Deep Learning Semi-Supervised Strategy for Gamma/Hadron Classification of Imaging Atmospheric Cherenkov Telescope Events*

Diego Riquelme[1][a], Mauricio Araya[1], Sebastian Borquez[1], Boris Panes[2] and Edson Carquin[1]

[1]*Universidad Técnica Federico Santa María (for the CTA Consortium), Av. España 1680, Valparaíso, Chile*

[2]*Instituto de Astrofísica, Pontificia Universidad Católica de Chile, Avenida Vicuña Mackenna 4860, Santiago, Chile*

Keywords: Deep Learning, Self-Supervised, Cherenkov, Classification, High-Energy Astronomy, Convolutional.

Abstract: The new Cherenkov Telescope Array (CTA) will record astrophysical gamma-ray events with an energy coverage range, angular resolution, and flux sensitivity never achieved before. The Earth's atmosphere produces Cherenkov's light when a shower of particles is induced by a high-energy particle of astrophysical origin (gammas, hadrons, electrons, etc.). The energy and direction of these gamma air shower events can be reconstructed stereoscopically using imaging atmospheric Cherenkov detectors. Since most of CTA's scientific goals focus on identifying and studying Gamma-Ray sources, it is imperative to distinguish this specific type of event from the hadronic cosmic ray background with the highest possible efficiency. Following this objective, we designed a competitive deep-learning-based approach for gamma/background classification. First, we train the model with simulated images in a standard supervised fashion. Then, we explore a novel self-supervised approach that allows the use of new unlabeled images towards a method for refining the classifier using real images captured by the telescopes. Our results show that one can use unlabeled observed data to increase the accuracy and general performance of current simulation-based classifiers, which suggests that continuous improvement of the learning model could be possible under real data conditions.

1 INTRODUCTION

The Cherenkov Telescope Array (The CTA Consortium, 2019) [1] (CTA) is a ground-based assemblage of tens of Imaging Atmospheric Cherenkov Telescopes (IACT). These telescopes were designed to study high and very high energy (>20 GeV up to 300 TeV) gamma rays. Since gamma rays are produced in violent, highly active regions of the universe, such as supernovae remnants, pulsar wind nebulae, and supermassive black holes in distant galaxies, their study could provide important information about these sources. Gamma rays interact with the Earth's atmosphere, generating a cascade of secondary particles called an extensive air shower. The highly relativistic process stimulates the emission of Cherenkov light in the atmosphere, which is reflected by the telescope's mirrors into a high-speed camera sensor, effectively imaging the shower of particles.

But Cherenkov light is not exclusive to gamma-ray showers. Cherenkov light events observed by the telescopes are dominated by extensive air showers produced by cosmic rays, whereas gamma-ray detections are only a fraction. CTA will provide a factor of 5 to 10 improvement in sensitivity compared to current IACT telescopes, which requires an efficient rejection of cosmic rays in a highly imbalanced classification scenario. Since the particle level processes underlying the generation of both types of extended showers are not identical, the morphology of both extensive air showers is also different. Existing IACTs can reject the background noise by reducing the image into geometrical parameters that are handed to classical classifiers, such as Random Forest (RF) (Albert, 2007) or Boosted Decision Trees (BDT) (Krause et al., 2017; Becherini et al., 2011; Ohm et al., 2009).

Following the current trends in digital image processing, deep-learning techniques have also been explored in high-energy physics. In particular, convolutional neural networks (CNN) use the information of raw images, which potentially provides an advantage to image parameterization as it allows for processing whole event images at high speed. This line of work has been studied, and promising results are already available (Grespan et al., 2021; Mangano et al., 2018; Miener et al., 2021b; Nieto et al., 2019; Nieto et al., 2017; Shilon et al., 2019; Jacquemont et al.,

[a] https://orcid.org/0000-0003-3363-7720

*For the CTA Consortium

[1] https://www.cta-observatory.org/

Riquelme, D., Araya, M., Borquez, S., Panes, B. and Carquin, E.
Deep Learning Semi-Supervised Strategy for Gamma/Hadron Classification of Imaging Atmospheric Cherenkov Telescope Events.
DOI: 10.5220/0011611500003411
In *Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2023)*, pages 725-732
ISBN: 978-989-758-626-2; ISSN: 2184-4313

2019; Juryšek et al., 2021). Most of these advances use Monte Carlo simulations of the showers to train and assess supervised models. Simulations are useful since they can be generated in any desired proportion, number of events, and different configurations. But once real data becomes available, an adaptive strategy is needed to train the models further.

In this work, we present a proof of concept for allowing semi-supervised training once real data become available in CTA, in quantities that will enable proper training and testing. This is relevant because real observations are naturally unlabeled in the context of IACTs, as there is no other independent nor systematic way to confirm the type of the detected particle. Moreover, several deep learning classification approaches proposed so far could benefit from our strategy of further training with unlabeled real data. So our contribution is complementary to previous efforts in the field. Also, we propose analyzing the latent space of the deep learning model to understand the properties of the input data parameters in a lower dimensional representation. This may give us insights into the connection to parametric descriptions of the input data (such as Hillas parameters) and may impact the speed of effective models for rapid classification tasks.

This paper is organized as follows. First, a general idea of the current research on gamma/hadron classification is given. Then, we discuss the data features and introduce the real data problem. Subsequently, we describe our CNN architecture and its features. Later, the experiments are described. Ensuing, the results are shown. Finally, we briefly explore the Latent Space and propose some interesting lines of work.

2 DEEP LEARNING FOR GAMMA/HADRON CLASSIFICATION

2.1 Related Work

As mentioned, previous existing Cherenkov telescope arrays have developed techniques based on machine learning: VERITAS and H.E.S.S use Boosted Decision Trees (BDT) while MAGIC uses Random Forests (RF) (Becherini et al., 2011; Ohm et al., 2009; Krause et al., 2017; Albert, 2007), both based on Hillas parameterization (Hillas, 1985). These techniques have reasonable performance, but new approaches based on artificial neural networks, such as CNN, promise a leap forward by directly using the charge and arrival time of pixels of the images captured by the telescope. In fact, some authors have already studied the effectiveness of deep learning approaches tested on real data. Pipelines that include a framework for deep learning techniques are being designed and used for stereoscopic reconstruction (Miener et al., 2021b). While most studies were conducted with simulated data, there are some of them that compares with real data (Shilon et al., 2019; Miener et al., 2021a; Vuillaume et al., 2021) or trained/tested with a combination of both (Albert, 2007; Lu, 2013).

Recently, some preliminary efforts have been made in the domain adaptation field in the context of gamma/hadron separation (Drew, 2021). They confront the problem with real data available from different observatories.

2.2 Convolutional Neural Networks with CTA Data

For this work we used simulated data (Bernlohr, 2008; Heck et al., 1998) for the CTA south site(The CTA Consortium, 2019). In particular, we used the Prod5 simulation dataset to enable reproducible research; further information can be found on CTA's website. The dataset was subdivided into training and testing. Training classes are balanced, while testing has twice the number of protons than gamma events. This was done to emulate class imbalance in real scenarios, even though real class imbalance is much larger. For instance, the event acquisition frequency in VERITAS is mainly dominated by protons (350 Hz), while gamma occurrence is 1 Hz (Krause et al., 2017). For the classification task, protons and diffuse gamma sources are commonly used. Ideally, every other particle could be used for training, but since protons are the dominant source, is common to use them as the negative class in the classification task.

In (Grespan et al., 2021), a CNN was trained for reconstruction, obtaining state-of-the-art performance. Similarly, in (Shilon et al., 2019), the authors validate deep learning techniques for classification. Also, the authors explore the impact of feeding real data to a model trained with simulated data. They concluded that there is a significant performance loss, which motivates a more detailed study on using real data when available.

CTA models used for event reconstruction are expected to be enhanced thanks to information from real data. Still, since we cannot be truly sure about the actual source of the shower, we can't rely on the existence of labelled real data. This is why we propose a self-supervised step for training a CNN model. Since it can only be trained with simulated data, a semi-supervised method could enable a model to use real data further when they become available to improve the latent space representation of the events.

We used two quality cuts to limit the number of events and filter poor-quality images. These quality cuts are based on the Hillas intensity parameter, which comprises the total charge in photo-electrons (phe) captured by the sensors of the telescope (Grespan et al., 2021). The two cuts are (1) events with a Hillas intensity parameter higher than 1000 phe (reconstructed photo electron) and (2) events higher than 300 phe. This allows us to learn a model with filtered high-intensity images and then expose the model to lower-intensity images unseen in the initial training for a self-supervised refinement of the networks.

Even though there are recent advances on how to perform stereoscopic reconstruction with heterogeneous telescopes, (Brill et al., 2019; Miener et al., 2021b; Shilon et al., 2019), our proposal focuses on single-telescope images from LST telescopes. However, since our proof of concept uses the available data rather than a novel architecture, this technique could also be used over heterogeneous stereoscopic classification techniques.

Specifically, we used a small and straightforward CNN model. Activations were ReLu, and hyperparameters were tuned with a grid search methodology. We also use Cyclical Learning Rate (CLR), which leads to faster convergence, and better metrics (Smith, 2017). All models were trained on Wilkes 3 HPC, with A100 GPU, using the Tensorflow package (Developers, 2022). We mounted the model over *GERUMO* [2] pipeline, and thus, all the hyperparameters and training information are available in the repository. Layer shapes and dimensions of the CNN architecture employed are shown in Table 1. As for preprocessing, we only normalized the images.

Table 1: Summary of CNN Architecture.

Layer	Shape
Input	(55,47,3)
ConvLayer	(53,45,64)
ConvBock	(26,22,128)
ConvBlock	(13,11,256)
ConvBlock	(6,5,512)
ConvLayer	(6,5,512)
Flatten	(15360)
Dense	128
Dense	128
Dense	64
Output	2

[2]https://github.com/sborquez/gerumo

Table 2: ConvBlock Layer.

ConvBlock	Kernel	Filters
ConvLayer	128	5x5
	256	3x3
	512	3x3
Dropout	0.25	
ConvLayer	256	5x5
	512	3x3
	1024	3x3
BatchNormalization		
MaxPool		2x2

2.3 Pseudo-Labeling Strategy

Pseudo-labeling is a semi-supervised strategy that allows supervised training with labelled and unlabelled data. It produces a hybrid loss between the supervised and self-supervised labels, balanced by a mixing coefficient (Lee, 2013). Throughout the training, the balancing coefficient (weighting) changes in favour of the unbalanced data, which usually is available in more significant quantities.

The self-supervised loss uses predicted classes for unlabeled data as if they were true labels. This is done by assigning labels with the predictions from the model, by equation 1:

$$y'_i = \begin{cases} 1 & \text{if i} = argmax f_i(x), \\ 0 & \text{otherwise} \end{cases} \tag{1}$$
$$i \in gamma, other$$

where y'_i is the pseudo label generated for the unlabeled data x, and $f_i(x)$ is the model's output for the $i'th$ label. This means that given unlabeled data x, the algorithm will use the prediction of labels as if they were the true labels.

Initially, the model does not predict the classes accurately, which is why the balancing coefficient starts in favour of supervised learning. Once the model's representation of the classes becomes more precise in the network's latent space, the predictions become more accurate. Therefore the balancing coefficient starts moving toward the self-supervised data until ending with nothing but an unlabeled loss.

Pseudo labelling makes two assumptions on the latent space (Chapelle et al., 2009): *Continuity Assumption* which states that points that are close to each other are more likely to share a label, and *Cluster Assumption* which asserts that data tends to form discrete clusters, and points that belong to the same cluster, are likely to share a label.

Pseudo-labeling is a sound strategy for tackling the CTA gamma/hadron classification problem, but it requires a novel approach in our context. Since labelled data is only available through simulations, once the observatory begins operations, the mayority of available data is going to be real unlabeled data. With this in mind, we propose the following idea as a proof of concept to be used by CTA.

First, we train a supervised model with high-intensity simulated data. Then, we initiate the self-supervised training with an uncertainty threshold as shown in Equation 2, where α is the uncertainty threshold. Ideally, we would use real images here, but since no real data is available, we use lower-intensity images as proof of concept. The uncertainty threshold is to avoid very complex examples that, even though they contain very relevant information for the network to learn, could mislead the training path of the network. But once the network gets modified through learning, those examples could meet the threshold cut in another round.

$$self\_supervised\_data = \begin{cases} x \in dataset & \text{if } f_i(x) > \alpha, \\ x \notin dataset & \text{otherwise} \end{cases} \tag{2}$$

$$i \in gamma, other$$

3 EXPERIMENTS

To test our proof of concept, we first need to validate that we can train and further improve the model with labels created by the same model and tune the α parameter. This is done before taking the second step, by emulating the context of new images. Next, a description of the performed experiments, metrics and datasets is given.

3.1 Validation Experiment

The first experiment assesses our variation of the pseudo-labelling technique, which will be referred to as the *Validation Experiment*. It begins by training a supervised model with a small dataset. The performance reached by this model should be enough to generate distinguishable clusters in the latent space. Then, the same model is trained in a self-supervised fashion by generating the labels with its prediction and applying the uncertainty threshold on a single epoch. The details of the training events are shown in Table 3.

For the Validation Experiment, both supervised and self-supervised datasets have a quality cut on Hillas intensity over 1000 phe. The training datasets are balanced and evenly distributed for both particles. This is done to train the CNN only using the morphology of the image. The two models were tested over the same testing dataset. The testing dataset also had a quality cut of 1000 phe on Hillas intensity.

3.2 Application Experiment

The second experiment, namely the *Application Experiment*, consists of training a model in a supervised manner, just like in the first experiment, but with a larger subset to improve training. Then, this model is further trained in a self-supervised fashion with the same strategy as in the previous experiment but with different quality events. In this experiment, the self-supervised step is done with 300 phe cuts on Hillas intensity instead of 1000 phe. Also, the second used dataset overlaps with the first to soften the training gradients. Since the model was already trained with events from the 1000 phe quality cut, the performance improvement comes from the second quality cut (300 phe). The details of the training events are shown in Table 3.

Both models were tested over the same testing dataset, which also has the 300 phe quality cut on Hillas intensity. All training and validation datasets were balanced in particle class, i.e., proton and gamma, but the testing dataset is unbalanced in a 2:1 proportion, respectively. This is to simulate a more realistic proportion of the classes. A general overview of the experiments and datasets is shown in Table 3.

4 RESULTS

In this section, we present the main results of both experiments. The metrics considered for the analysis are accuracy, recall, precision, and f1-score (Zeugmann et al., 2011).

4.1 Validation Experiment

In the validation experiment, the model was trained using an early stopping mechanism until a plateau was reached in the loss. Since we used CLR, the network converges faster than conventional methods.

For the Supervised step, we trained for 23 epochs, but the minimum loss was reached on epoch 17. The joint results for testing are shown in table 4.

While for the self-supervised method, all the training was done on a single epoch to avoid overfitting. Since applying an uncertainty threshold to the network could easily overfit the examples that the model already classifies with low uncertainty.

Table 3: Summary of data used for the experiments.

Experiments	Training (strategy: #events)	Validation	Testing
Validation	supervised: 74945[a] self-supervised: 541911 [a]	7898[a]	234276[a]
Application	supervised: 541911 [a] self-supervised: 1519741 [b]	108945 [a]	605424[b]

[a] greater than 1000 phe, [b] greater than 300 phe

The α parameter was found by grid search as stated in Equation 2. The model's learning capacity seems sensitive to this parameter, but our best performance was reached with $\alpha = 0.9$, thus applied for both experiments.

As shown in table 4, the self-supervised strategy improved the model's performance in all metrics. In other words, for the same image quality (same Hillas intensity cut), the model learned a better representation using the predicted labels as if they were true. Consequently, we can conclude that the general semi-supervised strategy is an excellent candidate to further study with different image qualities.

4.2 Application Experiment

For the Application experiment, it can be seen that performance also improved for all metrics except for precision, which decreased by 1%. However, the increase in recall and accuracy tell us that the improvement in correctly identifying gamma comes at the price of a slight rise in the type-I error.

Table 4: Summary of the results of both experiments.

Experiments	Accuracy	Recall	Precision	f1-score
Validation Experiment				
Supervised	0.88	0.85	0.82	0.82
self-supervised	0.90	0.92	0.85	0.85
Control (1000 phe)	0.91	0.90	0.83	0.87
Application Experiment				
Supervised	0.81	0.76	0.70	0.73
self-supervised	0.83	0.87	0.69	0.77
Control (300 phe)	0.85	0.89	0.73	0.80

For all experiments, Recall and Precision are focused on gammas. Also, a control experiment was done by training the model in a supervised fashion on all the training datasets (complete information scenario). The idea is to have a broader vision of where the semi-supervised strategy lies.

4.3 Considerations

Our experimental design used different-quality unlabeled data, which might be very different from real event data. Real data can diverge significantly from simulated data (Shilon et al., 2019), considering that our cuts are different instances of the same larger dataset and not an independent data source, further experiments must be done on the difference of real/simulated data (Jacquemont et al., 2021). However, we assume that a trained classifier with carefully-crafted simulated data will have a reasonable performance on real data (Miener et al., 2021a; Vuillaume et al., 2021). This assumption must be verified with real data; therefore, we suggest a latent space tool to further help with this purpose.

The difference between real/simulated data is a matter of constant study. Some of the studied differences consider dead/bright pixels and camera modules, imperfect calibration, imperfect NSB/moonlight modelling, atmospheric modelling, ageing of the telescopes, and the difficulty in reliable modelling of hadronic interactions. Our strategy studies a possible approach to take advantage of using such complex real images, but further research should be done specifically on real/simulated data integration.

5 LATENT SPACE

We assume that a model that reaches a good performance also learns a good representation of the phenomenon, meaning that the images generate a nonlinear transformation which leads to separable classes to predict in a latent space. Since CNN reduces the dimension with depth, it brings an opportunity to explore this latent space representation of the network.

While convolutional layers extract certain features that become more complex with depth, logic layers combine these features to fulfil the task, in this case, class prediction. We could extract the result of any intermediate layers to explore the latent space. Then, to visualize, we can apply dimensional reduction techniques, such as UMAP (McInnes et al., 2020), t-SNE (Laurens Van der Maaten and Hinton, 2008) or PCA. We applied PCA to show how the network distinguishes between classes and confirm our assumptions on pseudo-labelling.

As shown in Figure 1, classes are separable and misclassified cases are accumulated on the edges of the clusters, as was expected. We also confirm that the *Cluster assumption* and *Continuity assumption* for the pseudo labelling hold.

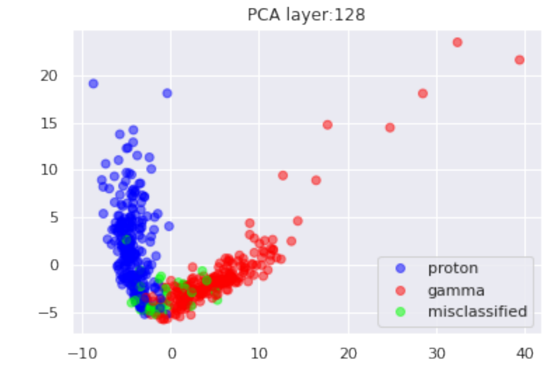

Figure 1: PCA applied to the first logic Layer (1st Dense 128).

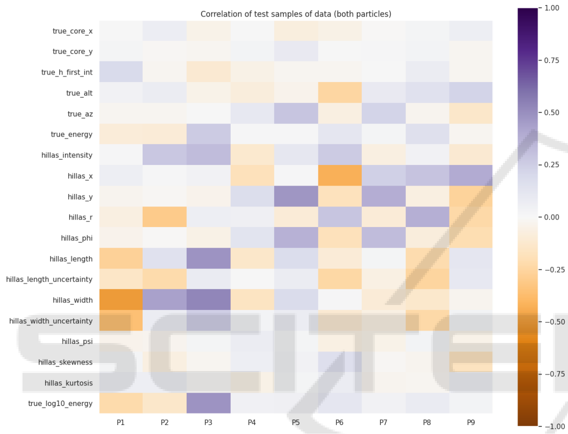

Figure 2: Correlation between PCA and Hillas Parameters.

Additionally for PCA, we set a threshold of 70% of the explained variance to get the number of components for representation on the first logic layer to find common ground between the CNN representation and Hillas parameters. In Figure 2, we show the correlation between the Hillas parameters and the PCA components for the testing set from the Validation Experiment.

As shown in Figure 2, the first PCA component P1 is inversely correlated with Hillas length and width. This means that the features learned by the CNN correlate with Hillas parameters that are known to be essential to discriminate whether it is gamma or proton (Albert, 2007; Krause et al., 2017).

Further investigations need to be done regarding these latent spaces. It also could be used to diagnose and study the structural differences between real and simulated data (Shilon et al., 2019).

6 CONCLUSION

We proposed a proof of concept for self-supervised training for CTA. We validated our modification of the pseudo-labelling strategy and then retrained the model with images with different quality cuts. The model learned from fainter pseudo-labelled images and improved the identification of gamma-ray showers in a statistically relevant amount.

These results suggest that our proposal is a suitable candidate to stimulate this line of research and a potential approach for the actual operation of IACTs. The potential for classification performance on real data could be enhanced by augmenting simulated training data with real event images.

Our brief exploration of the latent space could be used as common ground to connect Convolutional Neural Networks and the Hillas parametrization. With a deeper exploration, we expect to understand how the networks classify the data, contributing to further improvements in simulation. For example, training a CNN to classify real and simulated images as in (Shilon et al., 2019) and then studying its latent space could reveal a path to improve simulations further.

ACKNOWLEDGEMENTS

This work was performed using resources provided by the Cambridge Service for Data Driven Discovery (CSD3) operated by the University of Cambridge Research Computing Service (www.csd3.cam.ac.uk), provided by Dell EMC and Intel using Tier-2 funding from the Engineering and Physical Sciences Research Council (capital grant EP/T022159/1), and DiRAC funding from the Science and Technology Facilities Council (www.dirac.ac.uk).This work used IRIS computing resources funded by the STFC.

This work was conducted in the context of the CTA Analysis and Simulations Working Group, and this paper has been through internal review by the CTA Consortium. We gratefully acknowledge financial support from the agencies and organizations listed here: http://www.cta-observatory.org/consortium_acknowledgments.

The work was also support by ANID PIA/APOYO AFB180002, ANID-Basal Project FB0008 and project FONDECYT 1190886.

REFERENCES

Albert, J. (2007). Implementation of the Random Forest Method for the Imaging Atmospheric Cherenkov Telescope MAGIC. Publisher: arXiv Version Number: 2.

Becherini, Y., Djannati-Ataï, A., Marandon, V., Punch, M., and Pita, S. (2011). A new analysis strategy for detection of faint gamma-ray sources with Imaging Atmospheric Cherenkov Telescopes. Publisher: arXiv Version Number: 1.

Bernlohr, K. (2008). Simulation of Imaging Atmospheric Cherenkov Telescopes with CORSIKA and sim_telarray. Publisher: arXiv Version Number: 1.

Brill, A., Feng, Q., Humensky, T. B., Kim, B., Nieto, D., and Miener, T. (2019). Investigating a Deep Learning Method to Analyze Images from Multiple Gamma-ray Telescopes. In *2019 New York Scientific Data Summit (NYSDS)*, pages 1–4. arXiv:2001.03602 [astro-ph].

Chapelle, O., Scholkopf, B., and Zien, Eds., A. (2009). Semi-Supervised Learning (Chapelle, O. et al., Eds.; 2006) [Book reviews]. *IEEE Transactions on Neural Networks*, 20(3):542–542.

Developers, T. (2022). TensorFlow.

Drew, R. D. (2021). Deep unsupervised domain adaptation for gamma-hadron separation.

Grespan, P., Jacquemont, M., López-Coto, R., Miener, T., Nieto-Castaño, D., and Vuillaume, T. (2021). Deep-learning-driven event reconstruction applied to simulated data from a single Large-Sized Telescope of CTA. arXiv:2109.14262 [astro-ph].

Heck, D., Knapp, J., Capdevielle, J. N., Schatz, G., and Thouw, T. (1998). *CORSIKA: a Monte Carlo code to simulate extensive air showers.* Publication Title: CORSIKA: a Monte Carlo code to simulate extensive air showers ADS Bibcode: 1998cmcc.book.....H.

Hillas, A. M. (1985). Cerenkov light images of EAS produced by primary gamma. 3(OG-9.5-3):4.

Jacquemont, M., Vuillaume, T., Benoit, A., Maurin, G., Lambert, P., and Lamanna, G. (2021). First Full-Event Reconstruction from Imaging Atmospheric Cherenkov Telescope Real Data with Deep Learning. In *2021 International Conference on Content-Based Multimedia Indexing (CBMI)*, pages 1–6, Lille, France. IEEE.

Jacquemont, M., Vuillaume, T., Benoit, A., Maurin, G., Lambert, P., Lamanna, G., and Brill, A. (2019). GammaLearn: A Deep Learning Framework for IACT Data. In *Proceedings of 36th International Cosmic Ray Conference — PoS(ICRC2019)*, page 705, Madison, WI, U.S.A. Sissa Medialab.

Juryšek, J., Lyard, E., and Walter, R. (2021). Full LST-1 data reconstruction with the use of convolutional neural networks. arXiv:2111.14478 [astro-ph].

Krause, M., Pueschel, E., and Maier, G. (2017). Improved γ/hadron separation for the detection of faint gamma-ray sources using boosted decision trees. *Astroparticle Physics*, 89:1–9. arXiv:1701.06928 [astro-ph].

Laurens Van der Maaten and Hinton, G. (2008). Visualizing data using t-SNE. 9(11).

Lee, D.-H. (2013). Pseudo-Label : The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks. Technical report.

Lu, C.-C. (2013). Improving the H.E.S.S. angular resolution using the Disp method. arXiv:1310.1200 [astro-ph].

Mangano, S., Delgado, C., Bernardos, M., Lallena, M., and Vázquez, J. J. R. (2018). *Extracting gamma-ray information from images with convolutional neural network methods on simulated Cherenkov Telescope Array data*, volume 11081. arXiv:1810.00592 [astro-ph].

McInnes, L., Healy, J., and Melville, J. (2020). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv:1802.03426 [cs, stat].

Miener, T., López-Coto, R., Contreras, J. L., Green, J. G., Green, D., Mariotti, E., Nieto, D., Romanato, L., and Yadav, S. (2021a). IACT event analysis with the MAGIC telescopes using deep convolutional neural networks with CTLearn. arXiv:2112.01828 [astro-ph].

Miener, T., Nieto, D., Brill, A., Spencer, S., and Contreras, J. L. (2021b). Reconstruction of stereoscopic CTA events using deep learning with CTLearn. arXiv:2109.05809 [astro-ph].

Nieto, D., Brill, A., Feng, Q., Humensky, T. B., Kim, B., Miener, T., Mukherjee, R., and Sevilla, J. (2019). CTLearn: Deep Learning for Gamma-ray Astronomy. arXiv:1912.09877 [astro-ph].

Nieto, D., Brill, A., Kim, B., and Humensky, T. B. (2017). Exploring deep learning as an event classification method for the Cherenkov Telescope Array. arXiv:1709.05889 [astro-ph].

Ohm, S., van Eldik, C., and Egberts, K. (2009). Gamma-Hadron Separation in Very-High-Energy gamma-ray astronomy using a multivariate analysis method. Publisher: arXiv Version Number: 1.

Shilon, I., Kraus, M., Büchele, M., Egberts, K., Fischer, T., Holch, T., Lohse, T., Schwanke, U., Steppa, C., and Funk, S. (2019). Application of deep learning methods to analysis of imaging atmospheric Cherenkov telescopes data. *Astroparticle Physics*, 105:44–53.

Smith, L. N. (2017). Cyclical Learning Rates for Training Neural Networks. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 464–472, Santa Rosa, CA, USA. IEEE.

The CTA Consortium (2019). *Science with the Cherenkov Telescope Array*. WORLD SCIENTIFIC.

Vuillaume, T., Jacquemont, M., de Lavergne, M. d. B., Sanchez, D. A., Poireau, V., Maurin, G., Benoit, A., Lambert, P., Lamanna, G., and Project, C.-L. (2021). Analysis of the Cherenkov Telescope Array first Large-Sized Telescope real data using convolutional neural networks. arXiv:2108.04130 [astro-ph].

Zeugmann, T., Poupart, P., Kennedy, J., Jin, X., Han, J., Saitta, L., Sebag, M., Peters, J., Bagnell, J. A., Daelemans, W., Webb, G. I., Ting, K. M., Ting, K. M., Webb, G. I., Shirabad, J. S., Fürnkranz, J., Hüllermeier, E., Matwin, S., Sakakibara, Y., Flener, P., Schmid, U., Procopiuc, C. M., Lachiche, N., and Fürnkranz, J. (2011). Precision and Recall. In Sammut, C. and Webb, G. I., editors, *Encyclopedia of Machine Learning*, pages 781–781. Springer US, Boston, MA.