Applying Machine-Learning Methods to Laser Acceleration of Protons: Lessons Learned From Synthetic Data

Authors: Ronak Desai, Thomas Zhang, John J. Felice, Ricky Oropeza, Joseph R. Smith, Alona Kryshchenko, Chris Orban, Michael L. Dexter, Anil K. Patnaik

Publication: Contributions to Plasma Physics (2025)

Presented by Aadit Singal

Abstract

- This study compares three ML methods **Neural Network (NN)**, **Support Vector Regression (SVR)** and **Gaussian Process Regression (GPR)** trained on synthetic data from a modified Fuchs et al. (2005) model of Target Normal Sheath Acceleration (TNSA) of protons

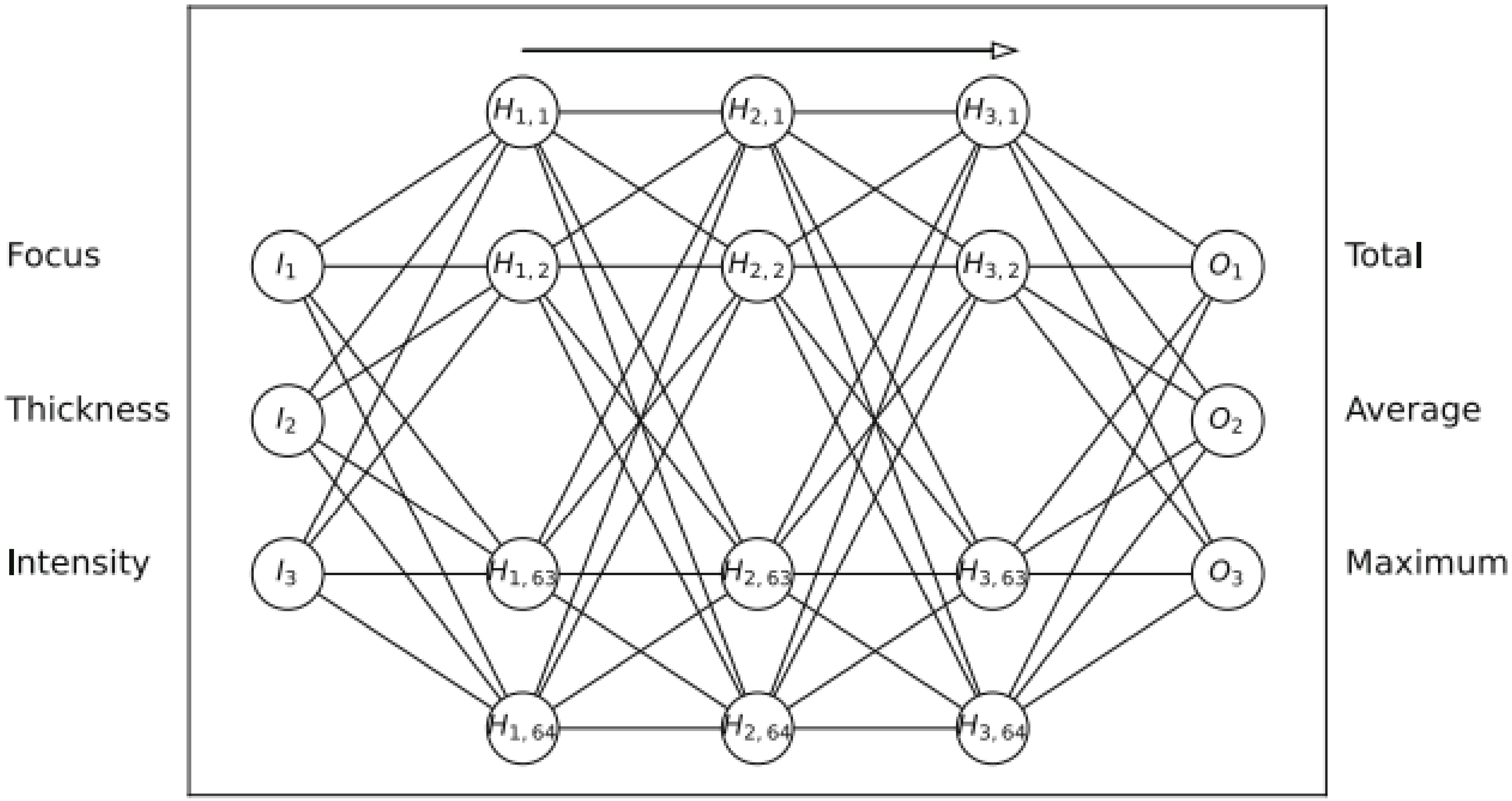
- Focus is on predicting peak, total and average proton energies plus laser to proton energy conversion efficiency to support optimization of laser parameters for desired proton spectra in high repetition rate systems.

- **Key finding:** SVR delivered the best balance of accuracy and GPU efficiency even with noise levels up to 30 percent and shows strong potential for real time control and optimization in upcoming high repetition rate laser experiments.

Synthetic Data Approach

- Based on the well-known **Fuchs et al. 2005** model of proton acceleration, but modified to better match experiments (multiplier increased from 1.3 to 4.0).
- **Inputs:** laser intensity, wavelength, pulse duration, target thickness, spot size, focal distance.
- **Outputs:** maximum proton energy, total proton energy, average proton energy, conversion efficiency from laser energy to proton energy.
- Generated 25,000 points with intensities from 10^17 to 10^19 W/cm², target thicknesses 0.5–10 μm, and focal distances -10 to +10 μm.
- Added 5–30% "Gaussian" noise to mimic real experimental uncertainty, trained using an 80-20 split between training and testing data.

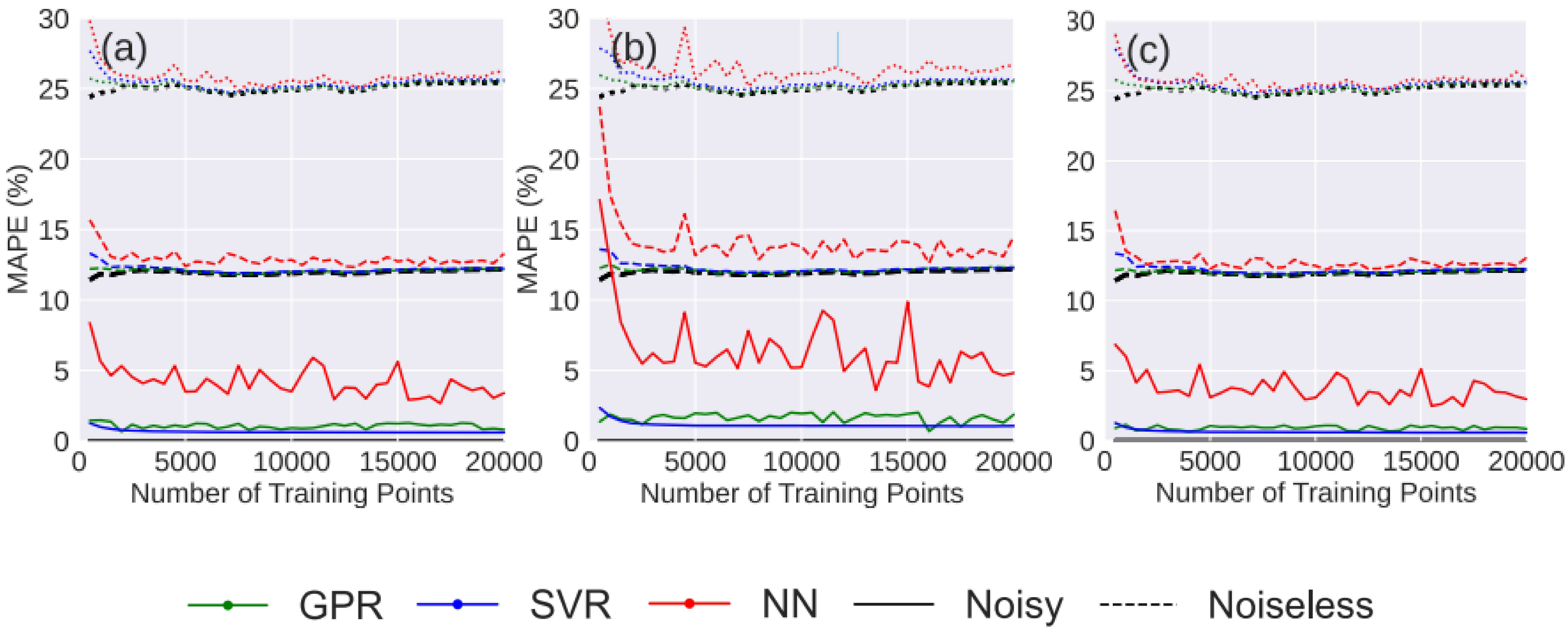
Neural Network Architecture



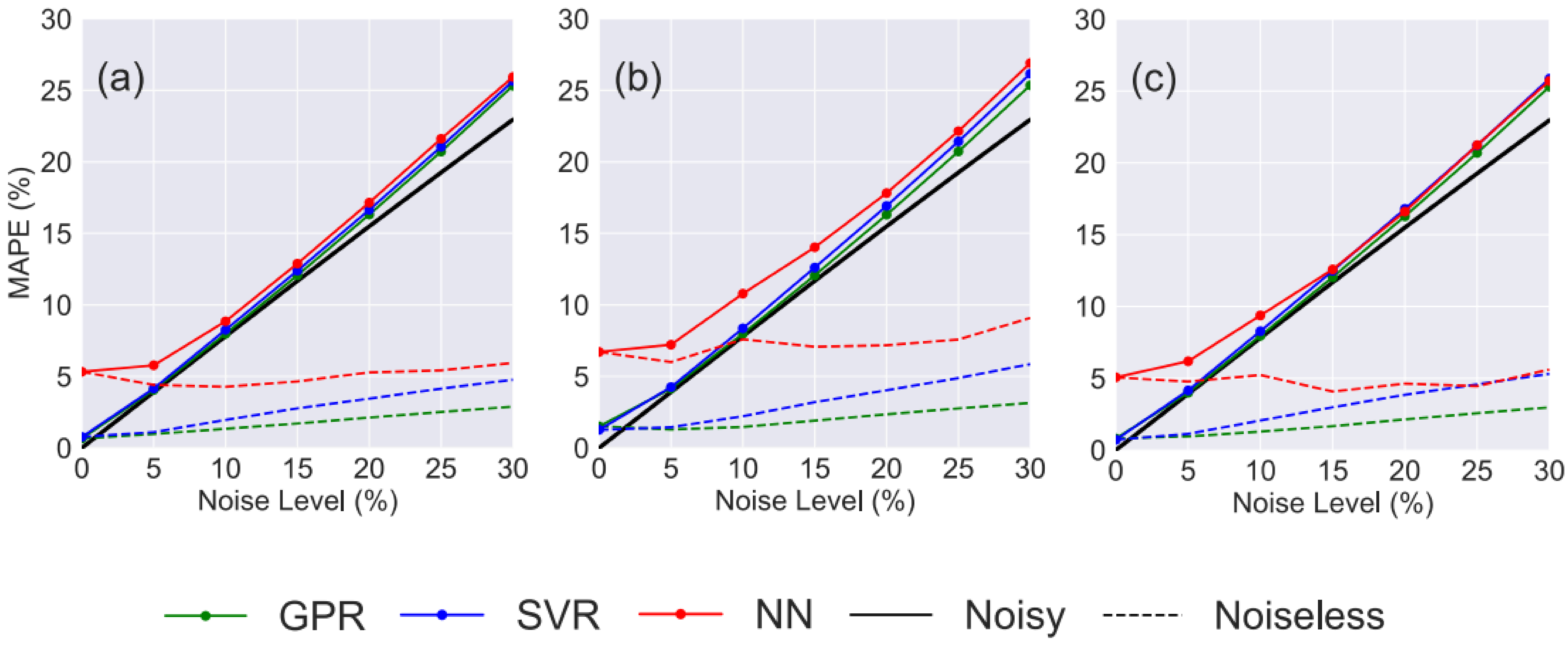
The three inputs (intensity, target thickness, and target focal position) and three outputs (maximum, total, and average proton energy) are fully connected by three hidden layers of 64 neurons each resulting in ***8771 total parameters***.

ML Models Overview

Model	Approach	Performance
Neural Network (NN)	3 hidden layers × 64 neurons, LeakyReLU activation, Adam optimizer. Learns input–output mapping directly from data.	Needs larger dataset. Underperformed at 20k points; lowest GPU memory but less accurate at low data/high noise.
Support Vector Regression (SVR)	Kernel-based regression (Radial Basis Function - RBF) with tolerance band. Implemented with cuML.	Best trade-off: highest accuracy and fastest training on single GPU at 20k points. Low memory use (~1.6 GiB).
Gaussian Process Regression (GPR)	Uses RBF kernel in GPyTorch LR = 0.2 Provides uncertainty estimates.	Comparable accuracy but much slower and far more memory-intensive (14 GiB, $O(N^3)$ scaling).



MAPE versus number of training points from ML model predictions for **(a) max proton energy**, **(b) total proton energy**, **(c) average proton energy** and **noisy testing data**. Each panel shows results from (solid) 0%, (dashed) 15% and (dotted) 30% added noise in the data. Black lines with different line types indicate the MAPE between the noisy and noiseless data. Because we only compare ML models to noisy data in this figure, these black lines indicate the best that any ML model could conceivably do.



Solid lines show the typical MAPE in **(a) maximum proton energy**, **(b) total proton energy**, and **(c) average proton energy** when the ML models (which were trained on 2000 synthetic data points with noise) are evaluated on data with different levels of noise. Dashed lines show the typical error when those same ML models are evaluated on noiseless test data. Black solid lines indicate the MAPE between the noisy and noiseless data.

Conclusion and Next Steps

- **Current Best:** SVR shines with 20k data due to lower MAPE and efficiency, outperforming both NN & GPR for now.
- **NN's Future:** NN will perform better with larger datasets to effectively utilize its 8771 parameters, addressing overfitting seen with 20k points and unlock its full potential.
- **GPR's Role:** GPR gives uncertainty insights but struggles with speed as data grows, due to its time complexity of O(N^3).

Enhancing the Neural Network

1. Larger Dataset - A bigger dataset will help the NN use its parameters fully for better accuracy.
2. Deeper/Wider Layers - We can experiment with more layers (4-5) and 64–128 neurons to match growing data.
3. Dropout Addition - 10–20% dropout for faster, consistent learning.
4. *Synthetic Data is smooth, Real Data is messier.* We can explore 1D PIC simulations as well.

Thank You