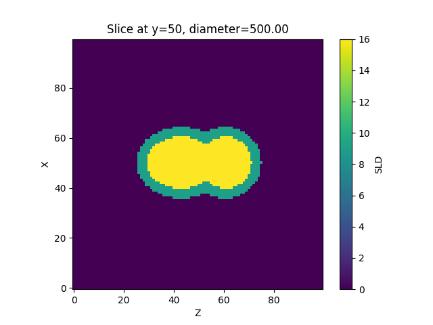
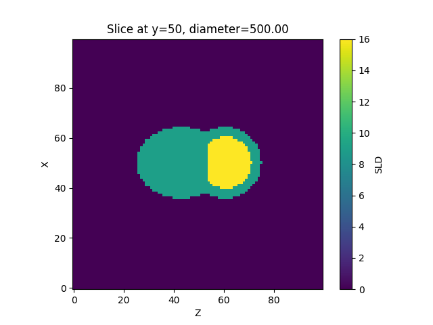
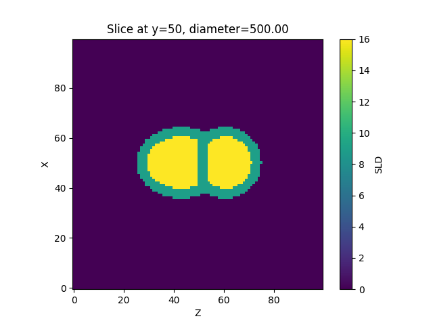
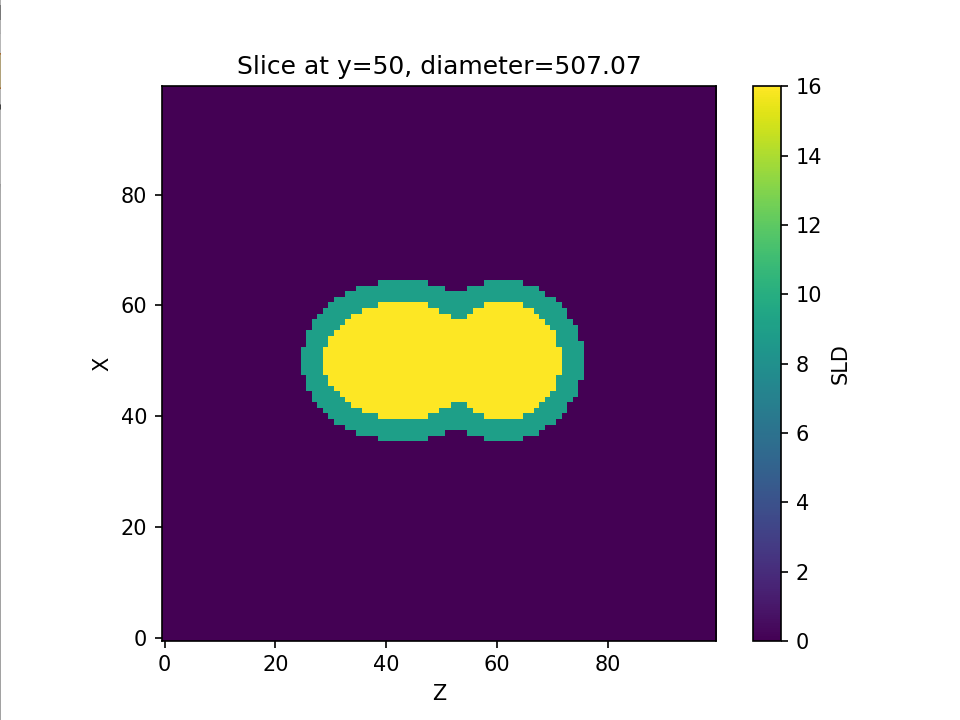
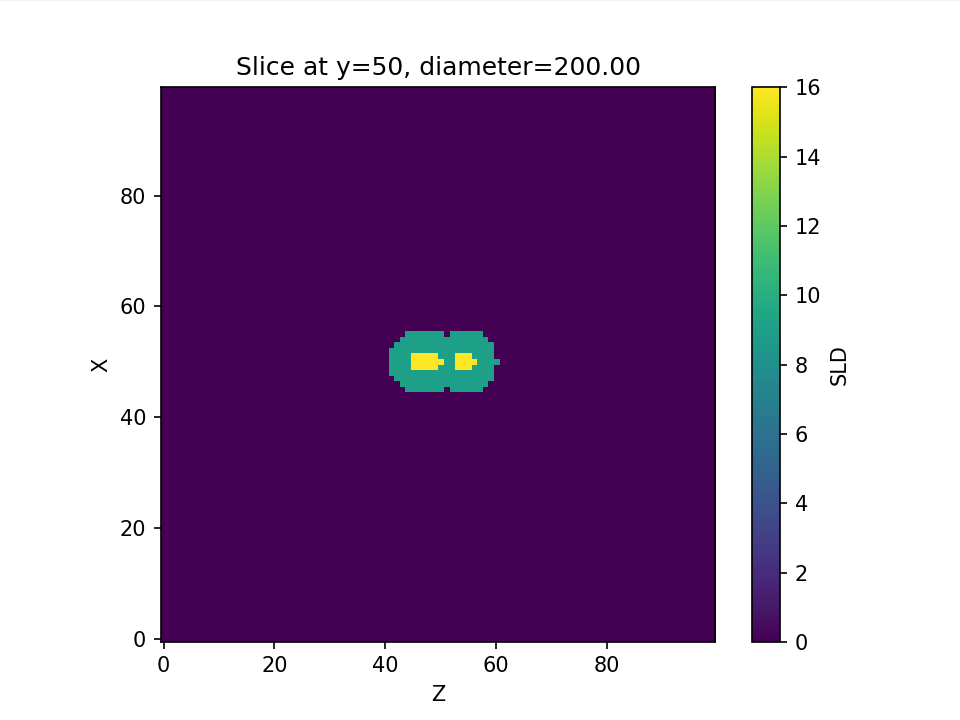
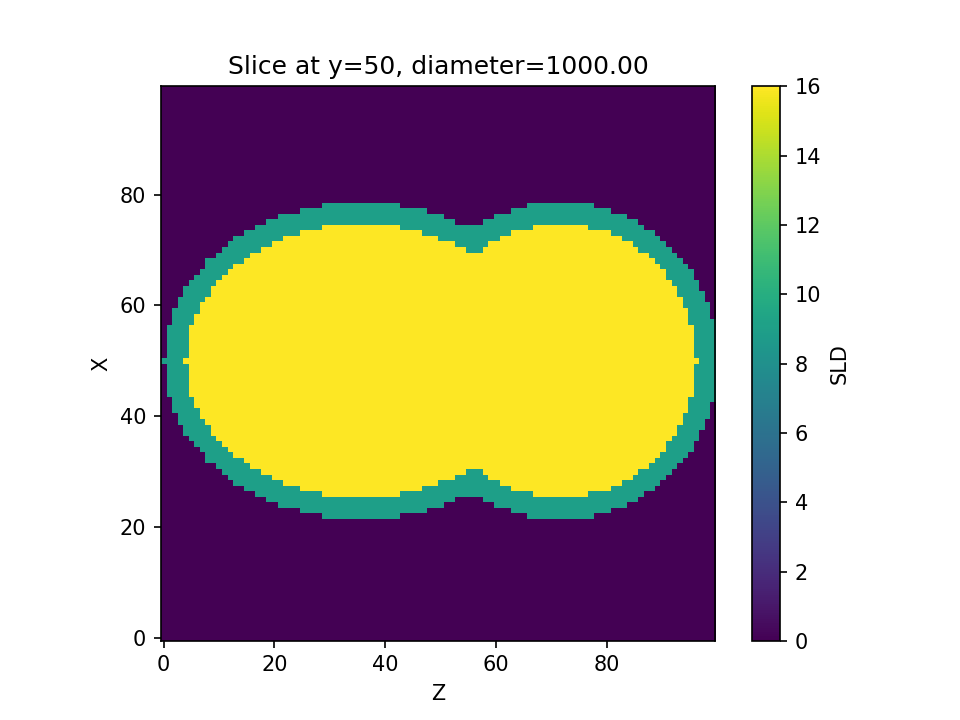
To make sure everyone understands the basic workflow of this project, here is a simple flow chart for this project:

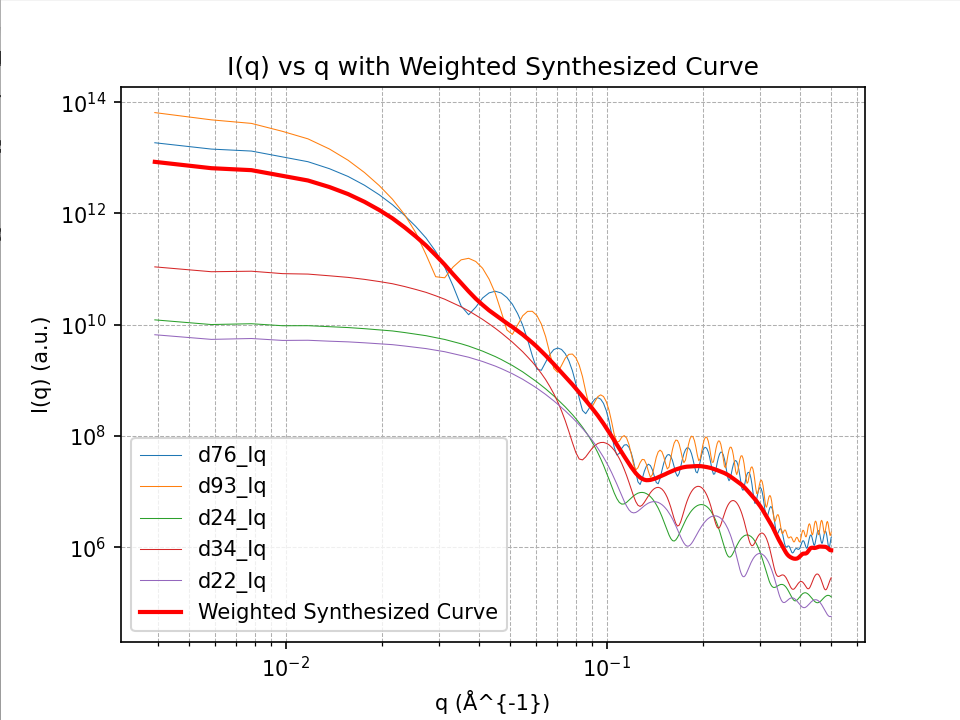
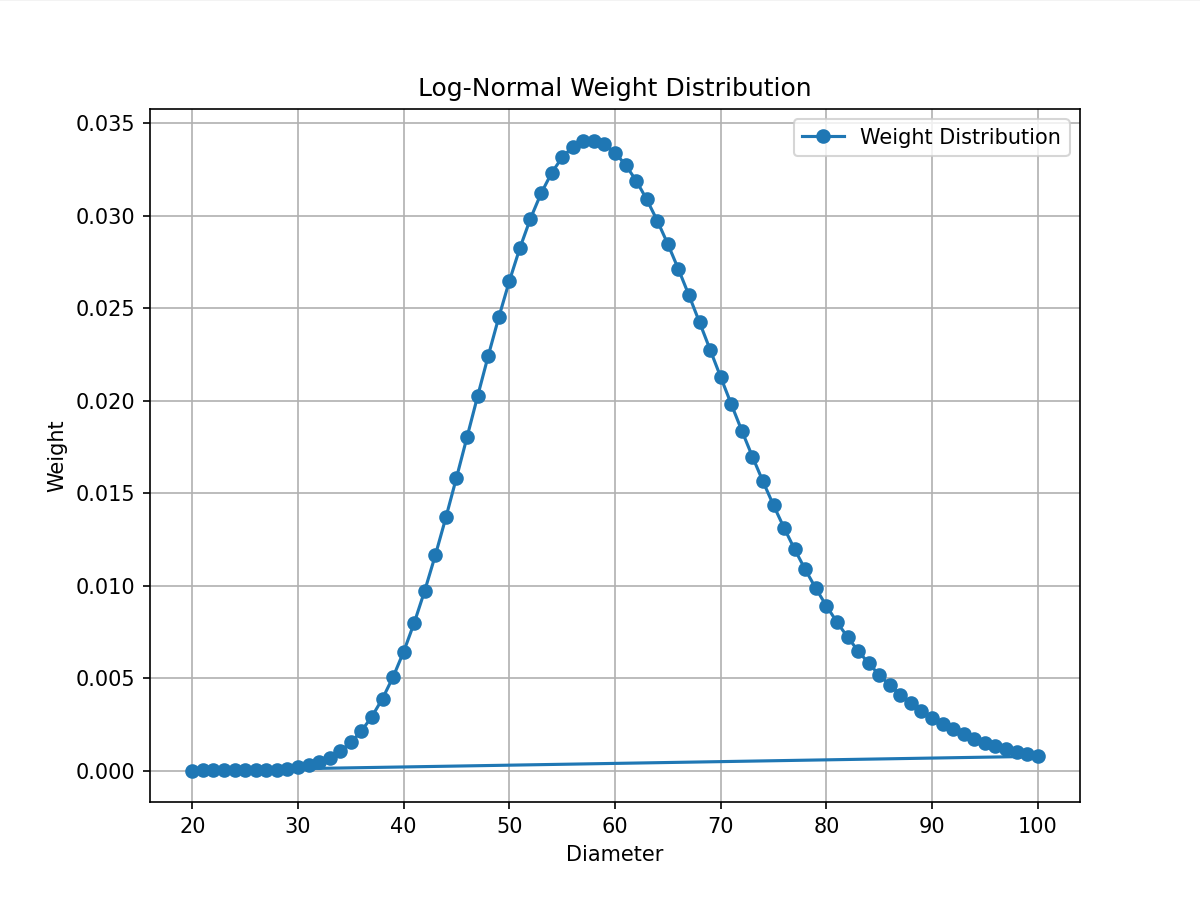
1. We have 3 types of LNP models with different shapes. At this stage, since we are focusing on predicting the size distribution of LNPs in sample solution of single type, only one model is considered.



1. According to [1], LNPs in the size range of 40 to 80 nm are more desired in production. Therefore, the prediction model is designed to predict the size distribution within the range of 20 to 100nm. At this stage, there is no other alternative way to enlarge this range. For example, if we want more simulated SAXS results from models in the size range of 5nm to 200nm, we can use 2 pixels to represent 1nm, which means a 5nm model will need 10 pixels, then a 200nm model will need 400 pixels. In this case, 400 pixels will cause an Edge effect during the calculation of FFT results, unless we can pad the model into a 2400-pixel-big box, which outranges the computation power of my PC.

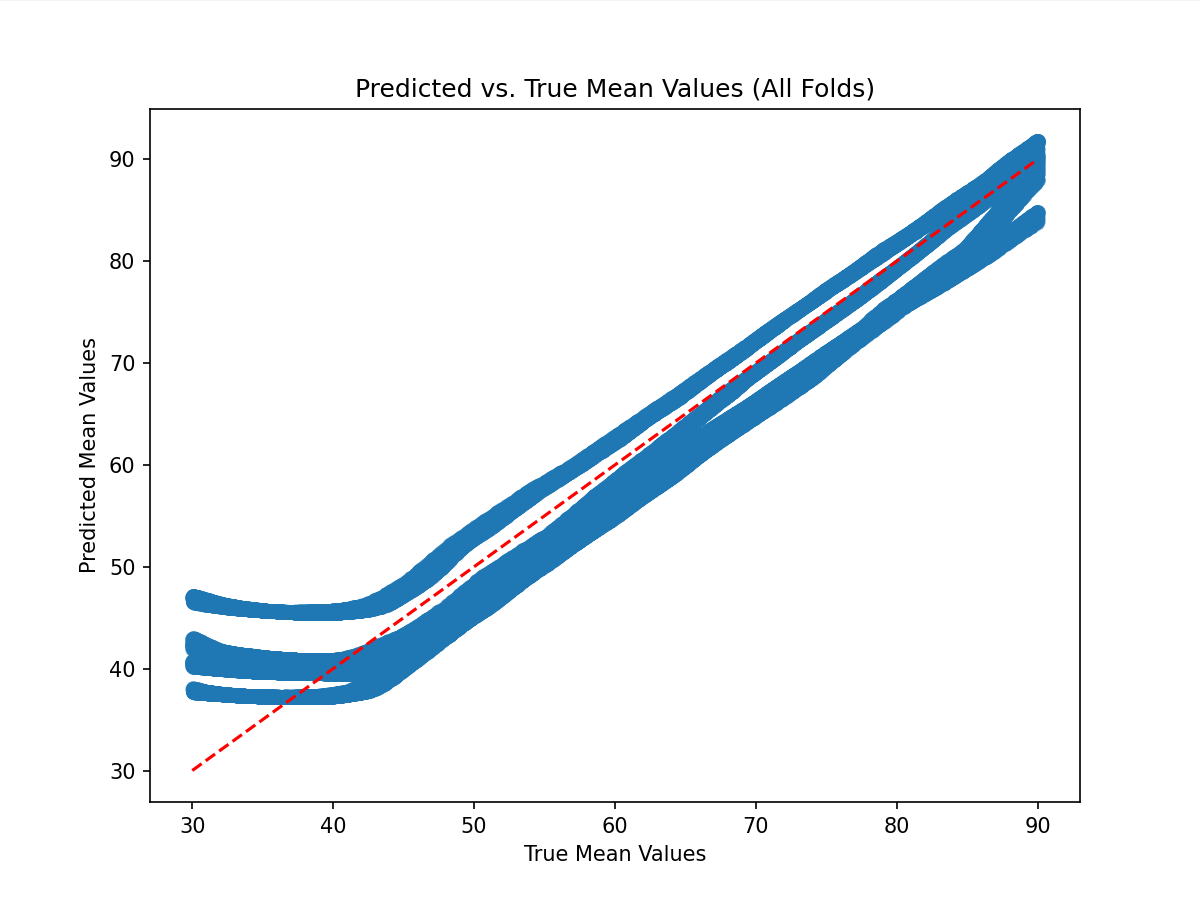
1. 100 models uniformly spaced in the size range of 20nm to 100nm are generated for SAXS simulation. As mentioned in (2), one pixel represents one-nanometer scheme. Therefore, all models are padded into a 512-pixel-big box before calculating the FFT results. After this process, 100 distinct SAXS curves are collected and stored as baseline data.
2. According to Prof. Doerschuk’s analysis and documentation of Sasview, the simulation algorithm of polydispersity is based on the relative number of particles, which means in our project, where the dilute aqueous solution is the subject, the size distribution should be also based on the number of particles. According to [2], the log-normal distribution is more common in number-of-particles-based size distribution. Thus, 100,000 pairs of (μ, σ) (uniformly distributed over 30-90 and 1-5) are generated corresponding to 100,000 different shapes of distribution. One example is shown left below. Each distribution has a “possibly unique mapping” to a specific synthesized curve shown right below (The bold red curve). Therefore, 100,000 data in total, each of them having two labels (μ, σ).



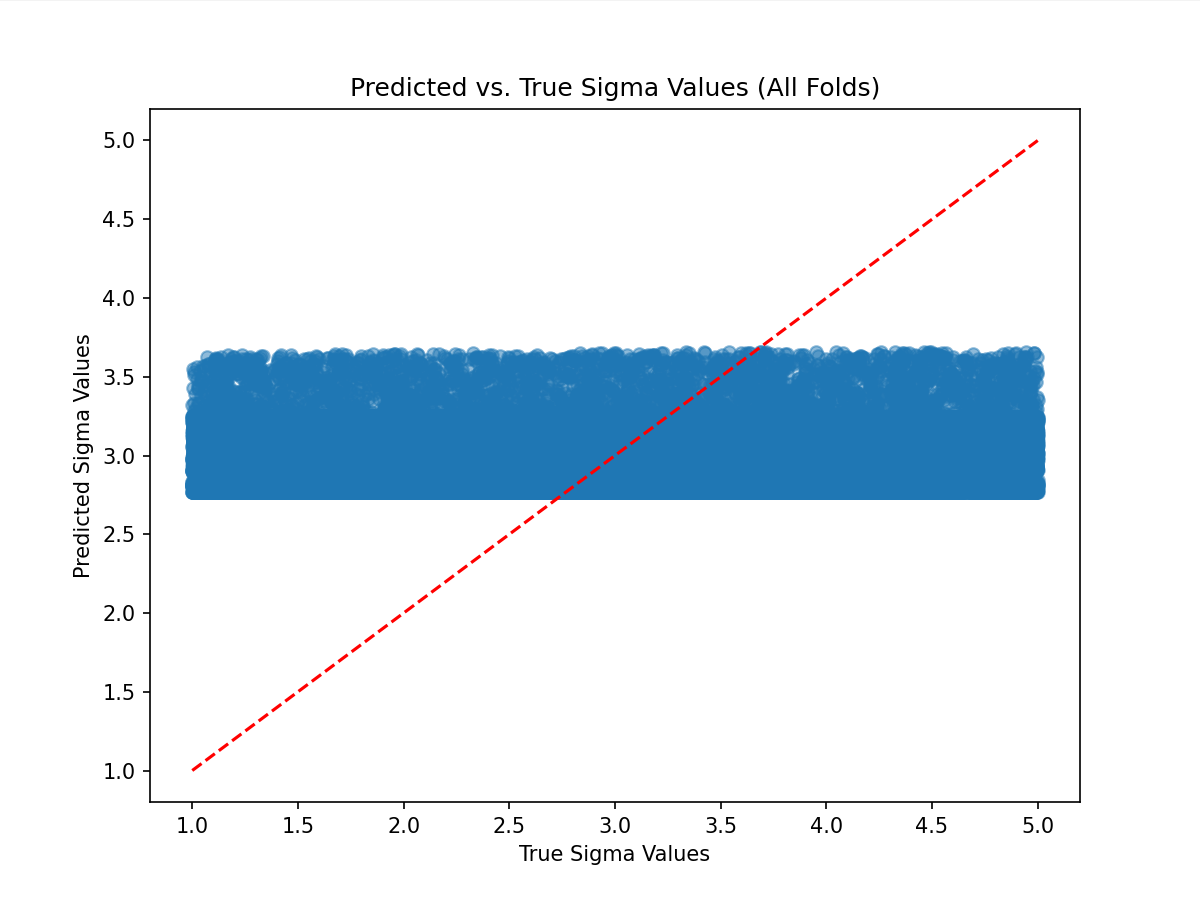
1. Our aim is to train a model (or models) that can predict (μ, σ) pair based on the input synthesized SAXS curve (bold red curve). The version 1 prototype prediction performance is shown below:

(4-fold / lr=1e-3 / BS=512 / 10 epochs / patience=4 / L1loss)

SimpleCNN Prediction for μ

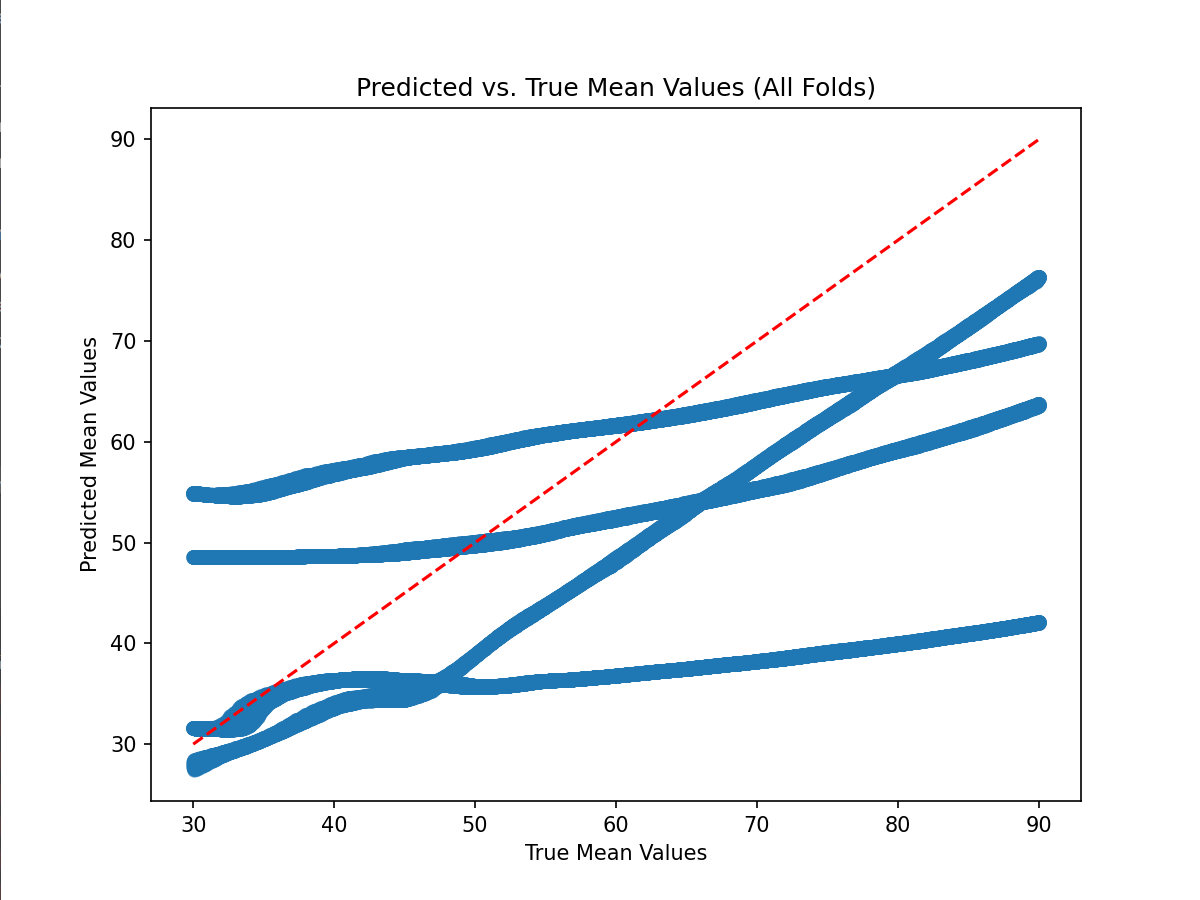


SimpleCNN Prediction for σ

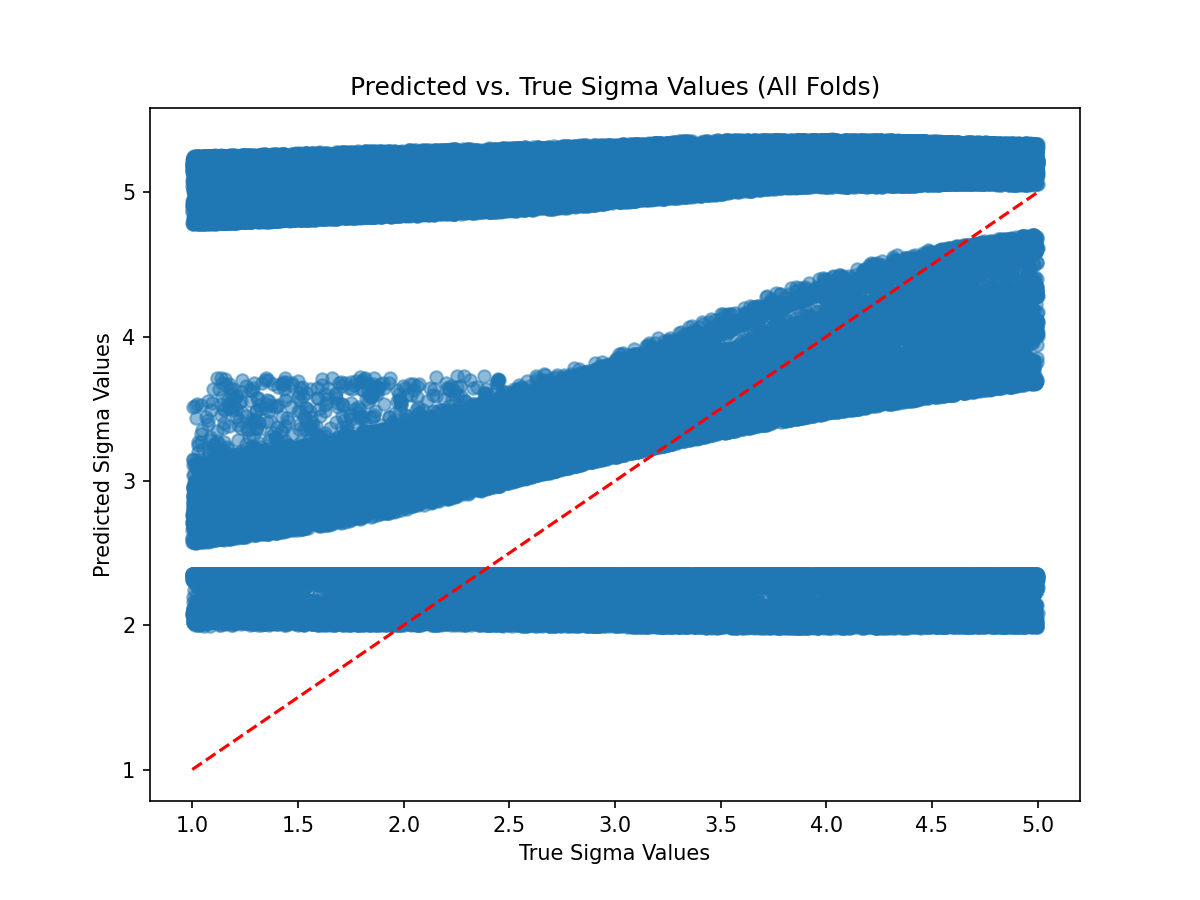


(4-fold / lr=1e-3 / BS=512 / 10 epochs / patience=4)

SimpleResNet Prediction for μ



SimpleCNN Prediction for σ



[1] Particles in Biopharmaceutical Formulations, Part 2: An Update on Analytical Techniques and Applications for Therapeutic Proteins, Viruses, Vaccines and Cells

[2] THE REAL ORIGIN OF LOGNORMAL SIZE DISTRIBUTIONS OF NANOPARTICLES IN VAPOR GROWTH PROCESSES