

# Particle identification using semi-supervised learning in the PICO-60 dark matter detector

Brendon Matusch<sup>1</sup> and Gevy Cao<sup>2</sup>

<sup>1</sup>SNOLAB, Lively, Ontario, P3Y 1N2, Canada

<sup>2</sup>Department of Physics, Queen's University, Kingston, Ontario, K7L 3N6, Canada

E-mail: [brendon.matusch@snolab.ca](mailto:brendon.matusch@snolab.ca), [gevy.cao@queensu.ca](mailto:gevy.cao@queensu.ca)

**Abstract.** Astrophysical observations have demonstrated the existence of dark matter over the past decades. Experimental efforts in the search for dark matter are largely focused on the well-motivated weakly interacting massive particles (WIMPs) as a dark matter candidate. Current experiments in direct detection are producing increasingly competitive limits on the cross section of WIMP-nucleon scattering. The main experimental challenge for all direct detection experiments is the presence of background signals. These backgrounds need to be either eliminated by providing sufficient shielding or discriminated from WIMP signals. In this work, semi-supervised learning techniques are developed to discriminate alpha recoils from nuclear recoils induced by WIMPs in the PICO-60 detector. The two semi-supervised learning techniques, gravitational differentiation and iterative cluster nucleation, maximize the effect of the most confidently predicted data samples on subsequent training iterations. Classifications using both techniques can reproduce the traditional acoustic parameter with accuracies over 98%. The best model yields an accuracy of 99.2% and a class-wise standard deviation value of 0.11. These techniques can reliably serve as an intermediate verification tool before the acoustic parameter is constructed in future detectors.

## 1. Introduction

The presence of dark matter has become evident over the past decades. Its existence is supported by numerous astrophysical observations, including the deviation of the galactic rotation curves from Newton's law of universal gravitation[1], the mass distribution of the bullet cluster due to gravitational lensing effect[2], and it is most strongly inferred by the anisotropy due to baryon acoustic oscillations in the cosmic microwave background[3]. Experiments are being conducted to search for dark matter directly through scattering with a nucleus, or indirectly through annihilation and subsequent decay into standard model particles[4]. Collider production of dark matter offers another gateway to probe its fundamental interactions. There are many dark matter candidates being considered, including axions[5] and primordial black holes[6], with the current detectors most well equipped to search for weakly interacting massive particles (WIMPs)[7][8].

Direct detection experiments have shown promising results in the field as they improve in sensitivity and scale[9]. To directly search for the hypothetical particle, many detectors are developed to be sensitive to nuclear recoils, where a dark matter particle elastically scatters off of a target nucleus via momentum transfer directly to the nucleus, causing the nucleus to

recoil. The energy deposited by the recoiling nucleus can then be detected through heat, charged particles, light or a combination of the three[10][11]. However, other particles such as neutrons, gamma rays and neutrinos can produce signals in the detectors that mimic those produced by a dark matter candidate. The experimental challenge is to either eliminate these backgrounds, or differentiate them from dark matter signals. In the last decade, experiments have been seeking deeper underground facilities to shield their detectors from cosmic rays. The remaining background signals can be rejected using calorimetry or machine learning techniques.

## 2. The PICO experiment

PICO, the joint effort between the original PICASSO and COUPP experiments[12][13], use superheated  $C_3F_8$  as the target fluid for WIMP-nucleon scattering. The detector is brought to a metastable thermodynamic (superheated) state at a set temperature by lowering the pressure past the saturation curve. The degree of superheat determined by the pressure and temperature choices sets the Seitz thermodynamic threshold in the detector. When the liquid is in the superheated state, any energy deposition above this threshold for nucleation will induce vaporization of the liquid at the interaction vertex[14].

PICO is located 2073 meters underground at the Sudbury Neutrino Observatory (SNOLAB) in Canada. Therefore, cosmic muons lose significant amount of energy for muon-induced neutrons to be a background in the detector. Additional neutron shielding is provided by immersing the detector in a water tank approximately 1.5 meters in radius. PICO detectors are extremely insensitive to electron recoils induced by gamma rays. The expected gamma background per live year is typically less than 1 event in the current-generation detector with appropriate threshold choices. The remaining background of concern is alpha recoils. Alpha particles are decay products of uranium, thorium and radon, which are naturally present in the lab and in the materials surrounding the detector. The decay of uranium or thorium in the materials around the detector can result in radon diffusion into the active liquid; radon can also cling onto the walls of the quartz jars holding the target fluid as a result of the exposure of the jars to lab air prior to assembly. Alpha recoils, like nuclear recoils, can induce nucleation in the target fluid. However, their acoustic properties are distinctly different. PICO detectors use piezos (Figure 1) to record the acoustic information of each event, and reject alphas with the integrated acoustic power (AP) in selected frequency bands. [15]

The technique provides excellent rejection for alpha recoils, shown in Figure 2. However, the disadvantage is that it requires manual tuning (e.g. frequency band selection, position corrections etc) for each new detector after large sets of calibration data. In some low threshold runs during the operation of PICO-60, ambiguous mid-AP events begin to emerge, which can obscure the detector's ability to discriminate nuclear recoils from alpha recoils. Early machine learning techniques utilizing supervised learning were developed for PICO-60 to complement the discrimination power of AP. In this work, novel techniques exploiting semi-supervised learning using PICO-60 data proved to reproduce AP with over 99% accuracy.

## 3. Semi-supervised learning for particle identification

Semi-supervised learning is an uncommon machine learning technique, relative to widely-used supervised and unsupervised learning. The concept is to train a machine learning model on a set of labeled data in addition to a set of unlabeled data. In the context of PICO-60, the labeled sets consist of the neutron calibration runs and the background radiation runs (which consist predominantly of alpha recoils). These labeled sets have a strong but not perfect correlation with particle types. The unlabeled data can consist of any mixture of particle types.

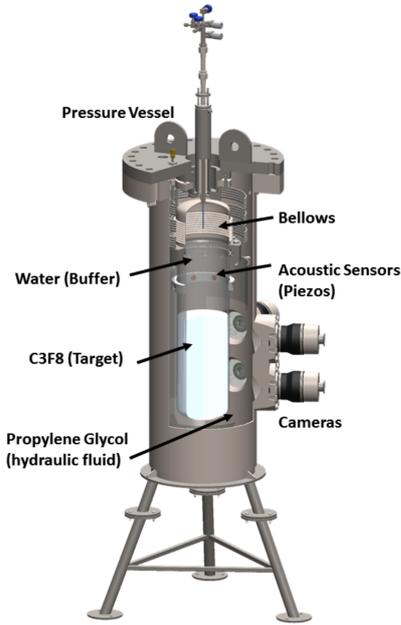


Figure 1: CAD model of the PICO-60 detector.

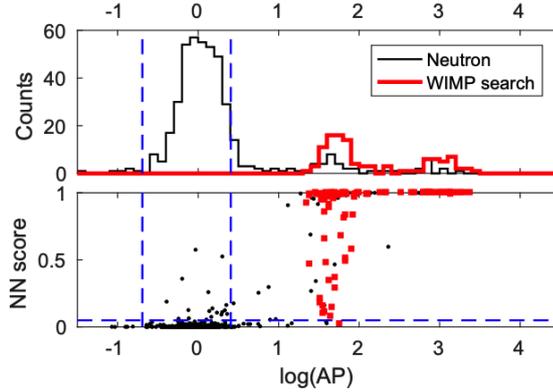


Figure 2: Top: Acoustic parameter of PICO-60 at a threshold of 3.3 keV. The peak of the AP distribution of nuclear recoil events are normalized to 1 from neutron calibration. The WIMP search data contain only alpha recoils. Bottom: signal classifications by a neural network using supervised learning (early development, not a part of this work). Nuclear recoil events are assigned a score of 0 and alpha recoil events are assigned a score of 1.

The primary advantage of semi-supervised learning is clear: it requires fewer labels to be collected than conventional supervised learning. PICO-60 data sets are relatively small after analysis cuts (in the order of hundreds of events per category), which makes semi-supervised learning particularly advantageous. A less obvious but powerful secondary advantage is that, since the learning process allows the network to reinforce its own decisions, the negative effect of impure training data is minimized, potentially allowing the resulting network to perform better than a network trained with supervised learning.

Two techniques for semi-supervised learning have been developed for this study: gravitational differentiation and iterative cluster nucleation. The essence of both algorithms is a positive feedback loop. First, the network is trained on a small amount of imperfectly labeled data (data drawn from the neutron calibration or WIMP search runs). Throughout the training process, the network runs predictions on unlabeled data. By using its most confident predictions in the training process, the desired function separating particle types is continuously reinforced, while less confident predictions are de-emphasized. This effectively draws new, unlabeled data into the training set for further iteration.

In this experimentation, multi-layer perceptrons were used as the basic neural network model. They were trained on a banded, integrated Fourier transform of the acoustic data from the two working piezos of the detector (Figures 3 and 4). All architectures were optimized using grid searches, in which every possible combination of hyperparameters within a fixed parameter space is tested [16]. The best architecture, consists of an input layer of 16 neurons, two fully-connected layers of decreasing number of neurons (12 and 8) and a binary output layer. The dropout probabilities between the last three layers are both 0.25. This architecture was selected using accuracy calculated on a randomly selected validation set.

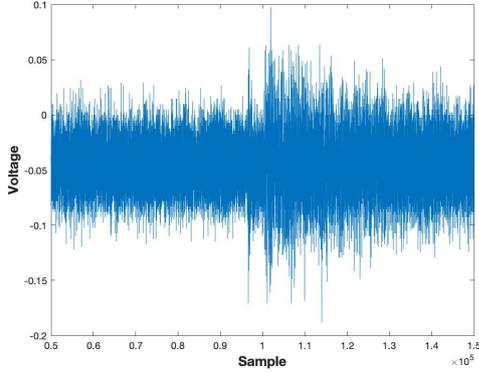


Figure 3: A typical acoustic waveform during bubble formation.

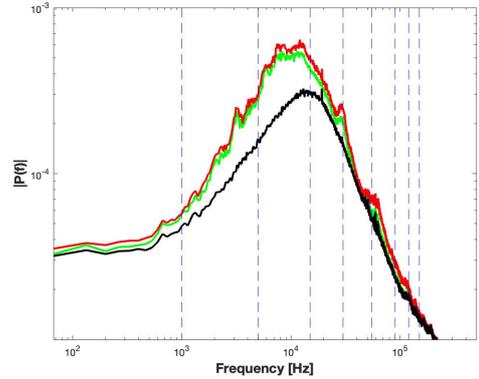


Figure 4: The average fourier power of alphas (red), nuclear recoil proxies (green) and noise (black). The frequency bands are defined by the dashed blue lines.

### 3.1. Gravitational Differentiation

Gravitational differentiation (GD) is a novel technique for training a neural network in a semi-supervised fashion on a relatively small set of imperfectly labeled data, while simultaneously incorporating the network’s changing predictions on a set of unlabeled data to encourage decisive classifications. In general terms, it amplifies the training effect associated with high-confidence predictions (those that most clearly show characteristics of either an alpha particle or a neutron) on unlabeled examples. Meanwhile, it minimizes the training impact of low-confidence predictions, such that the network will not be trained to make incorrect predictions.

Analogous to a gravitational effect, events with high-confidence predictions are attracted closer to their corresponding classifications. They are then used as training data, allowing the patterns they represent to be used more generally to make predictions on other events. The system accomplishes this using a new method for calculating gradients for the last layer of the neural network. Gradients are calculated using the gravitational differentiation function  $\text{GravDiff}(p, \psi, g)$ , which is parameterized by the network’s existing prediction  $p$  on the training example in question, the degree  $\psi$  of the piecewise exponential function used to distort the response of the gradient, and the gravitational multiplier  $g$ . The function is defined as follows:

$$\text{GravDiff}(p, \psi, g) = g \cdot \text{sgn}(p) \cdot \text{abs}(\tanh(2(p - 0.5)))^\psi \quad (1)$$

It is a transformation of the hyperbolic tangent that flattens the central range and comparatively exaggerates the asymptote on either side. The equation above can be described in the following steps and is further demonstrated in Figure 5:

- (i) Transform the network’s prediction  $p$ , so its range is 1 to -1 rather than 0 to 1.
- (ii) Apply the sigmoidal hyperbolic tangent, producing a value of 0 in the center and asymptotic slopes to -1 and 1 at the edges.
- (iii) Apply a relatively large exponent which is  $\psi$  (in the range of 3 to 11) to compress values close to 0 closer to 0. This ensures a shallow slope in the center (so low-confidence network outputs near 0.5 will produce very small output gradients).

This function is applied in a training algorithm that is parameterized by the set of imperfectly labeled data  $\varsigma$ , the set of unlabeled training data  $\nu$ , a binary classification neural network  $NN(x)$

(where  $x$  is the input data format), and the gravitational multiplier increment  $\delta_g$ . The distortion power  $\psi$  is assumed to be constant throughout a training run, and the gravitational multiplier  $g$  is initialized at 0, gradually increasing during training. It iterates as follows:

- (i) Train  $NN(x)$  for a single epoch on the combined set  $\zeta \cup v$ , using the predefined ground truths for  $\zeta$  (with a mean squared error loss function) and the most recently calculated gravitational gradients for each of the examples in  $v$ .
- (ii) Calculate predictions  $p$  with  $NN(x)$  on the entirety of  $v$ . Calculate  $\text{GravDiff}(p, \psi, g)$  and record the resulting gradients for the next iteration.
- (iii) Increment  $g$  by  $\delta_g$ . At the beginning, when  $g = 0$ , training will be entirely based on the labeled set  $\zeta$ , and progressively, over the course of a run, the gravitational effect increases with  $g$ .

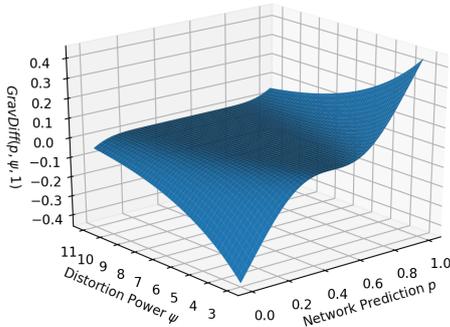


Figure 5: Demonstration of the principle of gravitational differentiation, where the  $\text{GravDiff}$  function distorts the gradient by  $\psi$  and pulls confident predictions closer to the binary categories (0 or 1), creating the positive feedback loop in the semi-supervised learning. The less certain predictions will have a gradient close to 0 in the next training iteration.

### 3.2. Iterative Cluster Nucleation

Iterative cluster nucleation (ICN) is a second semi-supervised learning algorithm developed as part of this study. It takes advantage of some amount of imperfectly labeled data, using it to classify the rest of an unlabeled training set while optimizing to produce an effective discriminator. The basic concept is similar to that of gravitational differentiation. Starting with a relatively small amount of labeled data from two classes (nuclear recoils and alpha particles, in this case), predictions are run on a set of unlabeled data, and those predictions are used to produce further training data. However, there are three key differences:

- (i) In iterative cluster nucleation, an unlabeled example is not included in the training process at all, until a very confident prediction is made on it. Once that occurs, it is added to the training set, expanding one of the two “clusters” of training data.
- (ii) Once an example is added to the active training set, it is never removed. This ensures that the highly confident predictions made at the very beginning continue to influence the training process.
- (iii) The unlabeled examples, once added to the active training set, are weighted as strongly as examples from the original labeled data set. This ensures imperfect labels in the labeled data do not outweigh the confident and more likely correct predictions from later in the training process.

### 3.3. Performance Comparison

The criteria that define the success of a network are: classification accuracy and confidence. The accuracy of the network is assessed against AP, assuming AP has perfect discrimination power. The confidence of the network on the predictions is characterized by the class-wise standard deviation (CWSD):

$$SD = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}} \quad (2)$$

$$CWSD = \frac{1}{2} \left[ \sqrt{\frac{\sum_{i=1}^{N_n} (x_{n,i} - \mu_n)^2}{N_n \cdot SD}} + \sqrt{\frac{\sum_{i=1}^{N_a} (x_{a,i} - \mu_a)^2}{N_a \cdot SD}} \right] \quad (3)$$

where  $N, N_n, N_a$  are the number of data samples in total, from neutron calibration set and background set respectively.  $x_n$  are predictions in the neutron calibration set and  $x_a$  are predictions in the background set. The  $\mu$ 's are the mean value of their respective dataset. A more confident neural net will predict a CWSD closer to 0 on the scale of 0 to 1.

Table 1 directly compares performance of all semi-supervised learning configurations, with supervised learning, the original PICO-60 ML analysis, and AP provided as baselines for comparison. For every  $\varsigma$  size except 32, ICN performs better than supervised learning; for 128 and 256, GD also performs better. For 32, 64, and 128 labeled examples, ICN proves to be more effective than GD; however, when 256 examples are labeled, GD provides the best accuracy, at 99.2%. Most notably, iterative cluster nucleation with only 128 labeled examples achieves accuracy 1.2 percentage points higher than supervised learning with all 496 training examples labeled. It also makes more confident predictions (with a class-wise standard deviation of only 0.21, as opposed to 0.26).

Table 1: Summary of the most successful semi-supervised learning models. Supervised learning models are included for comparison.

Initial $\varsigma$ Size	Technique	Accuracy	Precision	Recall	CWSD
32	GD	71.3%	70.1%	98.0%	0.73
32	ICN	85.2%	83.4%	98.4%	0.54
32	Supervised	90.4%	89.0%	97.1%	0.53
64	GD	79.7%	75.5%	98.6%	0.63
64	ICN	94.3%	92.2%	99.8%	0.34
64	Supervised	92.7%	90.4%	98.9%	0.41
128	GD	97.0%	96.3%	99.6%	0.23
128	ICN	98.2%	97.7%	99.4%	0.21
128	Supervised	94.7%	92.6%	99.5%	0.37
256	GD	99.2%	99.0%	99.7%	0.11
256	ICN	98.2%	97.6%	99.5%	0.22
256	Supervised	96.5%	94.6%	99.7%	0.28
496	Supervised	97.0%	95.4%	99.9%	0.26
278	PICO-60 ML	80.7%	76.9%	100.0%	0.65
N/A	AP	100%	100%	100%	0.43

## 4. Discussion and conclusion

The neural networks developed in this study using semi-supervised learning provide unprecedented, automated discrimination power comparable to the acoustic parameter. The two

semi-supervised learning techniques, gravitational differentiation and iterative cluster nucleation, are developed based on the principle that the more confident predictions on unlabeled data have a larger effect on training in the next iteration. While gravitational differentiation produced classifications with the highest accuracy and confidence, iterative cluster nucleation has proven to be more data-efficient. In particular, these two semi-supervised learning techniques consistently perform better in both accuracy and confidence comparing to supervised learning with increasing initial sample sizes. The best network model reproduced the AP predictions with an accuracy of 99.2% and a confidence value (CWSD) of 0.11. These techniques can complement the results predicted by AP, and provide intermediate verifications on small calibration data sets before AP is calculated in future detectors. The results may also be used in assisting the placement of the AP cuts. The development of effective machine learning techniques can be a powerful tool in background discrimination for dark matter detectors, as experiments prepare to deploy ton-scale detectors with potential for WIMP detection.

## References

- [1] Corbelli E and Salucci P 2000 *Mon. Not. Roy. Astron. Soc.* **311** 441–447 (*Preprint astro-ph/9909252*)
- [2] Clowe D, Gonzalez A and Markevitch M 2004 *Astrophys. J.* **604** 596–603 (*Preprint astro-ph/0312273*)
- [3] Eisenstein D J *et al.* (SDSS) 2005 *Astrophys. J.* **633** 560–574 (*Preprint astro-ph/0501171*)
- [4] Klasen M, Pohl M and Sigl G 2015 *Prog. Part. Nucl. Phys.* **85** 1–32 (*Preprint 1507.03800*)
- [5] Peccei R D 2008 *Lect. Notes Phys.* **741** 3–17 [3(2006)] (*Preprint hep-ph/0607268*)
- [6] Carr B, Kühnel F and Sandstad M 2016 *Phys. Rev. D* **94**(8) 083504 URL <https://link.aps.org/doi/10.1103/PhysRevD.94.083504>
- [7] Griest K 1993 The Search for the Dark Matter: WIMPs and MACHOs *Texas/PASCOS '92: Relativistic Astrophysics and Particle Cosmology* vol 688 ed Akerlof C W and Srednicki M A p 390 (*Preprint hep-ph/9303253*)
- [8] Schumann M 2019 *Journal of Physics G: Nuclear and Particle Physics* **46** 103003 direct Detection of WIMP Dark Matter: Concepts and Status (*Preprint 1903.03026*)
- [9] Liu J, Chen X and Ji X 2017 *Nature Phys.* **13** 212–216 (*Preprint 1709.00688*)
- [10] Meunier P *et al.* 1999 *Appl. Phys. Lett.* **75** 1335 (*Preprint physics/9906017*)
- [11] Renner J *et al.* (NEXT) 2015 *Nucl. Instrum. Meth.* **A793** 62–74 (*Preprint 1409.2853*)
- [12] Bolte W J *et al.* 2006 *J. Phys. Conf. Ser.* **39** 126–128
- [13] Amole C *et al.* 2015 PICASSO, COUPP and PICO - search for dark matter with bubble chambers *European Physical Journal Web of Conferences* vol 95 p 04020
- [14] Behnke E *et al.* (COUPP) 2013 *Phys. Rev.* **D88** 021101 (*Preprint 1304.6001*)
- [15] Amole C *et al.* (PICO Collaboration) 2017 *Phys. Rev. Lett.* **118**(25) 251301 URL <https://link.aps.org/doi/10.1103/PhysRevLett.118.251301>
- [16] Matusch B, Amole C, Ardid M, Arnquist I J, Asner D M, Baxter D, Behnke E, Bressler M, Broerman B, Cao G, Chen C J, Chowdhury U, Clark K, Collar J I, Cooper P S, Coutu C B, Cowles C, Crisler M, Crowder G, Cruz-Venegas N A, Dahl C E, Das M, Fallows S, Farine J, Felis I, Filgas R, Girard F, Giroux G, Hall J, Hardy C, Harris O, Hillier T, Hoppe E W, Jackson C M, Jin M, Klopfenstein L, Krauss C B, Laurin M, Lawson I, Leblanc A, Levine I, Licciardi C, Lippincott W H, Loer B, Mamedov F, Mitra P, Moore C, Nania T, Neilson R, Noble A J, Oedekerck P, Ortega A, Piro M C, Plante A, Podviyanuk R, Priya S, Robinson A E, Sahoo S, Scallon O, Seth S, Sonnenschein A, Starinski N, Štekl I, Sullivan T, Tardif F, Vázquez-Jáuregui E, Walkowski N, Weima E, Wichoski U, Wierman K, Yan Y, Zacek V and Zhang J 2018 *arXiv e-prints* arXiv:1811.11308 (*Preprint 1811.11308*)