Dark matter halo analysis on N-body cosmological simulations with deep learning

Jazhiel Chacón Lavanderos[1,2], Isidro Gómez-Vargas[2], José Alberto Vázquez[2] and Jesús Alberto Martínez Castro[1]

[1]Centro de Investigación en Computación, IPN. [2]Instituto de Ciencias Físicas, UNAM. 28 de noviembre de 2022

VIII Essential Cosmology for the Next Generation

Introduction

Cosmological simulations are an important key to fully understand the nature of the evolution of the Universe. With the properties of the dark matter density field in the initial conditions and the final output of dark matter halos, we deployed a variety of Machine Learning algorithms to infer whether or not dark matter particles, traced back to the initial conditions, would fall into halos whose masses are above a given threshold. This problem might be posed as a binary classification task, where the initial conditions of the dark matter density field are mapped into classification labels provided by a halo finder software.

Training Process

The dataset consists of one million data points, each point being a dark matter particle associated with a 10-vector component, along with their respective label, whether the dark matter particle belongs to a halo given the mass threshold of ($10^{11} \leq M/M_{\odot} \leq 10^{14}$).

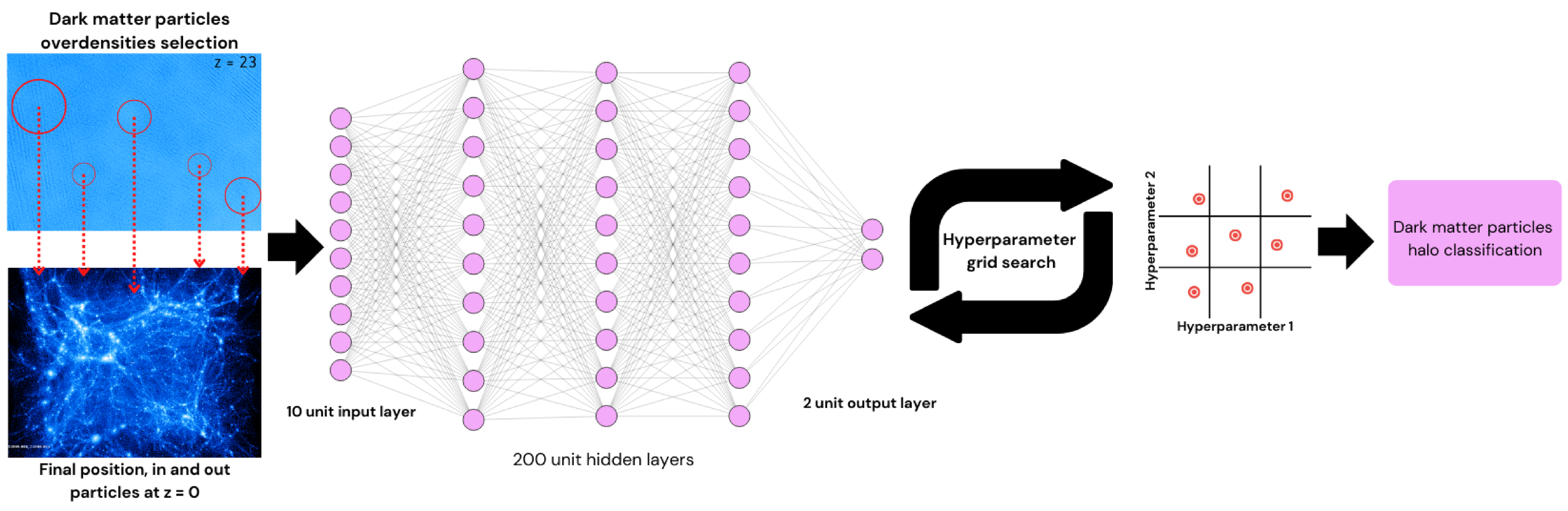
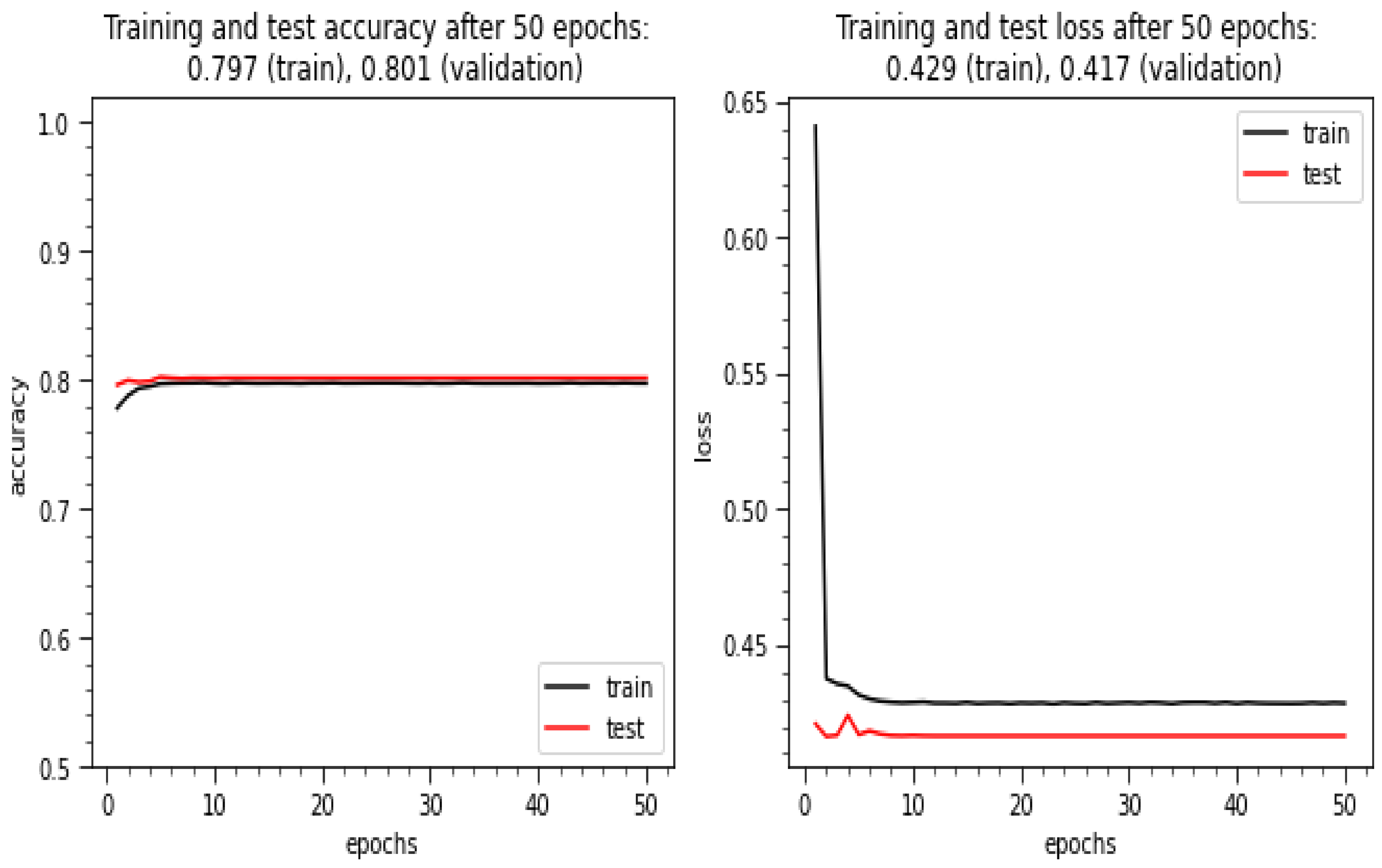


Figure 1: Architecture for our dark matter particle classification task. Features are selected in the initial conditions of the simulations, and we select ten components to train our ML algorithms.

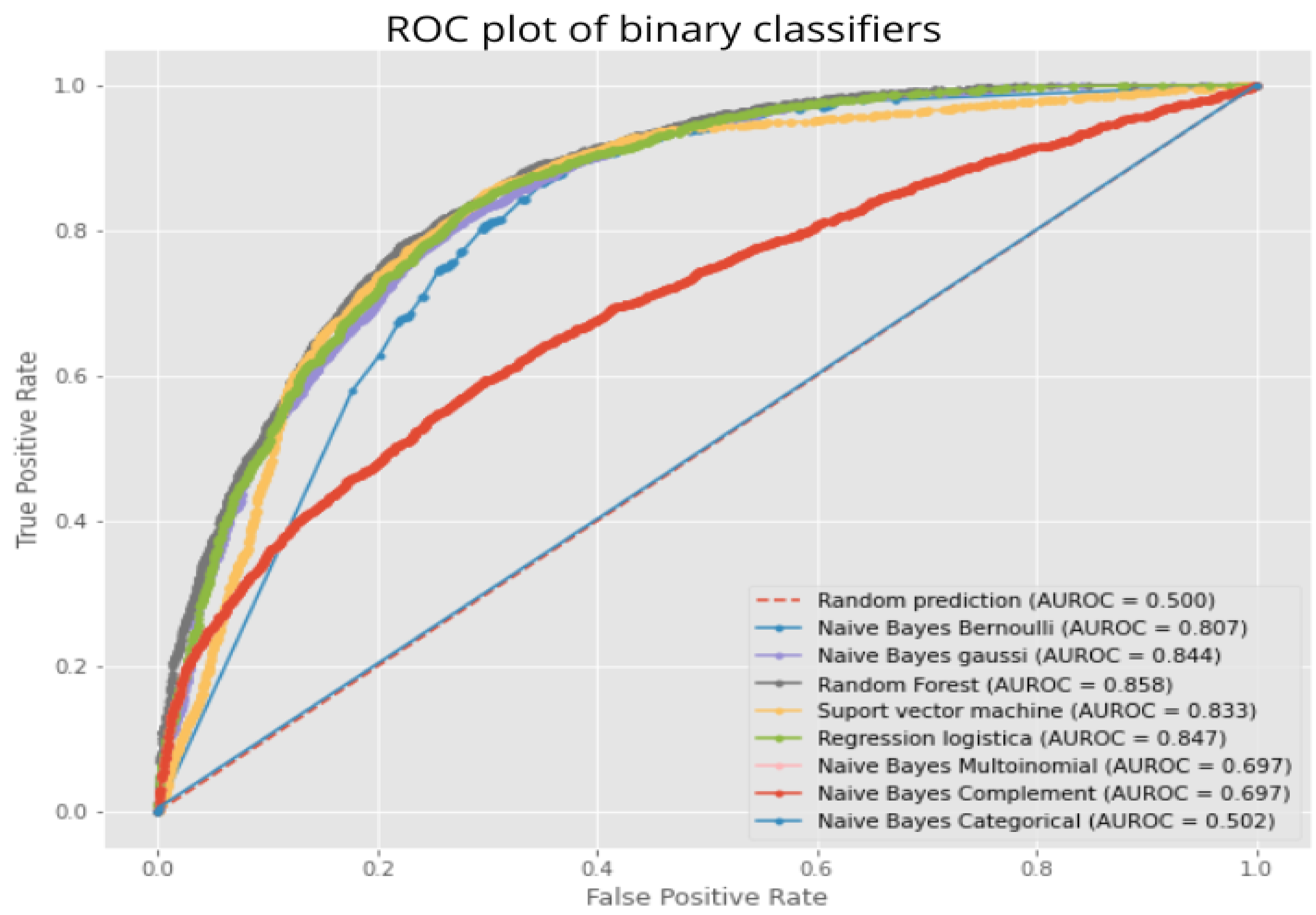


Metrics

Classifier	Accuracy	Precision	Recall	F_1	Time (s)
Logistic Regression	0.722	0.498	0.711	0.727	0.23
NB Gaussian	0.718	0.516	0.757	0.748	0.29
NB Bernoulli	0.699	0.495	0.757	0.730	0.26
NB Multinomial	0.611	0.316	0.554	0.527	0.18
NB Complement	0.608	0.316	0.554	0.527	0.18
NB Categorical	0.699	0.370	0.578	0.568	0.39
SVM	0.723	0.473	0.753	0.702	51
Random Forest	0.785	0.767	0.792	0.792	120
MLP	0.806	0.764	0.794	0.794	800

Training

We used the ROC curve in our ML algorithms which tells us the trade off between TPR and FPR on the test set.



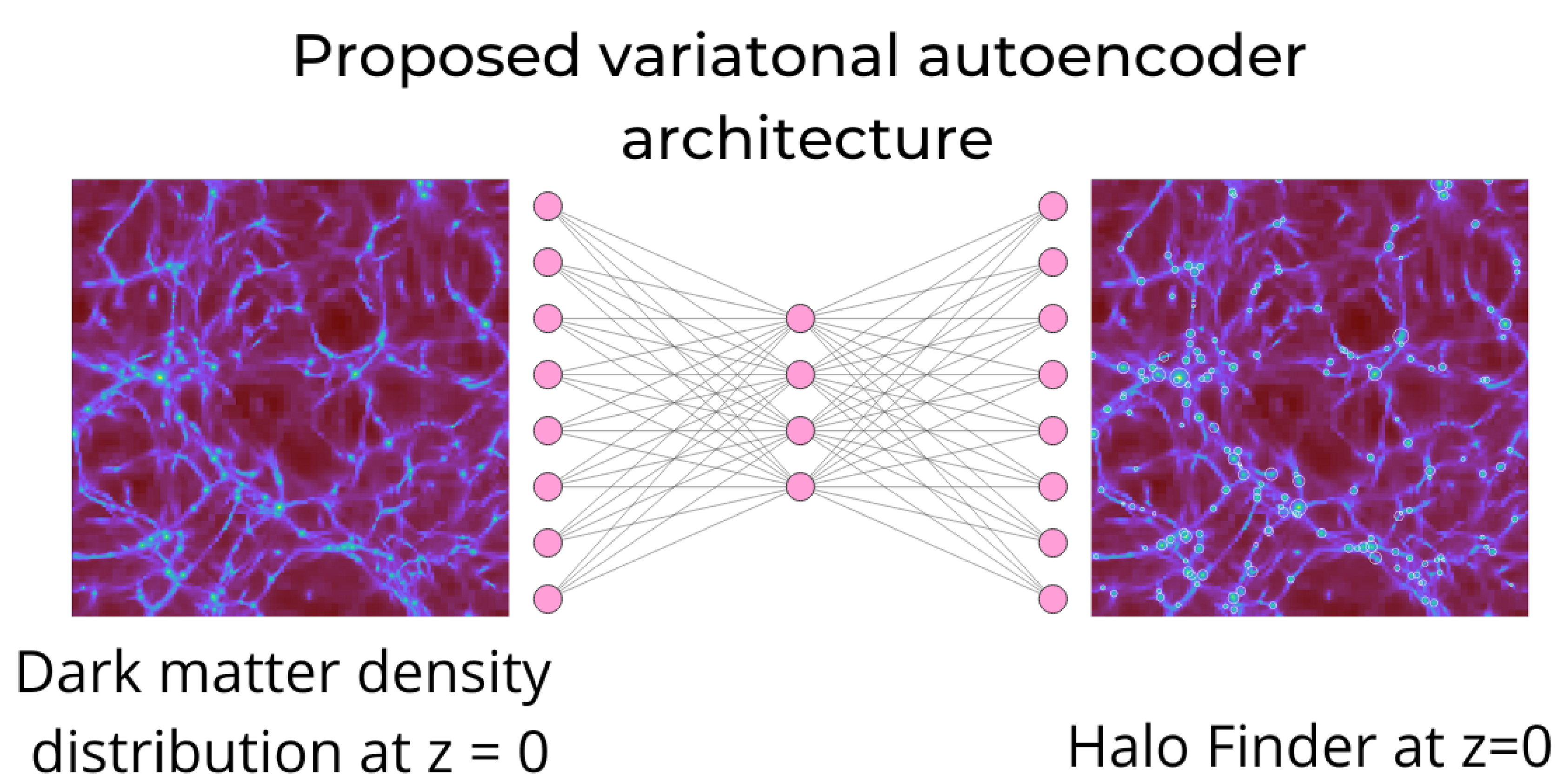
Conclusion

Encouraged by the good results observed in our previous work, we increased the classification algorithms as well as the dataset from a N-body cosmological simulation. Random forest and neural networks were the best performing models.

Our dataset, consisting on the feature selection of the initial conditions of the dark matter density field together with the final halo formation has enough information to provide insight about the algorithms used. Using the physical properties surrounding the initial conditions of the density field as features of our ML pipeline we are able to predict whether or not a dark matter particle will end up in a halo of a given mass threshold. With this result, we observed an improvement in performance compared to our previous results.

Future work

Using a denoising autoencoder architecture, we can feed the network with dark matter simulation outputs at any given redshift and teach the network how to find properties and characteristics of a certain cluster of pixels, this eventually will be part of a halo finding method model.



References

- Chacón, J., Vázquez, J.A. and Almaraz, E. (2022) https://doi.org/10.1016/j.ascom.2021.100527