Machine learning for galaxy clusters

Giacomo Riva

Journal club - 2022/05/12

What we will be talking about today...

Multiwavelength classification of X-ray selected galaxy cluster candidates using convolutional neural networks

Matej Kosiba [1,2★] Maggie Lieu,[2,3] Bruno Altieri,[2] Nicolas Clerc [4] Lorenzo Faccioli,[5] Sarah Kendrew,[6] Ivan Valtchanov,[7] Tatyana Sadibekova,[5,8] Marguerite Pierre,[5] Filip Hroch,[1] Norbert Werner,[9,1,10] Lukáš Burget,[11] Christian Garrel,[5] Elias Koulouridis,[12,5] Evelina Gaynullina,[8] Mona Molham [13] Miriam E. Ramos-Ceja [14] and Alina Khalikova[8]

Affiliations are listed at the end of the paper

Accepted 2020 June 9. Received 2020 June 8; in original form 2020 April 16

ABSTRACT

Galaxy clusters appear as extended sources in *XMM–Newton* images, but not all extended sources are clusters. So, their proper classification requires visual inspection with optical images, which is a slow process with biases that are almost impossible to model. We tackle this problem with a novel approach, using convolutional neural networks (CNNs), a state-of-the-art image classification tool, for automatic classification of galaxy cluster candidates. We train the networks on combined *XMM–Newton* X-ray observations with their optical counterparts from the all-sky Digitized Sky Survey. Our data set originates from the XMM CLuster Archive Super Survey (X-CLASS) survey sample of galaxy cluster candidates, selected by a specially developed pipeline, the `XAmin`, tailored for extended source detection and characterization. Our data set contains 1707 galaxy cluster candidates classified by experts. Additionally, we create an official Zooniverse citizen science project, *The Hunt for Galaxy Clusters*, to probe whether citizen volunteers could help in a challenging task of galaxy cluster visual confirmation. The project contained 1600 galaxy cluster candidates in total of which 404 overlap with the expert's sample. The networks were trained on expert and Zooniverse data separately. The CNN test sample contains 85 spectroscopically confirmed clusters and 85 non-clusters that appear in both data sets. Our custom network achieved the best performance in the binary classification of clusters and non-clusters, acquiring accuracy of 90 per cent, averaged after 10 runs. The results of using CNNs on combined X-ray and optical data for galaxy cluster candidate classification are encouraging, and there is a lot of potential for future usage and improvements.

Key words: galaxies: clusters: general – methods: data analysis – techniques: image processing.

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*

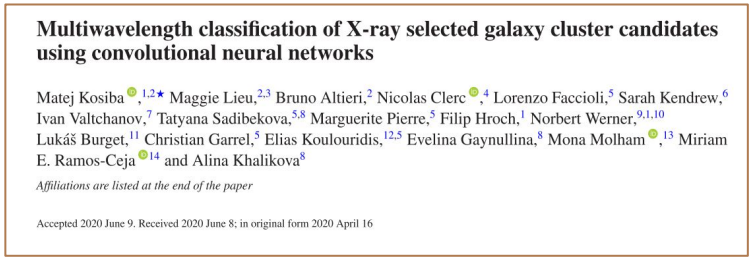
Accepted 2020 August 25. Received 2020 August 20; in original form 2020 July 6

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.

What we will be talking about today... and why

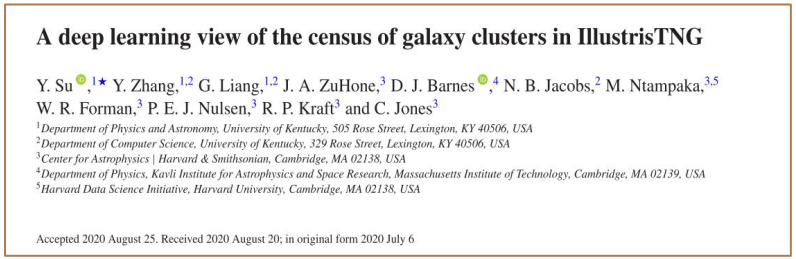


Multiwavelength classification of X-ray selected galaxy cluster candidates using convolutional neural networks

Matej Kosiba,[1,2★] Maggie Lieu,[2,3] Bruno Altieri,[2] Nicolas Clerc,[4] Lorenzo Faccioli,[5] Sarah Kendrew,[6] Ivan Valtchanov,[7] Tatyana Sadibekova,[5,8] Marguerite Pierre,[5] Filip Hroch,[1] Norbert Werner,[9,1,10] Lukáš Burget,[11] Christian Garrel,[5] Elias Koulouridis,[12,5] Evelina Gaynullina,[8] Mona Molham,[13] Miriam E. Ramos-Ceja[14] and Alina Khalikova[8]

Affiliations are listed at the end of the paper

Accepted 2020 June 9. Received 2020 June 8; in original form 2020 April 16



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

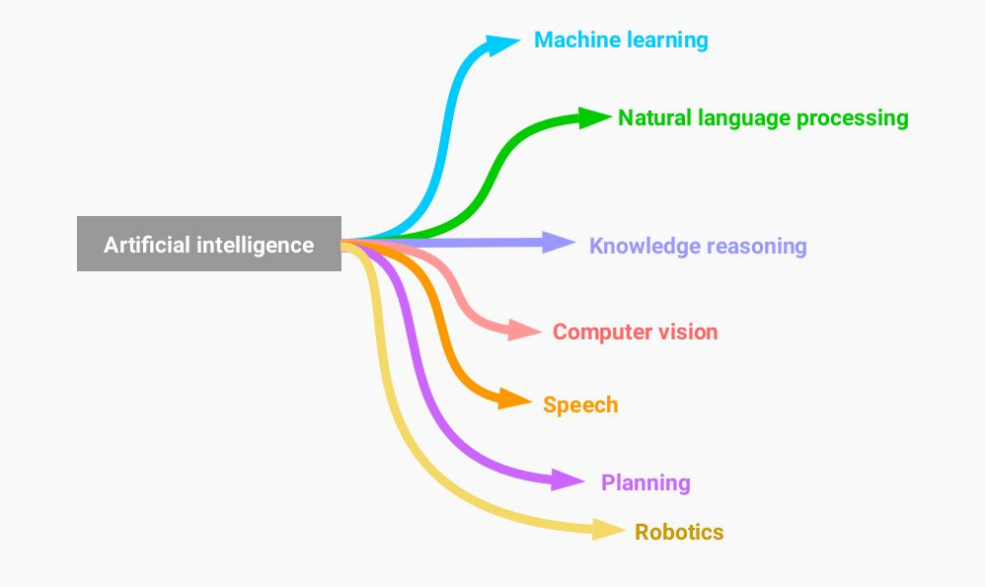
Y. Su,[1★] Y. Zhang,[1,2] G. Liang,[1,2] J. A. ZuHone,[3] D. J. Barnes,[4] N. B. Jacobs,[2] M. Ntampaka,[3,5] W. R. Forman,[3] P. E. J. Nulsen,[3] R. P. Kraft[3] and C. Jones[3]

[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

- Both works use ML techniques (CNNs, LeCun et al. 1999)
- Direct application to galaxy clusters
- I love citizen science ;)

Introduction to machine learning

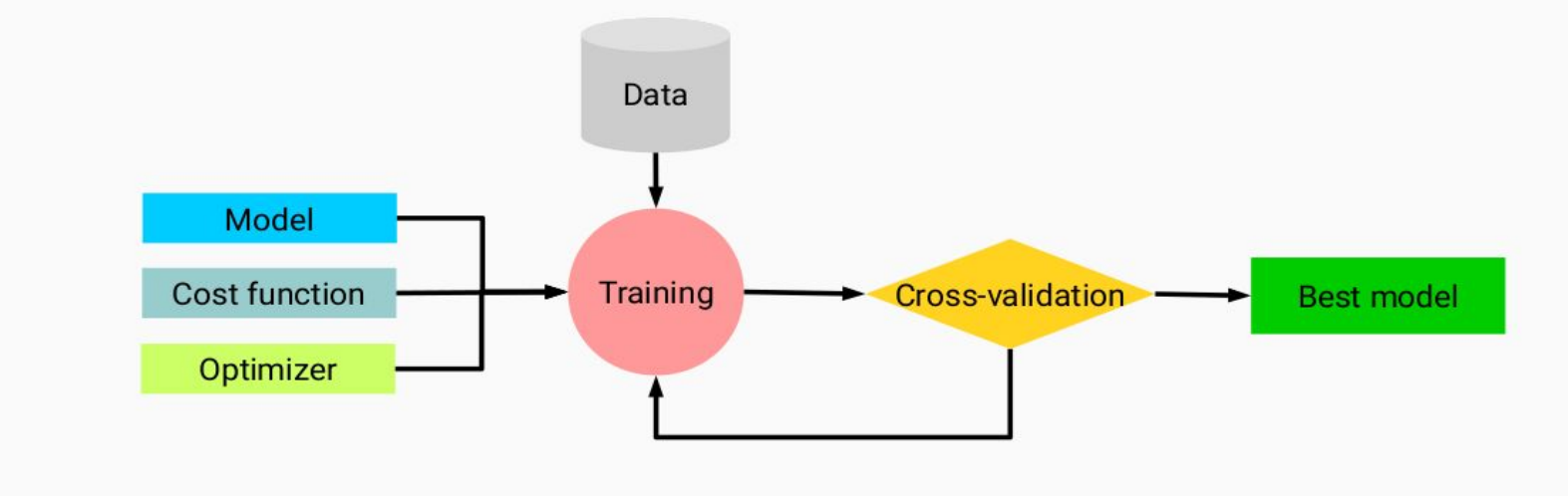


Definition of ML (T. Mitchell in 1998)

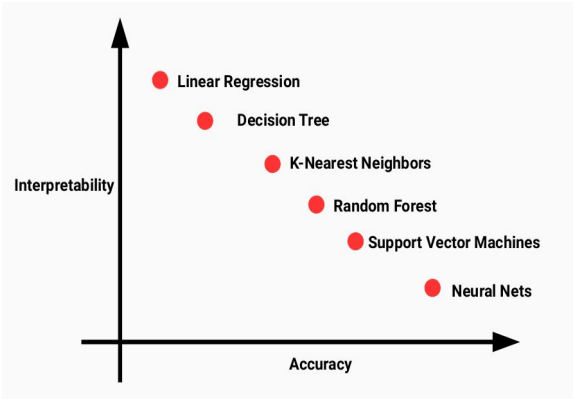
A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance on T , as measured by P , improves with experience E.

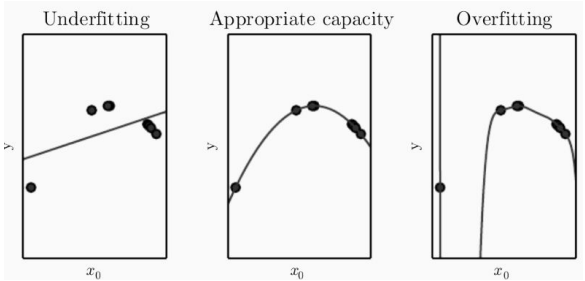
Introduction to machine learning

Different algorithms: supervised learning, unsupervised learning, reinforcement learning…

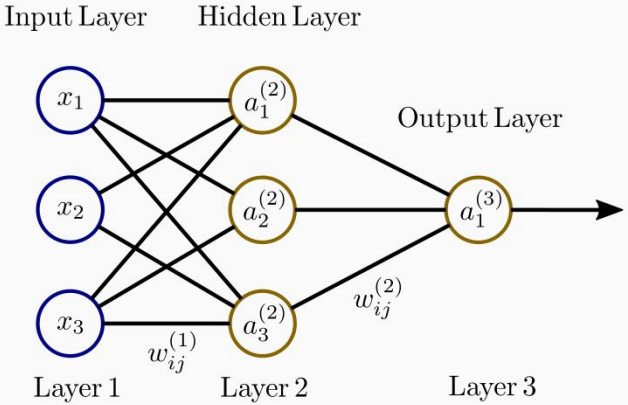


Selection of the model





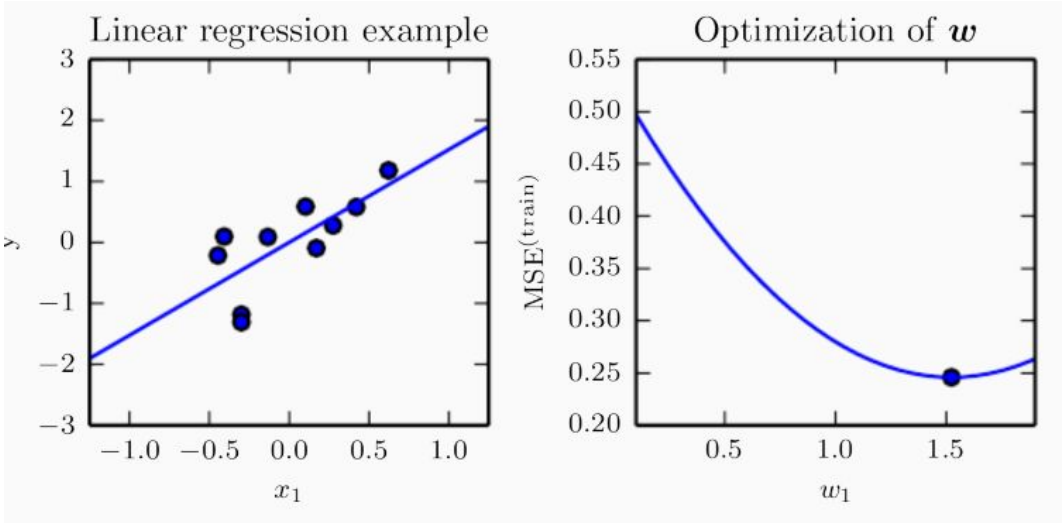
Neural network



- $a_1^{(2)} = g(w_{10}^{(1)} + w_{11}^{(1)}x_1 + w_{12}^{(1)}x_2 + w_{13}^{(1)}x_3)$
- $a_2^{(2)} = g(w_{20}^{(1)} + w_{21}^{(1)}x_1 + w_{22}^{(1)}x_2 + w_{23}^{(1)}x_3)$
- $a_3^{(2)} = g(w_{30}^{(1)} + w_{31}^{(1)}x_1 + w_{32}^{(1)}x_2 + w_{33}^{(1)}x_3)$
- **Output** $\rightarrow a_1^{(3)} = g(w_{10}^{(2)} + w_{11}^{(2)}a_1^{(2)} + w_{12}^{(2)}a_2^{(2)} + w_{13}^{(2)}a_3^{(2)})$

Model performance

Define metrics and statistical estimators for model performance (in classification problems: cost function, precision, accuracy, recall, ROC, AUC)

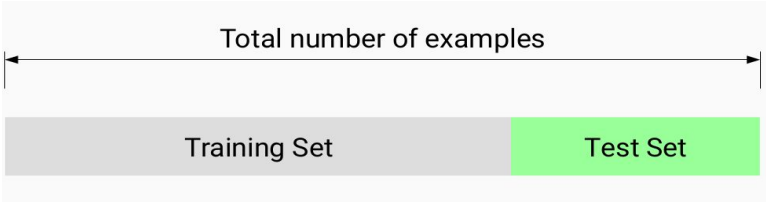


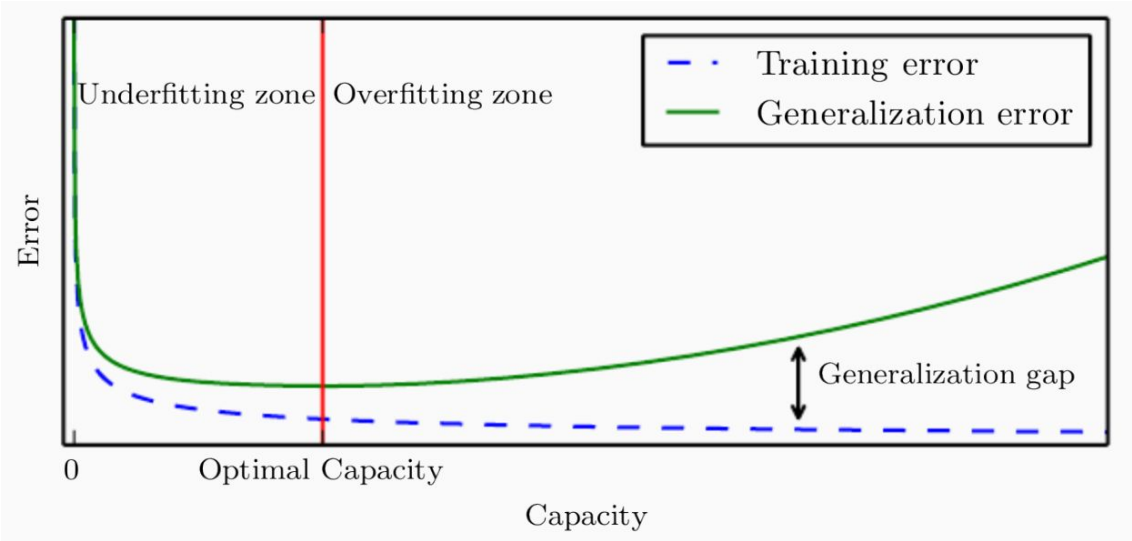
Mean squared error (MSE):

$$J(\boldsymbol{w}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_{\boldsymbol{w}}(\boldsymbol{x}_i))^2$$

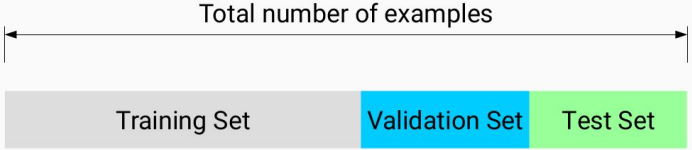
Need of the optimization algorithm!

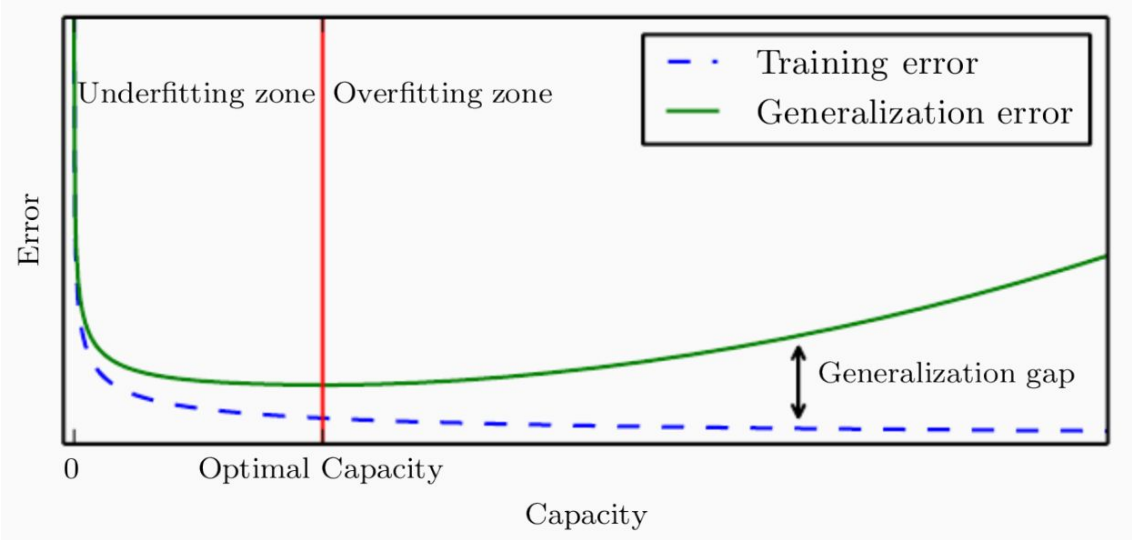
Train, validation and test set





Train, validation and test set





In many cases, part of the dataset is used for **validation** (fine tuning of the hyperparameters).

Convolutional Neural Networks:

https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

Multiwavelength classification of X-ray selected galaxy cluster candidates using CNNs (Kosiba et al. 2020)

Starting point:

ABSTRACT

Galaxy clusters appear as extended sources in *XMM–Newton* images, but not all extended sources are clusters. So, their proper classification requires visual inspection with optical images, which is a slow process with biases that are almost impossible to model. We tackle this problem with a novel approach, using convolutional neural networks (CNNs), a state-of-the-art image classification tool, for automatic classification of galaxy cluster candidates. We train the networks on combined *XMM–Newton*

Aim of the work:

Evaluate the ability of supervised learning with CNNs in classifying cluster and non cluster candidates.

How?

They compare the results obtained with CNNs with classifications made by experts and the *Hunt for Galaxy Clusters* project.

The Hunt for Galaxy Clusters project (link)

2 THE HUNT FOR GALAXY CLUSTERS

Our citizen science project, *The Hunt for Galaxy Clusters*,[1] was launched online as an official Zooniverse project on 2018 October 24. There were 1600 galaxy cluster candidates in the project that have been detected as extended X-ray sources by the `XAmin` wavelet-based pipeline (Pacaud et al. 2006). Each object was classified by at least 30 different volunteers, and this was completed by 2019 April 29. In total, 1227 volunteers participated in the project.

The Hunt for Galaxy Clusters project (link)

2 THE HUNT FOR GALAXY CLUSTERS

Our citizen science project, *The Hunt for Galaxy Clusters*,[1] was launched online as an official Zooniverse project on 2018 October 24. There were 1600 galaxy cluster candidates in the project that have been detected as extended X-ray sources by the `XAmin` wavelet-based pipeline (Pacaud et al. 2006). Each object was classified by at least 30 different volunteers, and this was completed by 2019 April 29. In total, 1227 volunteers participated in the project.

2.1 Data

The data in this work originate from the XMM CLuster Archive Super Survey (X-CLASS; Clerc et al. 2012), an X-ray galaxy cluster search in the archival data of the *European Space Agency's* X-ray observatory *XMM–Newton*, combined with corresponding optical counterparts from the Digitized Sky Survey POSS-II (DSS2). We used *XMM–Newton* data obtained between 2000 and 2015, employing selection criteria described in (Clerc et al. 2012), and excluding the data used by the XXL survey (Pierre et al. 2016).

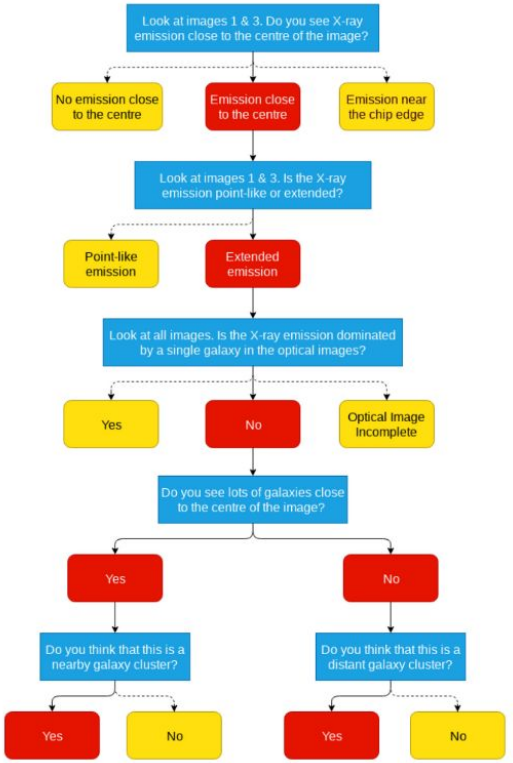
The Hunt for Galaxy Clusters project (link)

2 THE HUNT FOR GALAXY CLUSTERS

Our citizen science project, *The Hunt for Galaxy Clusters*,[1] was launched online as an official Zooniverse project on 2018 October 24. There were 1600 galaxy cluster candidates in the project that have been detected as extended X-ray sources by the `XAmin` wavelet-based pipeline (Pacaud et al. 2006). Each object was classified by at least 30 different volunteers, and this was completed by 2019 April 29. In total, 1227 volunteers participated in the project.

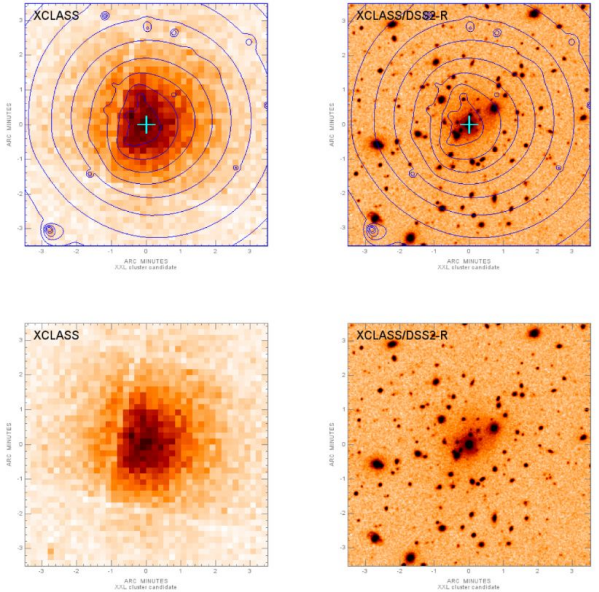
2.1 Data

The data in this work originate from the XMM CLuster Archive Super Survey (X-CLASS; Clerc et al. 2012), an X-ray galaxy cluster search in the archival data of the *European Space Agency's* X-ray observatory *XMM–Newton*, combined with corresponding optical counterparts from the Digitized Sky Survey POSS-II (DSS2). We used *XMM–Newton* data obtained between 2000 and 2015, employing selection criteria described in (Clerc et al. 2012), and excluding the data used by the XXL survey (Pierre et al. 2016).



The Hunt for Galaxy Clusters project (link)

Each volunteer is provided with four 7’ x 7’ images: two X-ray and two optical images.
The project uses six questions to help determine the class of a galaxy cluster candidate.



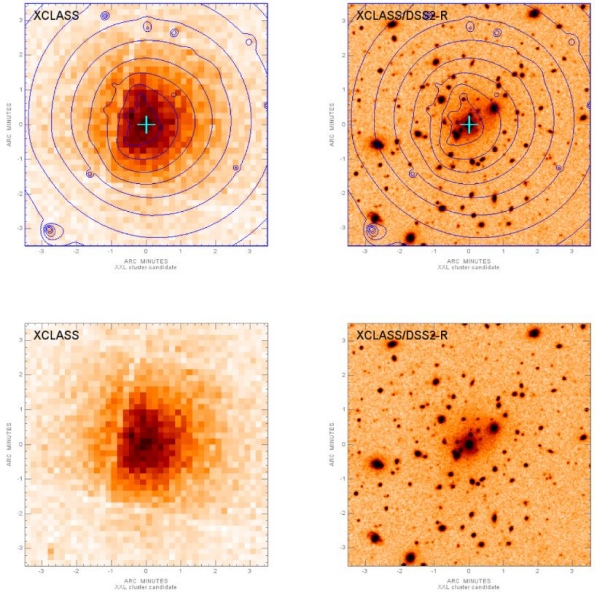
The Hunt for Galaxy Clusters project (link)

Each volunteer is provided with four 7’ x 7’ images: two X-ray and two optical images.
The project uses six questions to help determine the class of a galaxy cluster candidate.

Each object is classified by more than 30 volunteers.
Each classification is weighted according to the agreement of the majority:

$$G_i = \frac{C_i}{Q_i},\ i \in 1, ..., 6, \qquad (1)$$

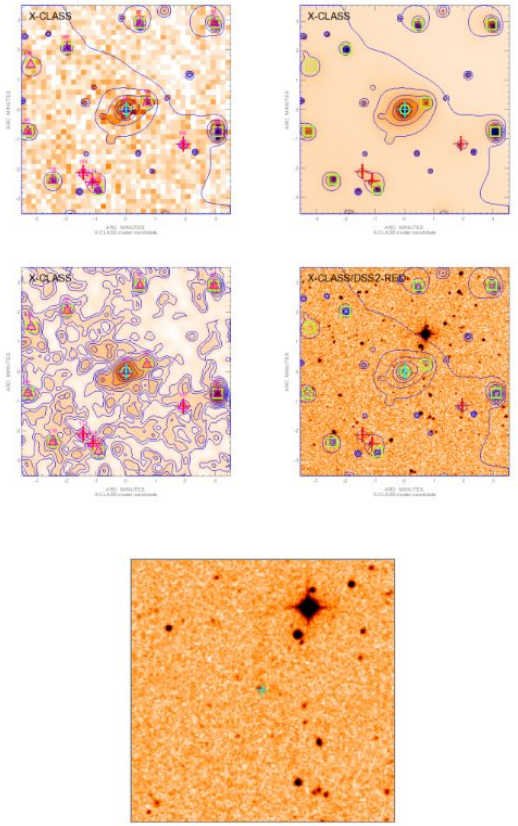
where G_i is the weight applied for an individual on question i, C_i is the number of answers to question i given by the individual that were in agreement with the majority, and Q_i is the total number of answers the individual has made for question i.



Classification by experts

Each galaxy cluster candidate is classified by two experts and three moderators make the final classification on conflicting decisions.

The experts classify objects as: low-redshift cluster ($0 < z < 0.3$), high-redshift cluster ($z > 0.3$), nearby galaxy, point source and some additional subclasses.



Machine learning approach

They use Convolutional Neural Network, with binary (cluster and non-cluster) classification + additional subcategories.

Layer	Layer type	Filter shape/stride	Input shape
1	Conv	3 × 3 × 64/(1, 1)	356 × 356 × 3
2	Max pool	2 × 2/(2, 2)	356 × 356 × 64
3	Conv	3 × 3 × 32/(1, 1)	178 × 178 × 64
4	Max pool	2 × 2/(2, 2)	178 × 178 × 32
5	Conv	3 × 3 × 32/(1, 1)	89 × 89 × 32
6	Max pool	2 × 2/(2, 2)	89 × 89 × 32
7	Conv	3 × 3 × 32/(1, 1)	45 × 45 × 32
8	Max pool	2 × 2/(2, 2)	45 × 45 × 32
9	Conv	3 × 3 × 32/(1, 1)	23 × 23 × 32
10	Max pool	2 × 2/(2, 2)	23 × 23 × 32
11	Conv	3 × 3 × 32/(1, 1)	12 × 12 × 32
12	Max pool	2 × 2/(2, 2)	12 × 12 × 32
13	Flatten	–	6 × 6 × 32
14	Dense	256	1152
15	Dense	2	256

Hyperparameters	Custom net	MobileNet
Batch size	10	20
Iterations	153 000	3825
Optimizer	SGD	Adadelta
Nest. Momentum	0.90	–
Rho	–	0.95
Initial lr.	0.0001	1.0
lr. decay	10^{-6}	0.95
Minimal lr.	10^{-4}	0.01
lr. red. patience	14	4
lr. red. factor	0.75	0.85
Dense dropout	0.65	0.65
Output activation	softmax	softmax
Loss function	cat. cross-entropy	cat. cross-entropy
Input image size	356 × 356	224 × 224

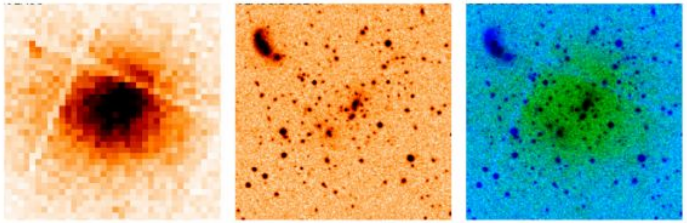
(link)

Machine learning approach

They use Convolutional Neural Network, with binary (cluster and non-cluster) classification + additional subcategories.

For each candidate cluster, a pair of X-ray and optical PNG images were merged into a single PNG image.

They also use data augmentation to reduce the probability of overfitting.



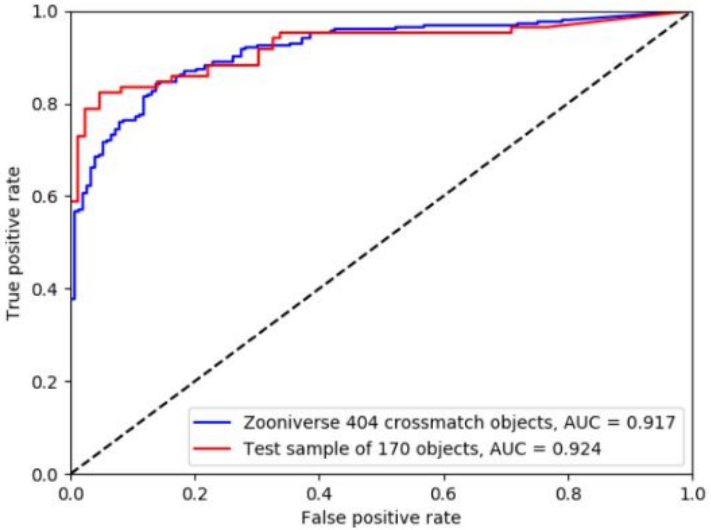
Layer	Layer type	Filter shape/stride	Input shape
1	Conv	3 × 3 × 64/(1, 1)	356 × 356 × 3
2	Max pool	2 × 2/(2, 2)	356 × 356 × 64
3	Conv	3 × 3 × 32/(1, 1)	178 × 178 × 64
4	Max pool	2 × 2/(2, 2)	178 × 178 × 32
5	Conv	3 × 3 × 32/(1, 1)	89 × 89 × 32
6	Max pool	2 × 2/(2, 2)	89 × 89 × 32
7	Conv	3 × 3 × 32/(1, 1)	45 × 45 × 32
8	Max pool	2 × 2/(2, 2)	45 × 45 × 32
9	Conv	3 × 3 × 32/(1, 1)	23 × 23 × 32
10	Max pool	2 × 2/(2, 2)	23 × 23 × 32
11	Conv	3 × 3 × 32/(1, 1)	12 × 12 × 32
12	Max pool	2 × 2/(2, 2)	12 × 12 × 32
13	Flatten	–	6 × 6 × 32
14	Dense	256	1152
15	Dense	2	256

Hyperparameters	Custom net	MobileNet
Batch size	10	20
Iterations	153 000	3825
Optimizer	SGD	Adadelta
Nest. Momentum	0.90	–
Rho	–	0.95
Initial lr.	0.0001	1.0
lr. decay	10^{-6}	0.95
Minimal lr.	10^{-4}	0.01
lr. red. patience	14	4
lr. red. factor	0.75	0.85
Dense dropout	0.65	0.65
Output activation	softmax	softmax
Loss function	cat. cross-entropy	cat. cross-entropy
Input image size	356 × 356	224 × 224

(link)

Performance measurements

ROC curve (link)



$$\mathrm{FPR} = \frac{FP}{FP + TN}$$

Accuracy is the most intuitive performance measurement. It is the ratio of correct predictions to all predictions and is defined as

$$A = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}, \quad (2)$$

where TP refers to the number of true positives, in our case the number of clusters correctly classified as clusters, TN is a number of true negatives (number of non-clusters correctly classified as non-clusters), FP is a number of false positives (number of non-cluster incorrectly classified as clusters), and FN states for a number of false negatives (number of clusters incorrectly classified as non-clusters).

Precision is the ratio of the correctly classified positives (i.e. clusters) and all objects classified as positives. This is defined as

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}. \quad (3)$$

Recall is the ratio of the correctly classified positives and all positives examples in the test data. It is defined as

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \quad (4)$$

The receiver operating characteristic (ROC) is a performance measurement of detection problems plotted as a true positive rate (recall) against the false positive rate, defined as

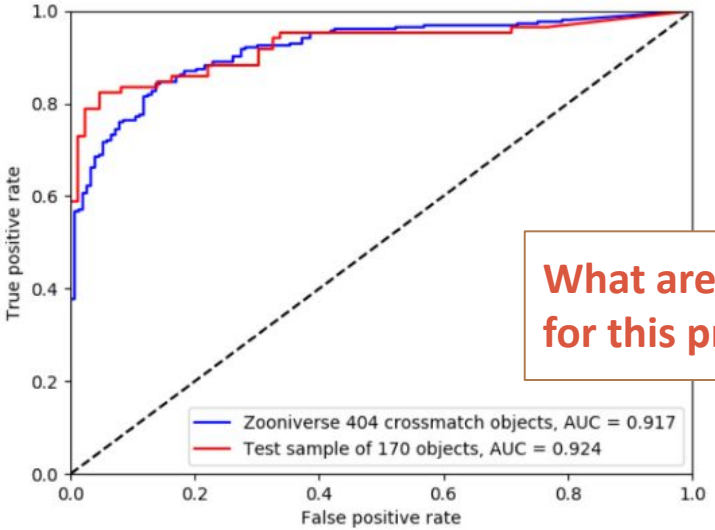
$$\mathrm{FPR} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \quad (5)$$

This is not FPR…

at various thresholds. The area under the curve (AUC) describes the model's capability to distinguish between two classification classes and is independent of the choice of the threshold. When reporting detection performance for a class (from the CNN output) in terms of ROC curve, we compare the posterior probability of the class to a varying detection threshold.

Performance measurements

ROC curve (link)



$$\mathrm{FPR} = \frac{FP}{FP + TN}$$

What are the thresholds for this problem?

Accuracy is the most intuitive performance measurement. It is the ratio of correct predictions to all predictions and is defined as

$$A = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}, \quad (2)$$

where TP refers to the number of true positives, in our case the number of clusters correctly classified as clusters, TN is a number of true negatives (number of non-clusters correctly classified as non-clusters), FP is a number of false positives (number of non-cluster incorrectly classified as clusters), and FN states for a number of false negatives (number of clusters incorrectly classified as non-clusters).

Precision is the ratio of the correct classified positives (i.e. clusters) and all objects classified as positives. This is defined as

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}. \quad (3)$$

is the ratio of the correctly classified positives and all examples in the test data. It is defined as

$$\frac{\mathrm{TP}}{+ \mathrm{FN}}. \quad (4)$$

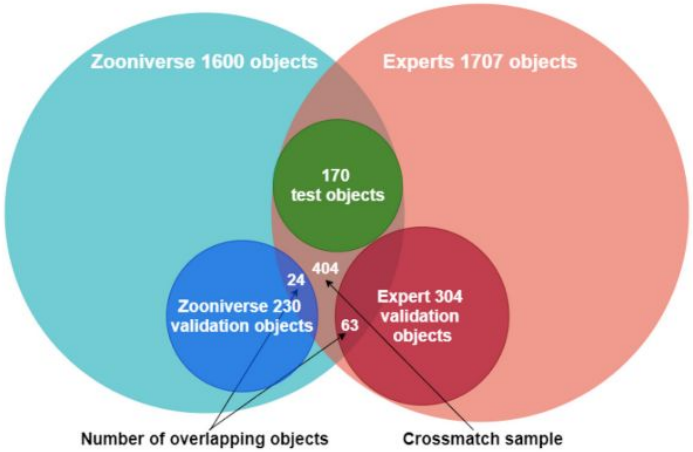
The receiver operating characteristic (ROC) is a performance measurement of detection problems plotted as a true positive rate (recall) against the false positive rate, defined as

$$\mathrm{FPR} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \quad (5)$$

at various thresholds. The area under the curve (AUC) describes the model's capability to distinguish between two classification classes and is independent of the choice of the threshold. When reporting detection performance for a class (from the CNN output) in terms of ROC curve, we compare the posterior probability of the class to a varying detection threshold.

The sample

Class	Zooniverse		Experts		
	Train	Validate	Train	Validate	Test
Cluster	320	130	845	200	85
Non-cluster	880	100	388	104	85
Total	1200	230	1233	304	170



Zooniverse sample: 1600 candidates
Expert sample: 1707 candidates
Crossmatch sample: 404 candidates

Test sample:

- 85 spectroscopically confirmed galaxy clusters
- 85 objects classified as non-clusters by experts

Results - The Hunt for Galaxy Clusters

Assuming that the expert classifications are the ground truth.

The Zooniverse volunteers performed better on the subsample of 170 objects.

$$A = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}},$$

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}.$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}.$$

Data set	Zooniverse classifications	TP	TN	FP	FN	Accuracy	Precision	Recall
404 objects	Unweighted	69	150	0	185	0.542	1.000	0.272
404 objects	Weighted	102	149	1	152	0.621	0.990	0.401
170 objects	Weighted	55	84	1	30	0.818	0.982	0.647

Results - The Hunt for Galaxy Clusters

Assuming that the expert classifications are the ground truth.

The Zooniverse volunteers performed better on the subsample of 170 objects.

In general, the Zooniverse volunteers preferentially classified objects as non-clusters.

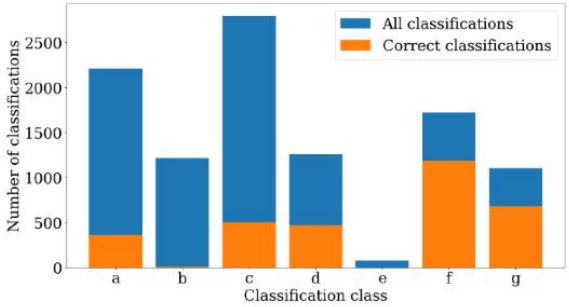


Figure 6. A quantification of the Zooniverse classifications for (a) no emission, (b) edge, (c) point, (d) nearby galaxy, (e) no optical image, (f) nearby galaxy cluster, and (g) distant galaxy cluster, assuming the ground truth is the expert classification.

Data set	Zooniverse classifications	TP	TN	FP	FN	Accuracy	Precision	Recall
404 objects	Unweighted	69	150	0	185	0.542	1.000	0.272
404 objects	Weighted	102	149	1	152	0.621	0.990	0.401
170 objects	Weighted	55	84	1	30	0.818	0.982	0.647

Results - The Hunt for Galaxy Clusters

The galaxy clusters found by Zooniverse volunteers populate all of the space, not showing bias.

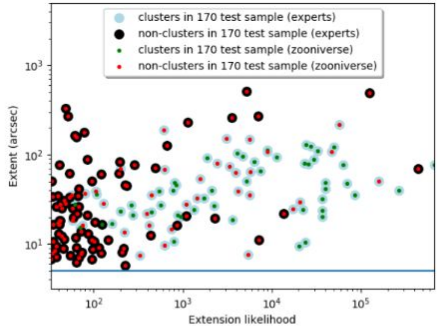


Figure B1. Extent–extension likelihood plane for objects of the 170 test sample classified by experts and the Zooniverse volunteers.

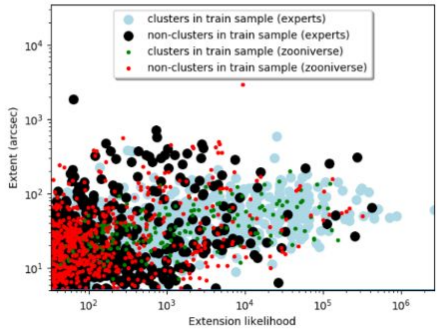


Figure B2. Extent–extension likelihood plane for objects of the experts train sample and the Zooniverse train sample.

Results - The Hunt for Galaxy Clusters

The galaxy clusters found by Zooniverse volunteers populate all of the space, not showing bias.

Even though the Zooniverse volunteers did not show a high accuracy compared to experts, misclassifying many galaxy clusters as other options, the sample of galaxy clusters they selected is pure. This makes us conclude that, via the Zooniverse project, the general public can help scientific research where a very pure sample of galaxy clusters is required, but it did not prove to be helpful in a case where a sample of galaxy clusters should be complete.

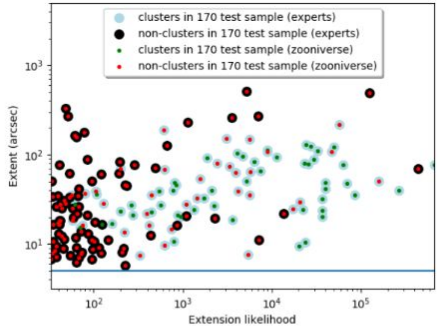


Figure B1. Extent–extension likelihood plane for objects of the 170 test sample classified by experts and the Zooniverse volunteers.

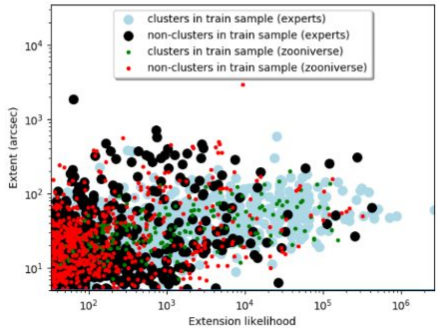


Figure B2. Extent–extension likelihood plane for objects of the experts train sample and the Zooniverse train sample.

Results - The Hunt for Galaxy Clusters

The galaxy clusters found by Zooniverse volunteers populate all of the space, not showing bias.

Even though the Zooniverse volunteers did not show a high accuracy compared to experts, misclassifying many galaxy clusters as other options, the sample of galaxy clusters they selected is pure. This makes us conclude that, via the Zooniverse project, the general public can help scientific research where a very pure sample of galaxy clusters is required, but it did not prove to be helpful in a case where a sample of galaxy clusters should be complete.

What about a ML approach also for the Zooniverse project?

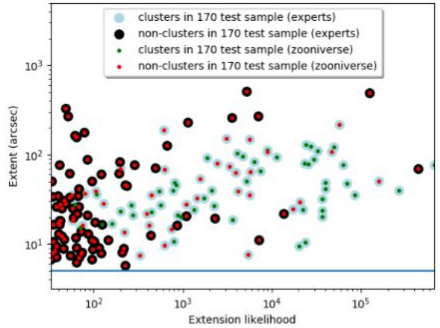


Figure B1. Extent–extension likelihood plane for objects of the 170 test sample classified by experts and the Zooniverse volunteers.

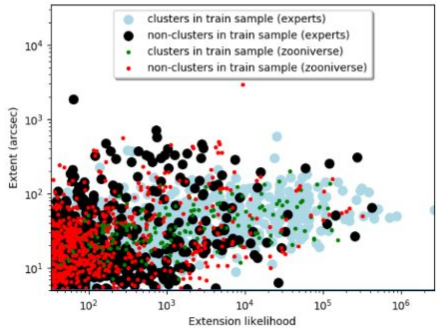


Figure B2. Extent–extension likelihood plane for objects of the experts train sample and the Zooniverse train sample.

Results - Machine learning approach

The best results are obtained with networks trained on the expert classified data set.

Training using the labels obtained in the Zooniverse project resulted in lower performance.

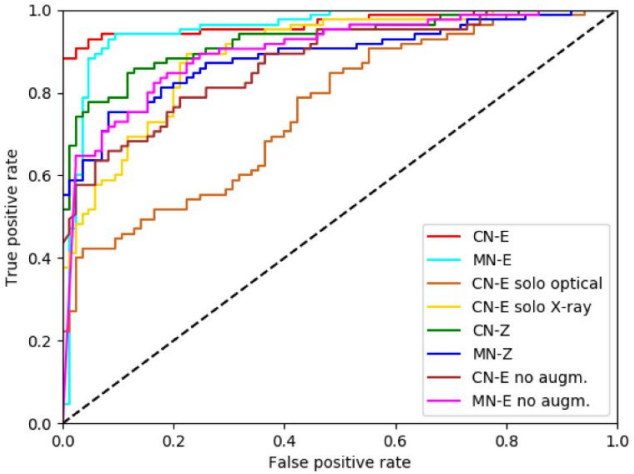
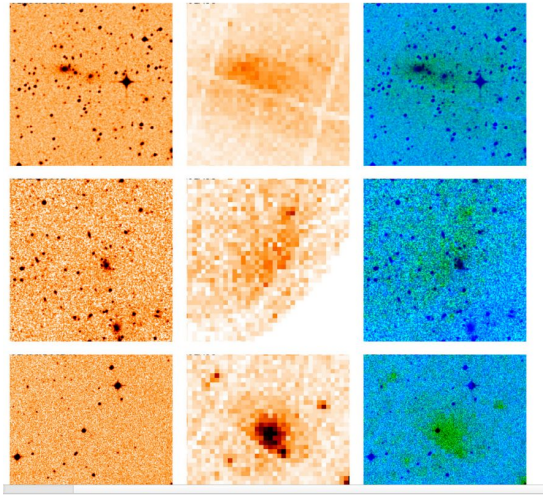


Table 5. Averaged galaxy cluster candidate classification results of the networks each trained 10 times with the exact same hyperparameters, only with a different seed for generation of random numbers during its initialization.

Network	A ± std	P ± std	R ± std	AUC ± std
CN-E	0.90 ± 0.03	0.89 ± 0.05	0.91 ± 0.03	0.96 ± 0.01
MN-E	0.88 ± 0.02	0.87 ± 0.03	0.91 ± 0.03	0.94 ± 0.01
CN-E solo optical	0.68 ± 0.02	0.64 ± 0.02	0.85 ± 0.04	0.77 ± 0.02
CN-E solo x-ray	0.81 ± 0.01	0.78 ± 0.03	0.86 ± 0.04	0.89 ± 0.01
CN-Z	0.82 ± 0.01	0.96 ± 0.01	0.67 ± 0.02	0.91 ± 0.01
MN-Z	0.79 ± 0.02	0.96 ± 0.03	0.62 ± 0.03	0.86 ± 0.02
CN-E no augm.	0.75 ± 0.02	0.70 ± 0.02	0.87 ± 0.03	0.87 ± 0.01
MN-E no augm.	0.81 ± 0.01	0.75 ± 0.02	0.91 ± 0.01	0.90 ± 0.02

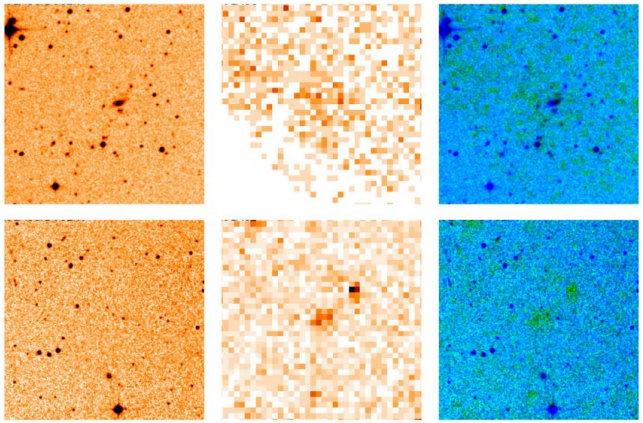
Input images are classified as galaxy clusters if the output probability is higher than 0.5.

Results - Machine learning approach

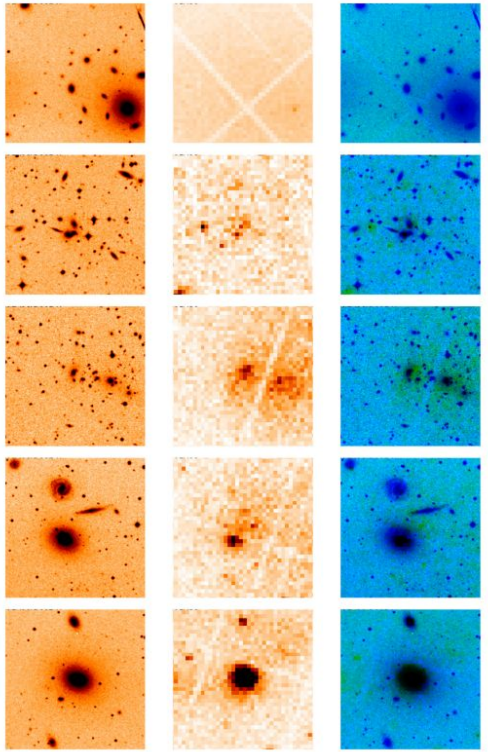


True positives

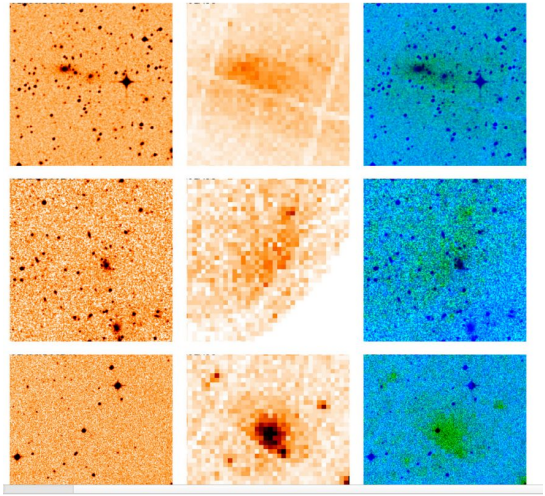
False positives



False negatives

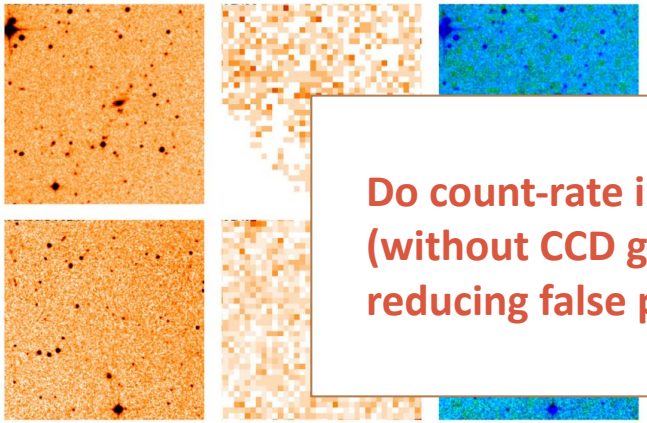


Results - Machine learning approach

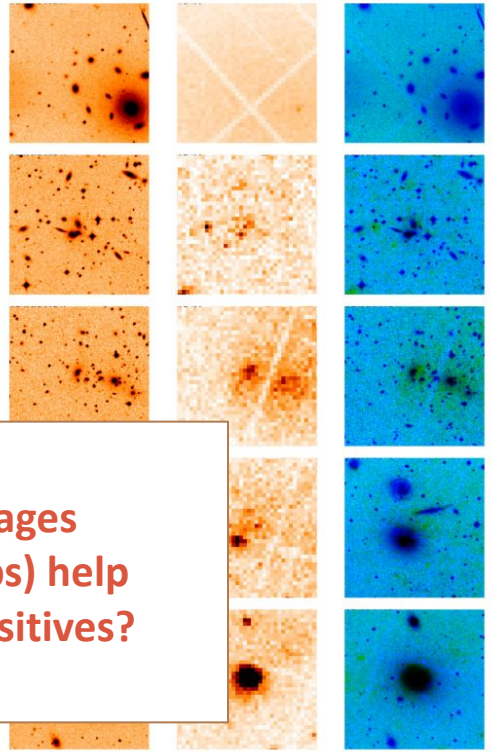


True positives

False positives



False negatives



Do count-rate images (without CCD gaps) help reducing false positives?

Results - Machine learning approach

Multiclass classification.

Average results:

- **MN:** (91±2)% AUC; (86±6)% acc.
- **CN:** (88±2)% AUC; (85±4)% acc.

Table 6. Results from the multiclass classification networks.

Class	*A*	*P*	*R*	AUC
MN greyscale				
Low-z cluster	0.77	0.62	0.94	0.93
High-z cluster	0.87	0.56	0.22	0.91
Point source	0.87	0.88	0.36	0.89
Nearby galaxy	0.90	0.70	0.73	0.92
Other	0.91	0.65	0.68	0.92
CN greyscale				
Low-z cluster	0.79	0.68	0.81	0.89
High-z cluster	0.84	0.44	0.65	0.89
Point source	0.84	0.75	0.27	0.88
Nearby galaxy	0.89	0.74	0.57	0.85
Other	0.87	0.52	0.64	0.88

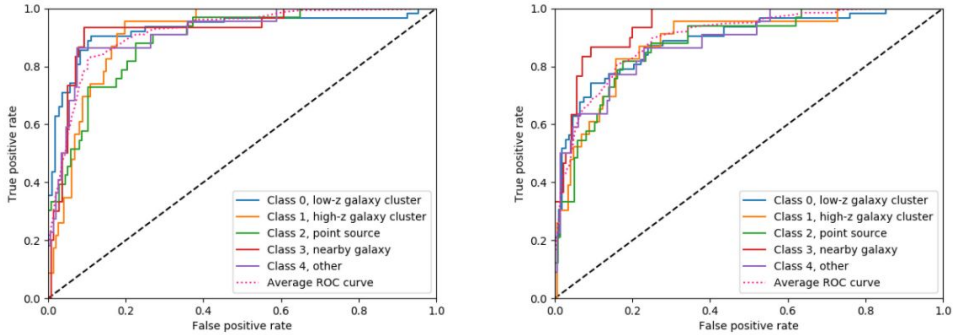


Figure 13. ROC curves for multiclass classification performed by the MobileNet architecture (*left*) and our custom network (*right*).

The ROC curves and performance measurements were calculated as one versus all problems.

Results - Machine learning approach

Multiclass classification.

Why is the AUC for the *Nearby galaxy* class the lowest?

Table 6. Results from the multiclass classification networks.

Class	*A*	*P*	*R*	AUC
MN greyscale				
Low-z cluster	0.77	0.62	0.94	0.93
High-z cluster	0.87	0.56	0.22	0.91
Point source	0.87	0.88	0.36	0.89
Nearby galaxy	0.90	0.70	0.73	0.92
Other	0.91	0.65	0.68	0.92
CN greyscale				
Low-z cluster	0.79	0.68	0.81	0.89
High-z cluster	0.84	0.44	0.65	0.89
Point source	0.84	0.75	0.27	0.88
Nearby galaxy	0.89	0.74	0.57	0.85
Other	0.87	0.52	0.64	0.88

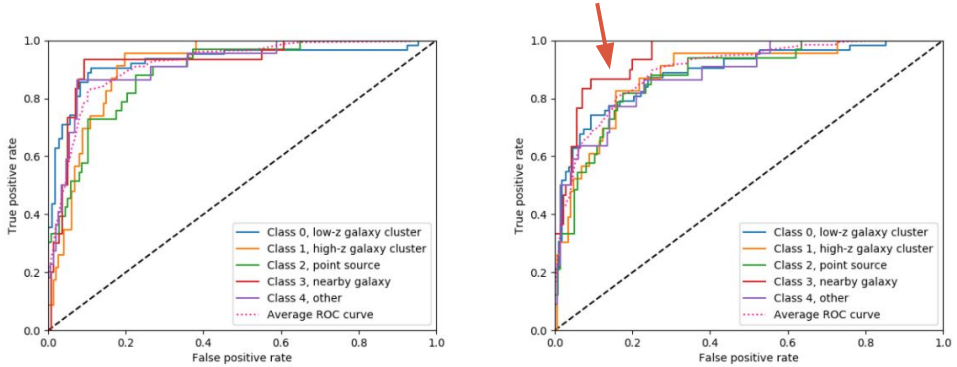


Figure 13. ROC curves for multiclass classification performed by the MobileNet architecture (*left*) and our custom network (*right*).

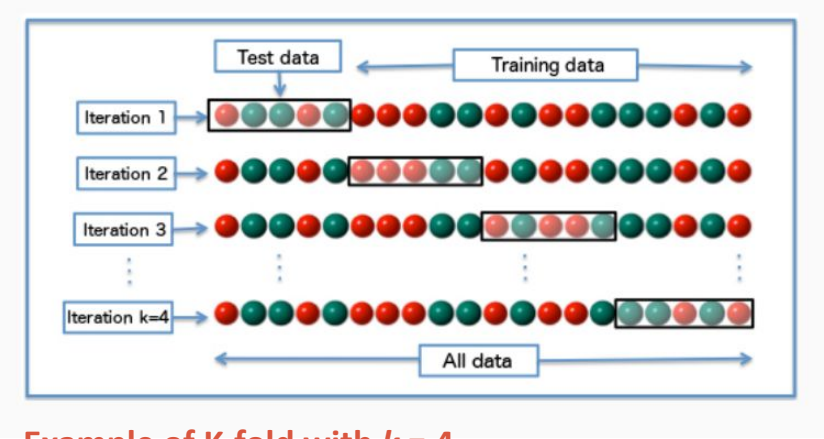
The ROC curves and performance measurements were calculated as one versus all problems.

Results - Machine learning approach

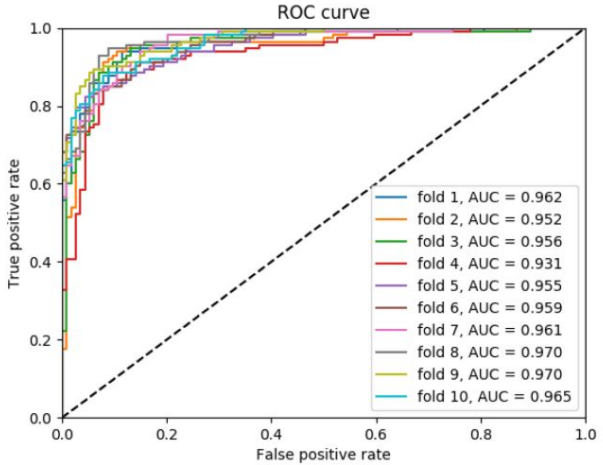
10-fold cross-validation ***(link)***

The original data set is randomly partitioned into *k* equal sized subsample. A single subsample is used as validation, the remaining *k-1* subsamples as training data.

→ **Compute the average over the k results.**



Example of K-fold with *k* = 4



Summary

Summary

The Hunt for Galaxy Cluster Project:

- Zooniverse volunteers classified 1600 cluster candidates, obtaining a 62% agreement with experts.
- The sample selected by Zooniverse volunteers is pure.

Summary

The Hunt for Galaxy Cluster Project:

- Zooniverse volunteers classified 1600 cluster candidates, obtaining a 62% agreement with experts.
- The sample selected by Zooniverse volunteers is pure.

Machine learning approach:

- The custom network obtained an average accuracy of 90% (in agreement with 10-fold cross-validation).
- The custom network gives good results also for multiclass classification.

A deep learning view of the census of galaxy clusters in IllustrisTNG (Su et al. 2020)

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.

This work presents a *"novel method (ML approach) of using X-ray images to identify cool core (CC), weak cool core (WCC) and non cool core (NCC) galaxy clusters".*

The sample

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.

The radiative cooling time is defined as:

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

Classifications obtained through the cooling time are the true labels for CNN training.

Average t(cool) are calculated from a 3D sphere within 0.012 R500 (Barnes et al. 2018)

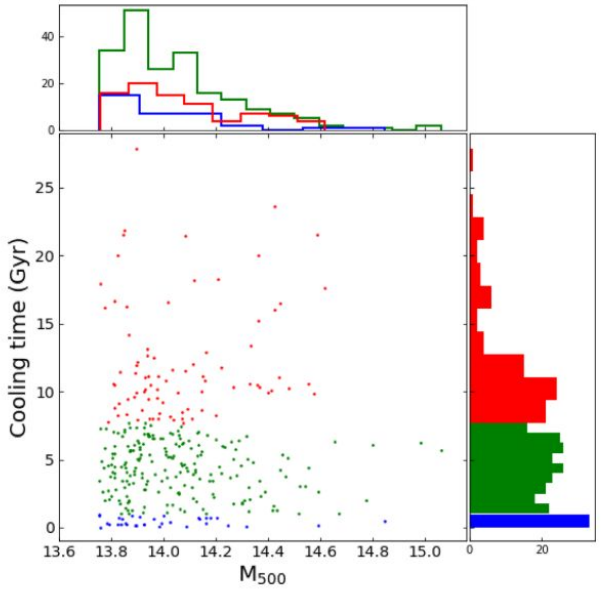


Figure 1. Distributions of central cooling times and $\log M_{500}/M_\odot$ of TNG300 clusters in our sample. Their central cooling times are in the range of 0.012–27.85 Gyr. We define CC and NCC clusters as those with cooling times shorter than 1 Gyr and longer than 7.7 Gyr, respectively. Clusters with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

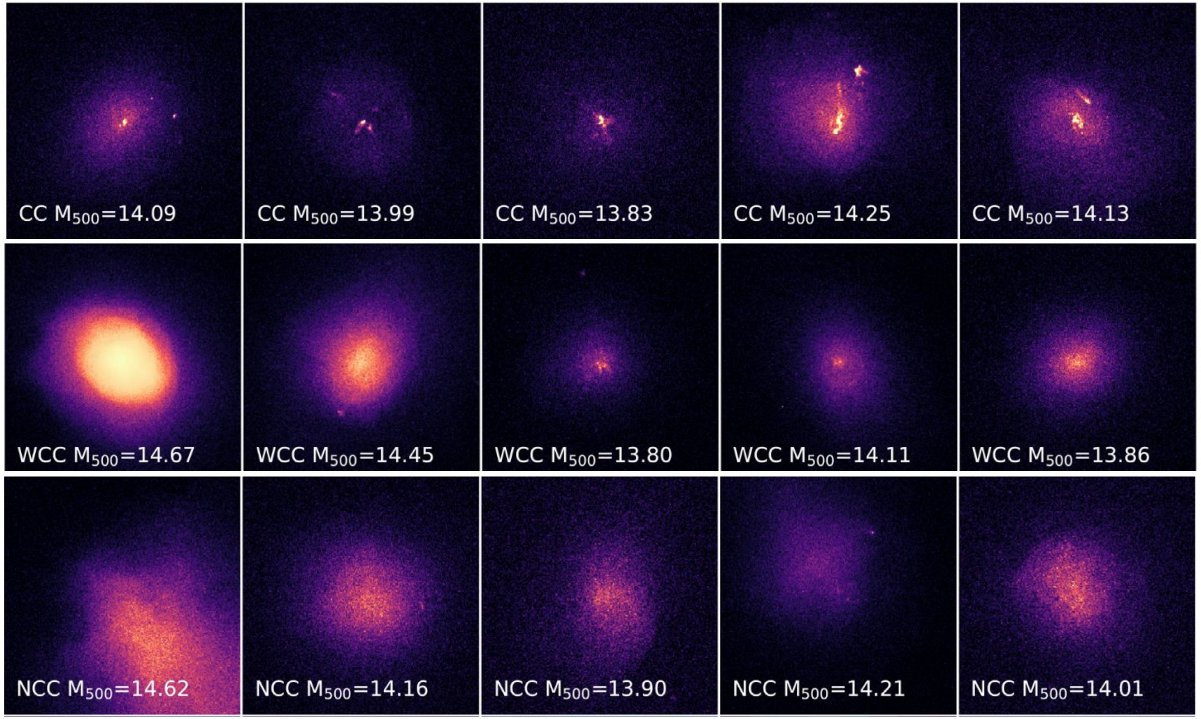
Barnes et al. 2018

The sample

CC

WCC

NCC



Final mock images have a dimension of 256 x 256 pixel^2

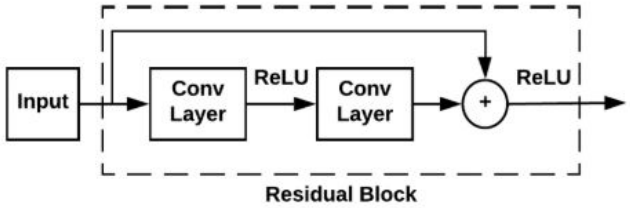
The spatial resolution is degraded to 3.9’’/pixel (8 times worse than Chandra ACIS resolution)

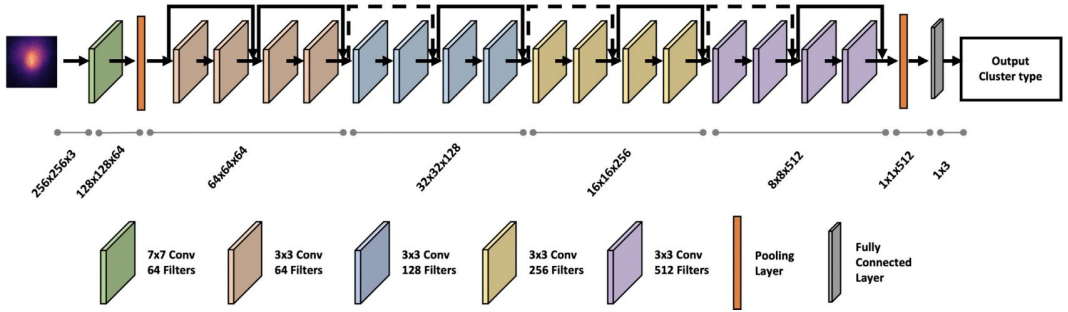
CNN architecture and training

The ResNet-18 network expects a 3-channel input image: each image is replicated three times to form a 256 x 256 x 3 image.

They apply a 10-fold cross-validation (8/10 training, 1/10 validation, 1/10 test).

They use data augmentation.





Model performance

As indicators for model performance, they use: precision, recall, BAcc, F1-score

Confusion matrix

True Positives (TP) e.g. 8	False Positives (FP) e.g. 2
False Negatives (FN) e.g. 4	True Negatives (TN) e.g. 20

Ideally, the confusion matrix is diagonal

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:

$$\text{Precision} = \frac{tp}{tp + fp}. \quad (3)$$

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

$$\text{Recall} = \frac{tp}{tp + fn}. \quad (4)$$

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

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (5)$$

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*:

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

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

Other conventional methods for cluster classification

Density

Concentration

Other conventional methods for cluster classification

Density

A rapidly cooling core implies a high central gas density.

The central electron number density is calculated within 0.012 R500 (Barnes et al. 2018):

- **CC**: ne > 0.015 cm-3
- **WCC**: 0.005 < ne (cm-3) < 0.015
- **NCC**: ne < 0.005 cm-3

Following: Barnes et al. 2018; Hudson et al. 2010

Concentration

Other conventional methods for cluster classification

Density

A rapidly cooling core implies a high central gas density.

The central electron number density is calculated within 0.012 R500 (Barnes et al. 2018):

- **CC**: ne > 0.015 cm-3
- **WCC**: 0.005 < ne (cm-3) < 0.015
- **NCC**: ne < 0.005 cm-3

Following: Barnes et al. 2018; Hudson et al. 2010

Concentration

The elevated ICM metallicity and density at the center of CCs produce a central peak in X-ray surface brightness.

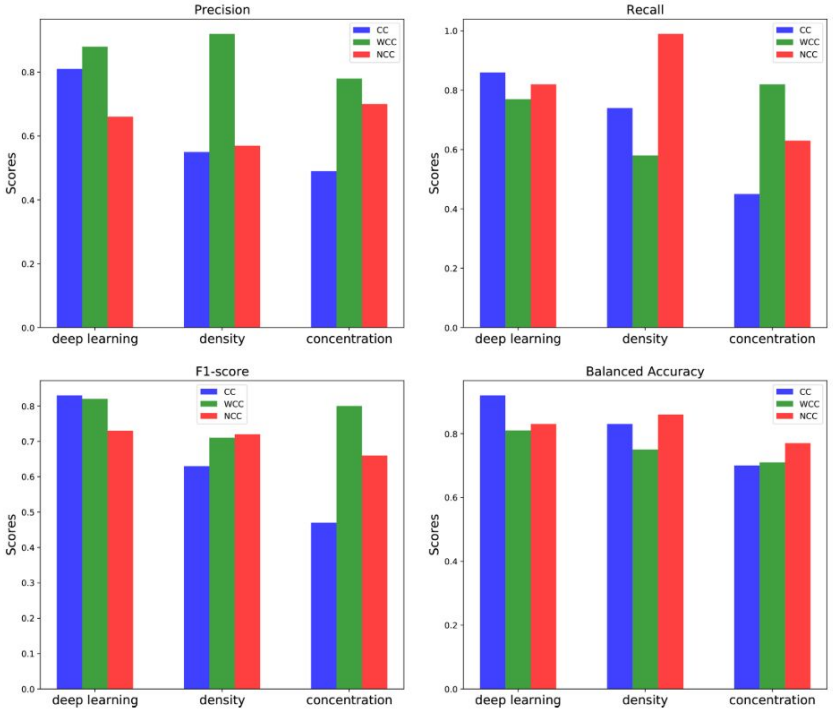
Concentration parameter (Santos et al. 2008):

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

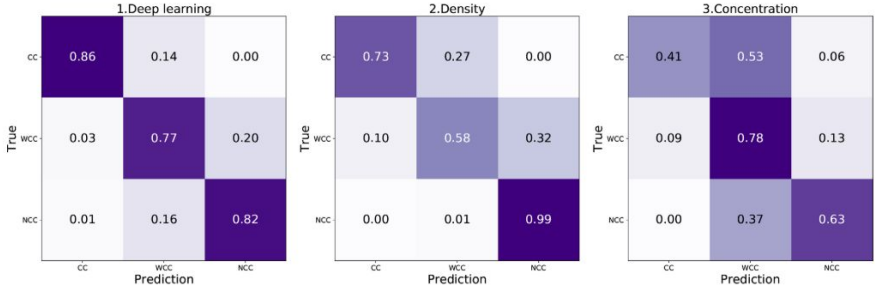
- **CC**: C > 0.155
- **WCC**: 0.075 < C < 0.155
- **NCC**: C < 0.075

Following: Barnes et al. 2018; Andrade-Santos et al. 2018

Results



The Deep learning approach leads to the highest F1-score!



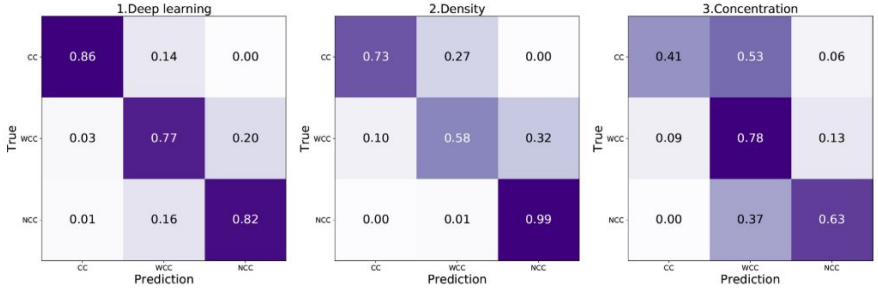
Method	Ave. F_1	Ave. BAcc	Class	Precision	Recall	F_1	BAcc
Deep learning	0.79	0.85	CC	0.81	0.86	0.83	0.92
			WCC	0.88	0.77	0.82	0.81
			NCC	0.66	0.82	0.73	0.83
Density	0.69	0.81	CC	0.55	0.74	0.63	0.83
			WCC	0.92	0.58	0.71	0.75
			NCC	0.57	0.99	0.72	0.86
Concentration	0.64	0.73	CC	0.49	0.45	0.47	0.70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

Results

The Deep learning approach leads to the highest F1-score!

On concentration-based classification…

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_{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



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

Class activation mapping

For each input image, the authors generated a Class Activation Map (CAM) to localize features that are most useful for the network.

Regions that are brighter are more informative for the network.

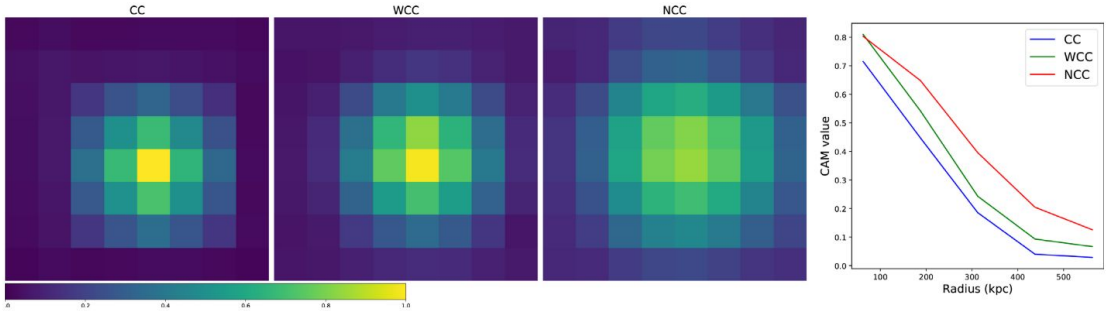
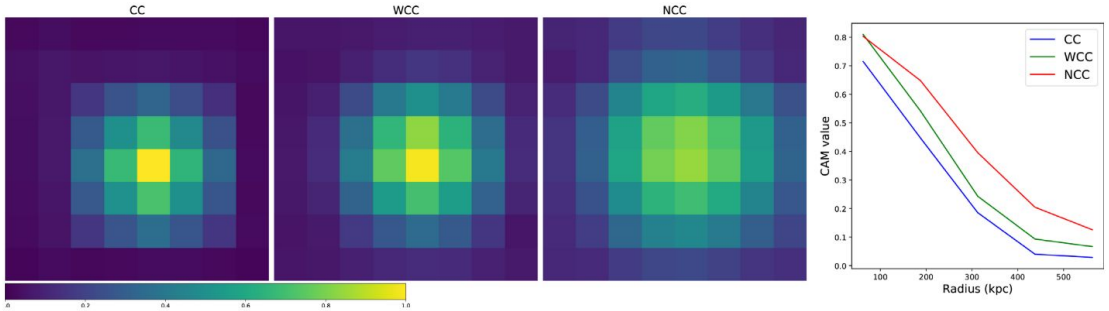


Figure 9. Class activation maps of the central $D = 1$ Mpc averaged over CC, WCC, and NCC clusters, respectively. All these clusters are predicted correctly with a probability above 0.9. The right-hand panel shows the radial profiles of the three CAM maps. The network utilizes relatively more information 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.

Class activation mapping

For each input image, the authors generated a Class Activation Map (CAM) to localize features that are most useful for the network.

Regions that are brighter are more informative for the network.

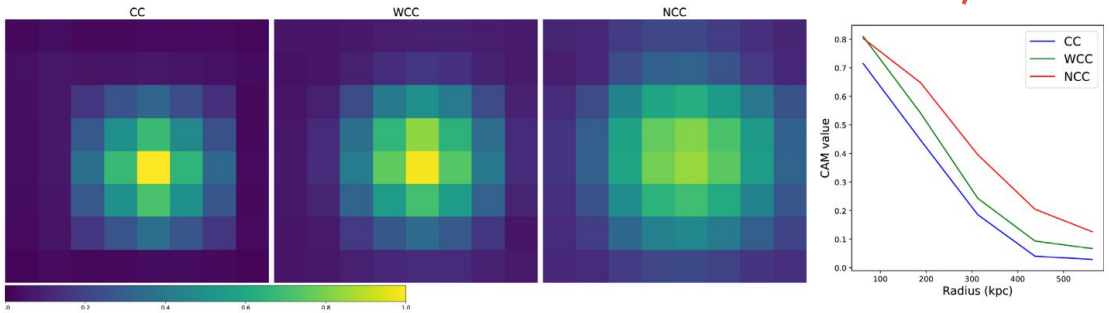


The network uses regions within r ≈ 300 kpc and r ≈ 500 kpc to identify CC and NCC clusters, respectively.

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.

Class activation mapping

Should the integral of these curves have the same value?



The network uses regions within r ≈ 300 kpc and r ≈ 500 kpc to identify CC and NCC clusters, respectively.

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.

Summary

- This paper shows a novel method for classifying CC, WCC and NCC clusters, from their X-ray images.
- ResNet-18 achieves an average precision, recall, F1-score, and BAcc of 0.78, 0.82, 0.79, and 0.85, respectively, well above a random prediction of 0.33.
- The deep learning algorithm outperforms the estimates given by the central gas densities and surface brightness concentration parameters.
- 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.