confidently but incorrectly. All the incorrect predictions with a probability above 0.9 are associated with WCC, which supports that WCC is a transitional phase between CC and NCC with intermediate morphologies that are more difficult to classify. Among them, two images are CC classified as WCC, eight are NCC classified as WCC, four are WCC classified as CC, and eight are WCC classified as NCC. Most of these clusters have cooling times close to the boundaries of $t_{\rm cool}\,({\rm CC|WCC}) = 1\,{\rm Gyr}$ and $t_{\rm cool}\,({\rm WCC|NCC}) = 7.7\,{\rm Gyr}$. However, six images from three clusters have typical coolings times for their type, as shown in Fig. A1. These cases appear to be outliers in their class. The top-left and top-middle images in Fig. A1 are from the same CC but it appears to be disturbed with a very asymmetric morphology and the network classified it as WCC. The top-right image is a WCC and it appears to be undergoing a merger. CAM suggests that the network may have noticed the subcluster in the lower right corner and classified it as NCC. The three images in Fig. A1 (bottom) are from the same NCC cluster but classified as WCC. It appears to be relatively relaxed. The double nuclei at the cluster centre may be mistaken as a bright filament. The network may not be well trained to identify these outliers due to the rarity of such atypical cases in our sample.

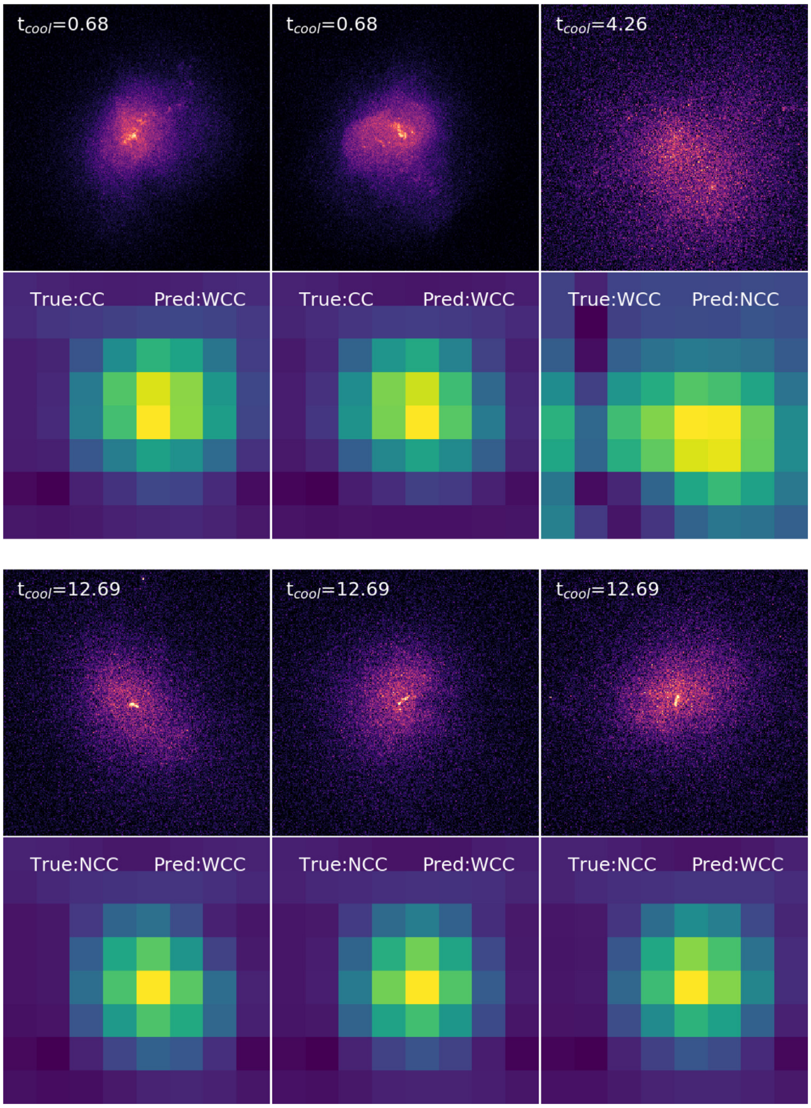

Figure A1. Same as Figs 2–4 but for clusters that are misidentified by the neural network. Their true cooling times are labelled in the *Chandra* X-ray images and their true and predicted cluster types are labelled in the class activation maps. All these clusters are predicted incorrectly with a probability above 0.9. Their X-ray morphologies appear to be atypical for their cluster types.

This paper has been typeset from a TEX/LATEX file prepared by the author.