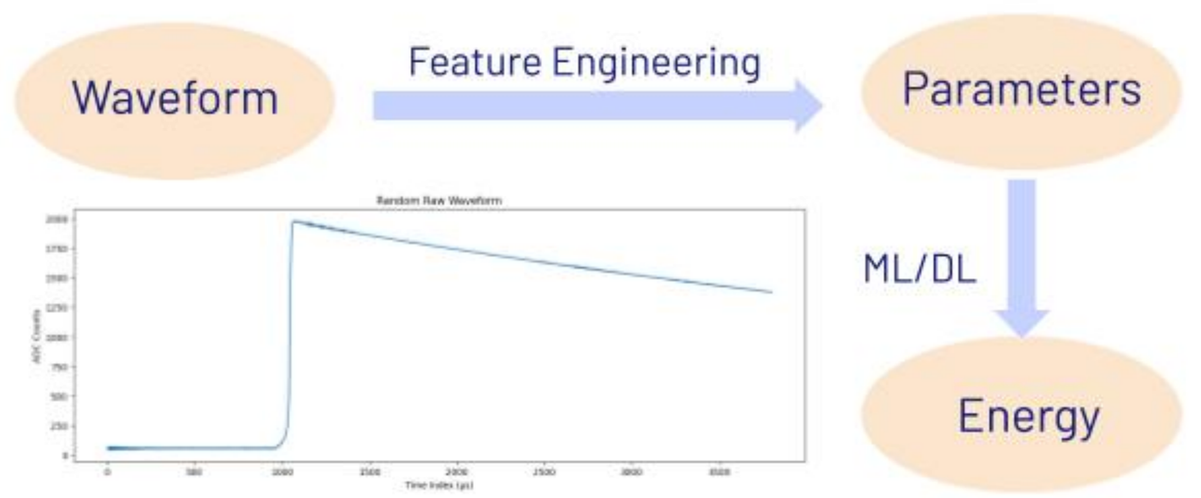
Finding Ghost Particles

Regression Subgroup

Ammie Xie axie@ucsd.edu • Haotian Zhu haz084@ucsd.edu • Matthew Segovia masegovia@ucsd.edu
Mentor: Aobo Li aol002@ucsd.edu

Introduction

- The standard model of particle physics suggest that each fermion particle has a corresponding antiparticle of opposite charge.
- The existence of neutrino particles defy this model, researching them may provide better insight into how to universe works.
- Majorana Demonstrator experiment searched for neutrinoless double beta decay ($0\nu\beta\beta$) through waveforms using high-purity Ge(HPGe) detectors.



Data

The processed train, test and NPML datasets contain 1.69M, 300k and 150k observations respectively. Each raw waveform data is feature engineered to extract parameters based on behaviors observed in the waveform.

Field	Description	Data Type
raw_waveform	Detector Waveform	array(size=(3800,) dtype=float)
energy_label	Analysis Label	float

Parameters

- Drift Time (tdrift)
- Late Charge (lq80)
- Late Charge Slope (AGR)
- Second derivative Inflection Points
- Rising Edge Slope
- Rising Edge Asymmetry (REA)
- Current Amplitude
- Energy Peak
- Tail Slope
- Delayed Charge Recovery (DCR)
- Fourier Transform
- Low Frequency Power Ratio (lfpr)

All of these parameters can be derived from a single waveform using principles of calculus and statistical analysis.

Model Building

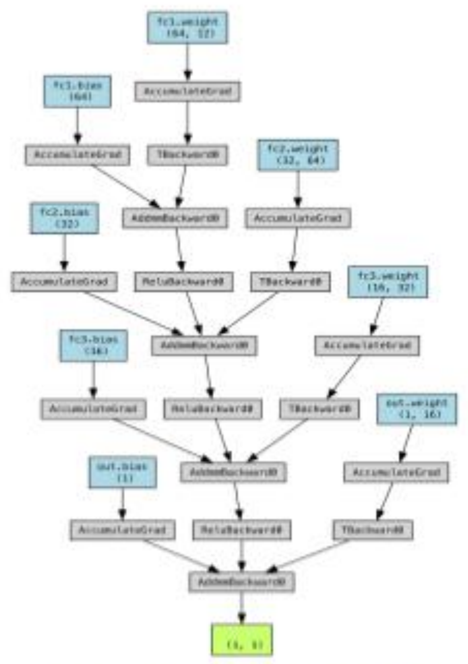
- **Tier 1**:
  - **Baseline** (Linear Regression)
  - **Ridge Regression**
  - **SVM**

  These simpler methods are not capable of capturing the complexities within the waveforms. Adding regularization via Ridge did not significantly enhance performance compared to baseline. Meanwhile, SVM also overfit or failed to generalize in this context.

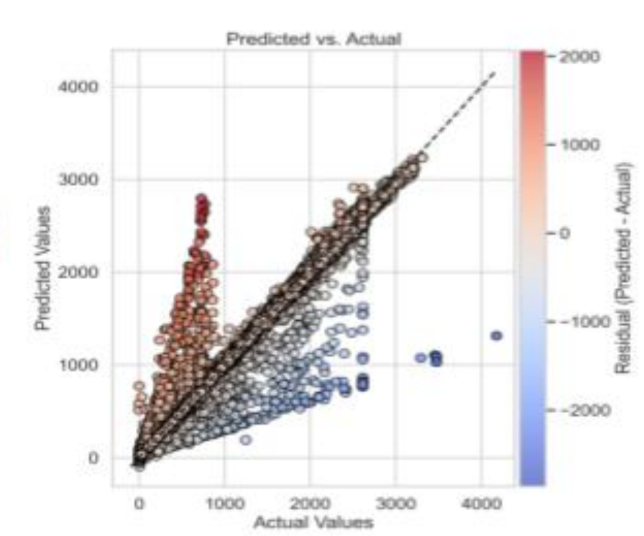
- **Tier 2**:
  - **Random Forest Regressor**

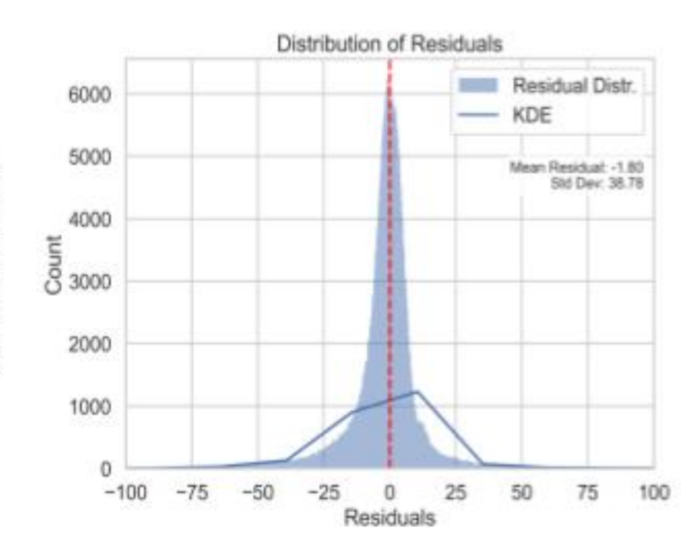
  Evaluated from different number of trees, yields the lowest MSE on the test set with 128 trees.

- **Best overall**:
  - **Neural Network**: Developed in PyTorch. This NN model features three hidden layers (64 → 32 → 16) with ReLU activations. Dropout was intentionally omitted to preserve subtle waveform characteristics, given the sensitive nature of particle interactions.



Architecture of NN



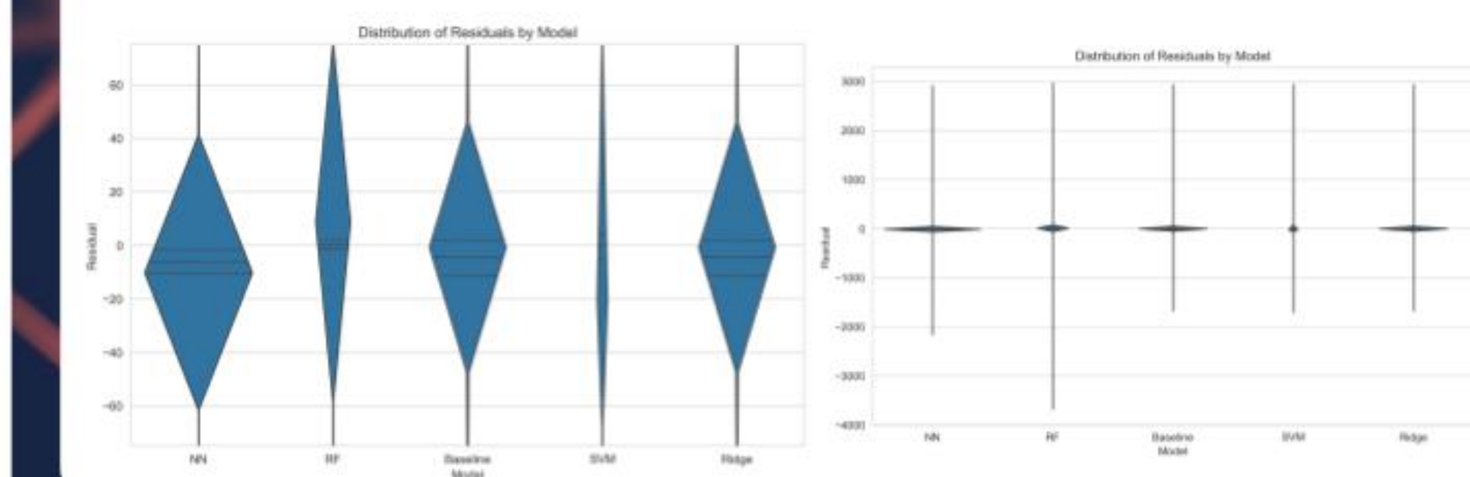


Results

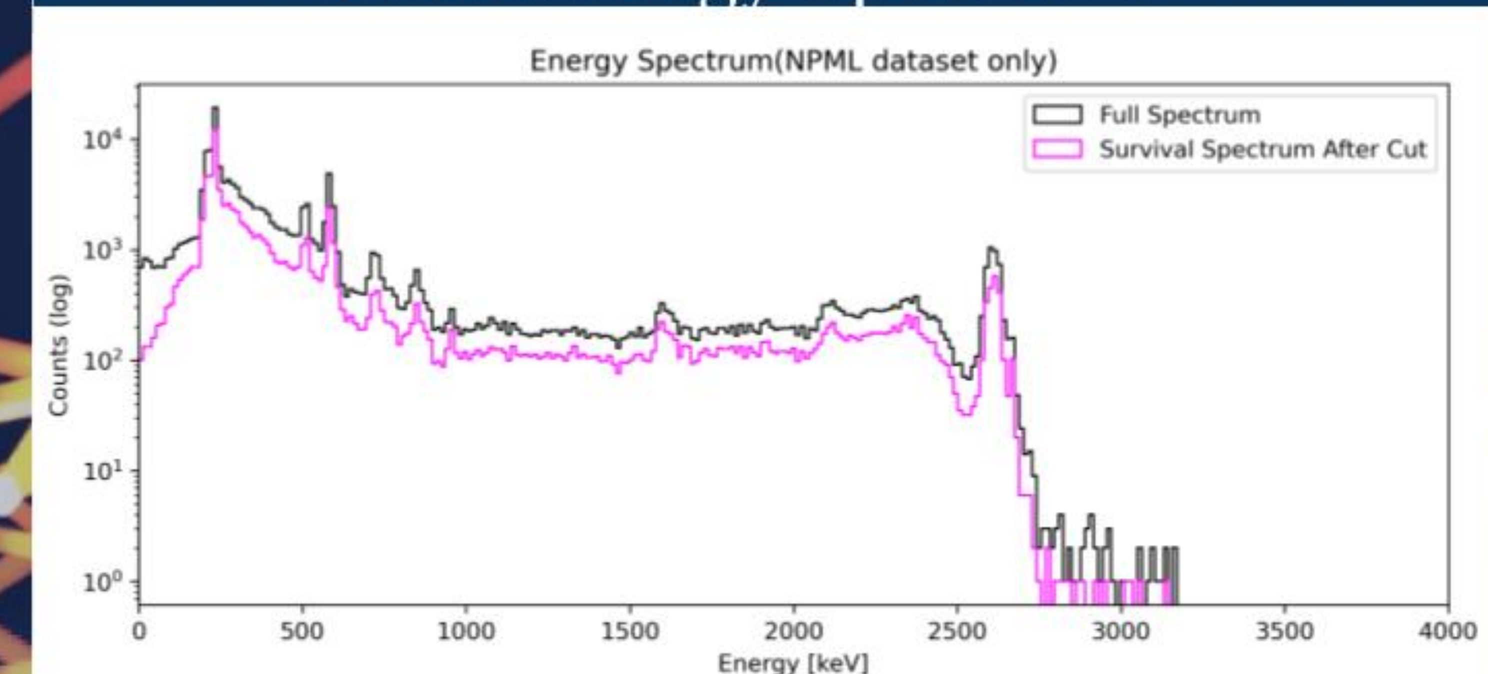
Model†pia	MSE	PMF (-75 to +75)
Baseline	4706.4220	0.9796
Ridge	4706.4220	0.9796
SVM	4759.4716	0.9774
RF	1860.8112	0.9878
NN	1507.1353	0.9896

Conclusion

- Neural Networks performed significantly better than the other regression models with the smallest variance in residuals.
- While the second best model being RF having a similarly low MSE, we can see in the violin plot below that there is a larger variance in the residuals.
- Neural Networks have a similar range of residuals compared to the other models with the exception of RF.



Final Energy Spectrum



The full spectrum is the distribution of our predicted energy labels on the original NPML which contains 160k rows The survival spectrum is the distribution of predicted energy labels on the subset of original NPML that passes all the PSD (classification) analysis labels.

UC San Diego
HALICIOĞLU DATA SCIENCE INSTITUTE