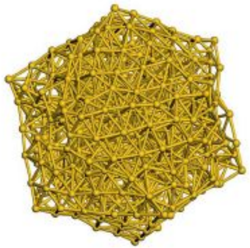
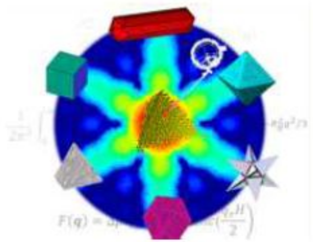
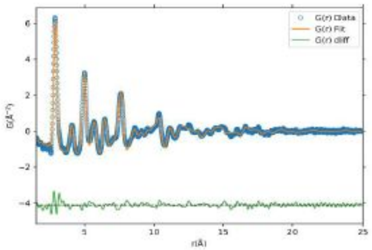
Machine learning-assisted structural characterization of nanoparticles using X Ray scattering







By:

Pierre Boissier

Maïmouna Gadji

OUTLINE

I.CONTEXT

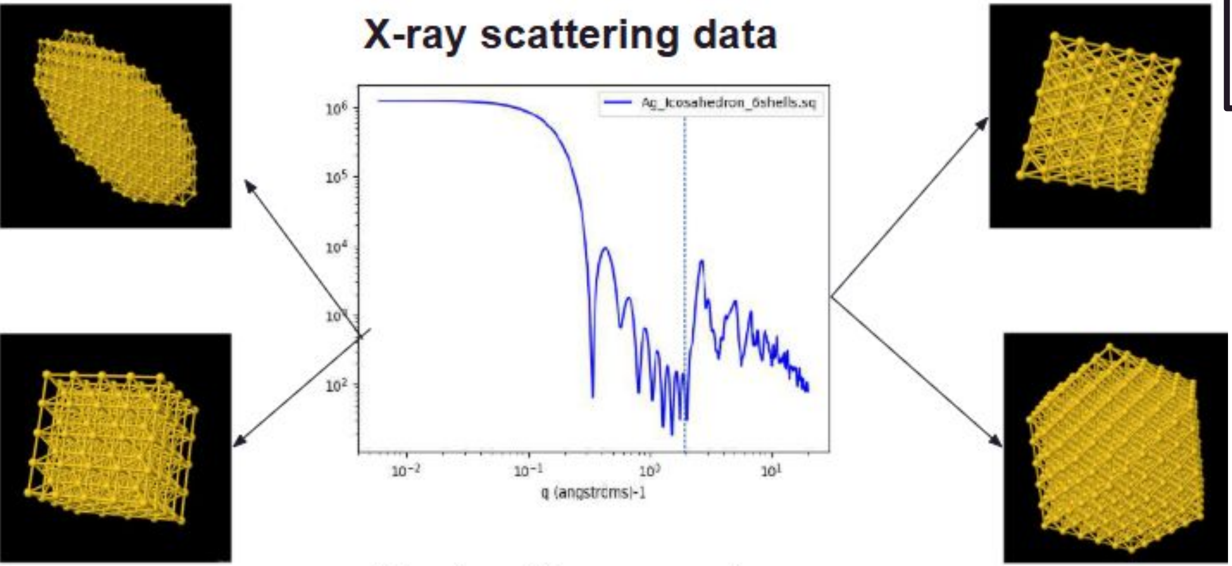
BiMAn Project:

- **Objective**: Develop efficient magnetic materials without rare-earth elements.

Various applications in **energy transition** like renewable energy system or data storage.

II. AIM OF THE PROJECT

Prediction of nanoparticle shape and size from X-ray scattering data



$$q = \frac{4\pi}{\lambda}\sin(\theta)$$

scattering angle

X-axis: q (Anstrome -1)
Y-axis: Intensities

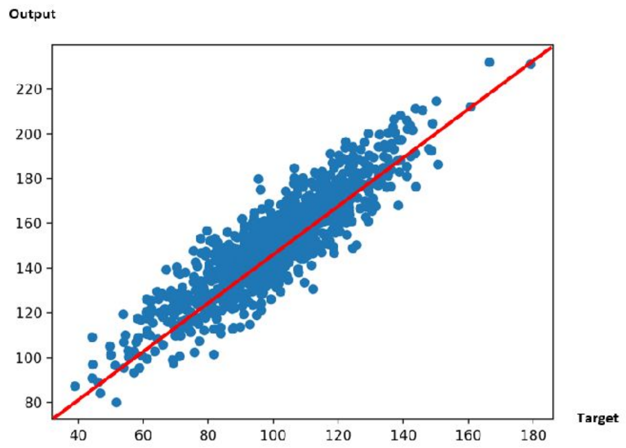
SWAXS METHOD

III.Size and Shape prediction

How?

- Size
  - n_atoms
  - Eq_radius
  - Size_parameter
- Shape
  - shape

III.1 Linear regression methods



Ridge Regression

$$J(\beta) = \|y - X\beta\|_2^2 + \alpha\|\beta\|_2^2$$

Logistic Regression

$$p_k = \frac{\exp(\beta_k^T x)}{\sum_{j=1}^{9} \exp(\beta_j^T x)}$$

$$\mathcal{L}(\beta) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{9} y_{i,k}\log(p_{i,k})$$

Idea of linear regression :

Find a linear relation between inputs and outputs

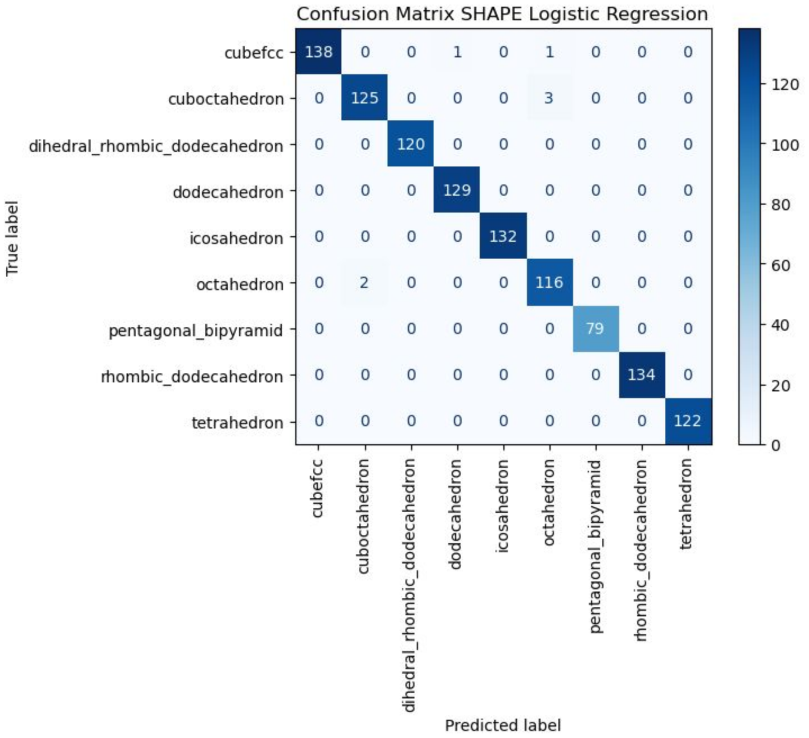
RESULTS: Linear reg

Size prediction = 3 quantitatives features to predict → Ridge Regression

Features	MSE	Variance of test data	Q2
log_n_Atoms	0.00015	3.196	0.999
Size_parameter_1	0.846	7.106	0.881
sqrt_log_eq_radius	0.310	0.695	0.554

RESULTS: Linear reg

Shape prediction = 1 qualitative features to predict → Logistic Regression



Accuracy = 99.4 %

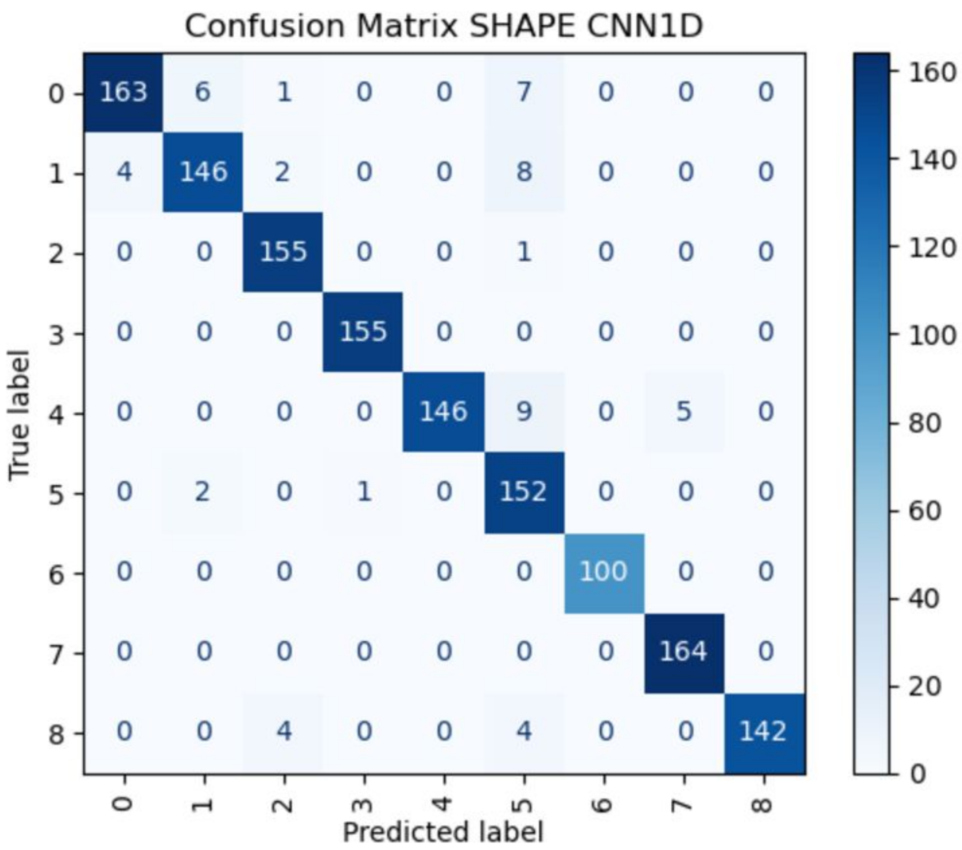
III.2 Convolutional neural network 1D (CNN 1D)

Layer (type)	Output Shape	Param #
conv1d_8 (Conv1D)	(None, 1997, 32)	128
max_pooling1d_8 (MaxPooling1D)	(None, 998, 32)	0
conv1d_9 (Conv1D)	(None, 996, 64)	6,208
max_pooling1d_9 (MaxPooling1D)	(None, 498, 64)	0
flatten_4 (Flatten)	(None, 31872)	0
dense_4 (Dense)	(None, 9)	286,857

Idea:

Searching for meaningful patterns in intensity curve and make the shape or size prédiction from them

RESULTS: CNN1D



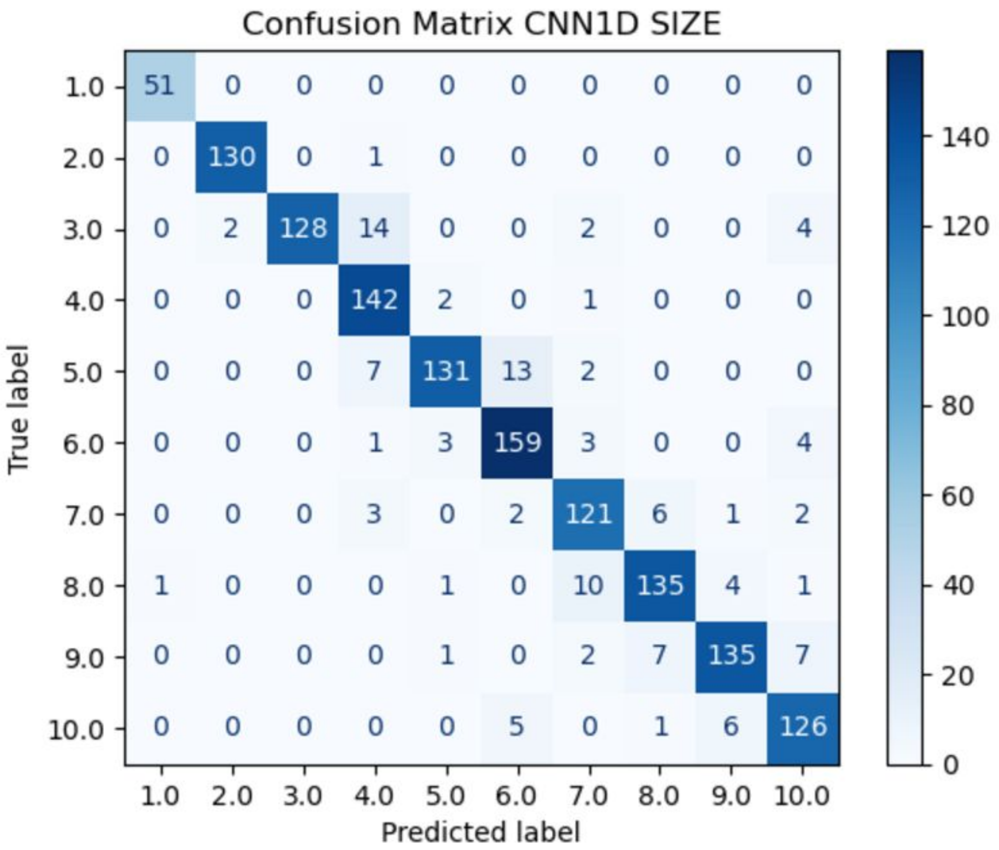
0	163	6	1	0	0	7	0	0	0
1	4	146	2	0	0	8	0	0	0
2	0	0	155	0	0	1	0	0	0
3	0	0	0	155	0	0	0	0	0
4	0	0	0	0	146	9	0	5	0
5	0	2	0	1	0	152	0	0	0
6	0	0	0	0	0	0	100	0	0
7	0	0	0	0	0	0	0	164	0
8	0	0	4	0	0	4	0	0	142
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**

Accuracy=96%

Labels

cubefcc 0
cuboctahedron 1
dihedral_rhombic_dodecahedron 2
icosahedron 3
octahedron 4
dodecahedron 5
pentagonal_bipyramid 6
rhombic_dodecahedron 7
tetrahedron 8

RESULTS: CNN1D



	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0
1.0	51	0	0	0	0	0	0	0	0	0
2.0	0	130	0	1	0	0	0	0	0	0
3.0	0	2	128	14	0	0	2	0	0	4
4.0	0	0	0	142	2	0	1	0	0	0
5.0	0	0	0	7	131	13	2	0	0	0
6.0	0	0	0	1	3	159	3	0	0	4
7.0	0	0	0	3	0	2	121	6	1	2
8.0	1	0	0	0	1	0	10	135	4	1
9.0	0	0	0	0	1	0	2	7	135	7
10.0	0	0	0	0	0	5	0	1	6	126

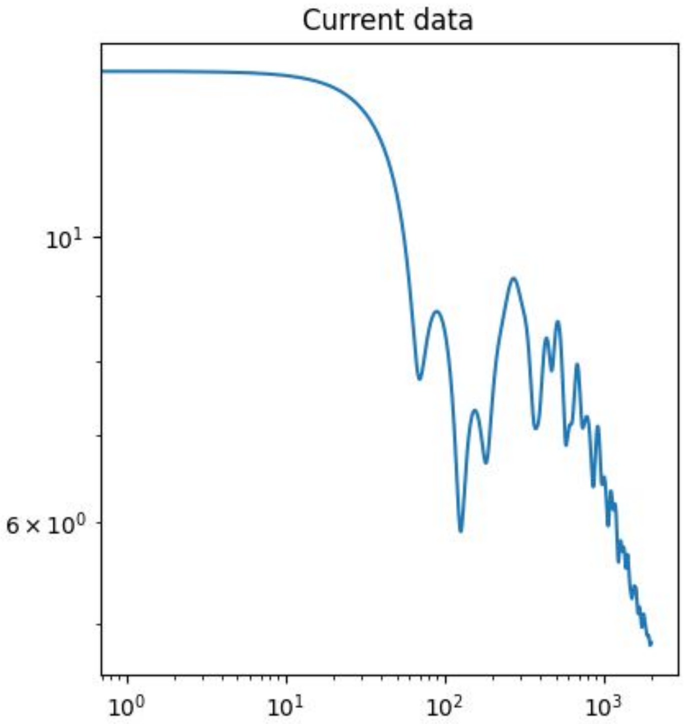
Accuracy=91%

IV. Testing the Robustness of Models

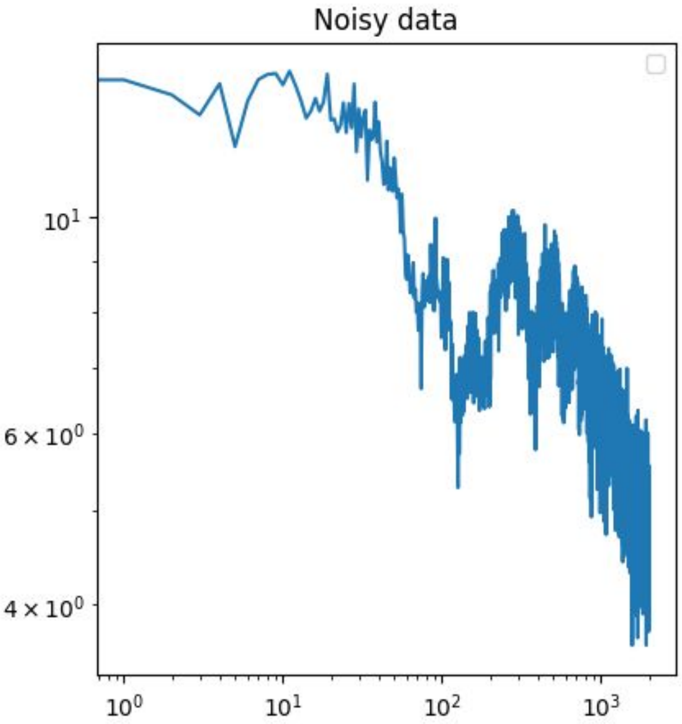
USING MORE REALISTIC DATA

Current data $\neq$ Realistic data.

I.V.1 Adding Noise

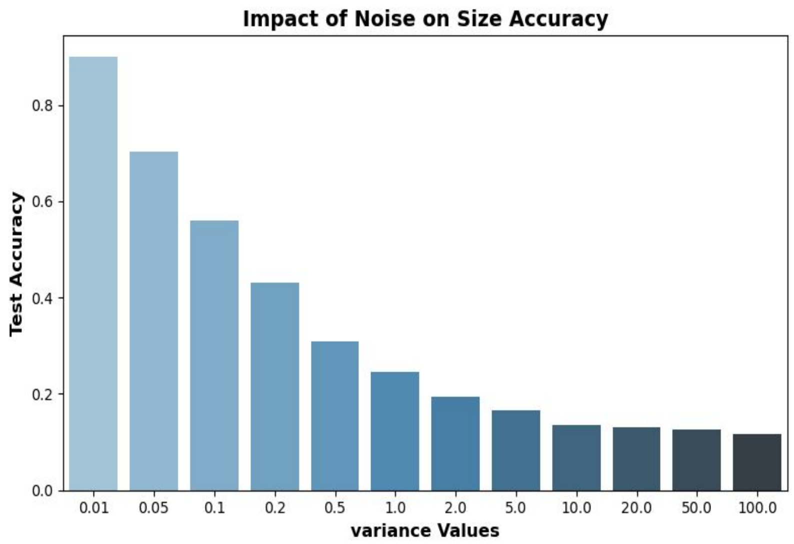


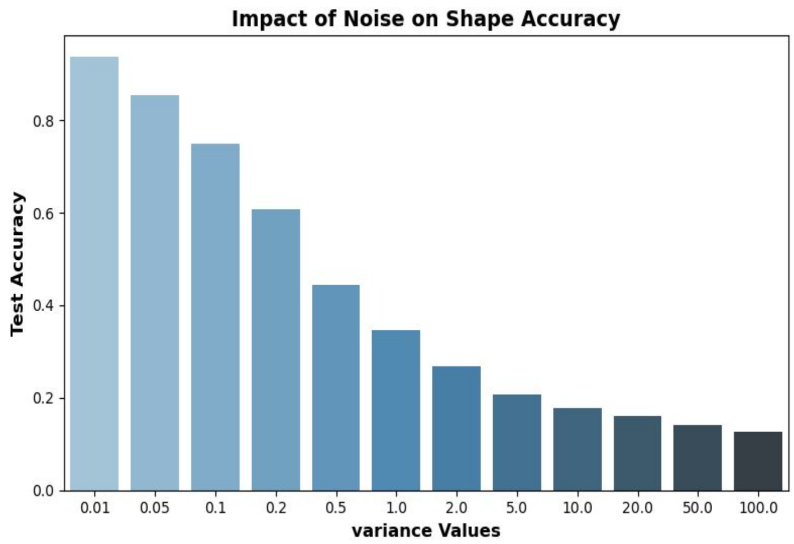
Gaussian noise



Impact of noise on prediction

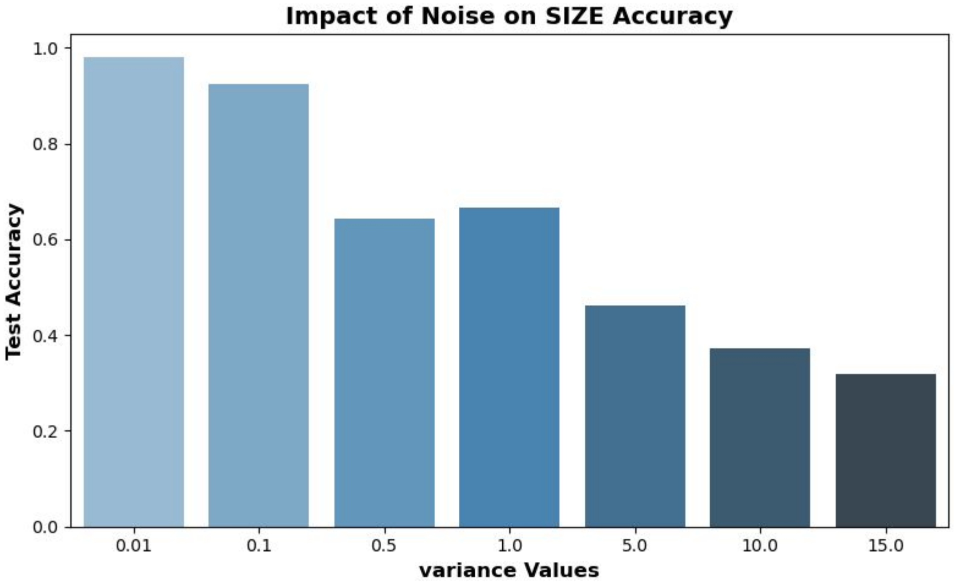
CNN1D model : Simple predictions on noisy data



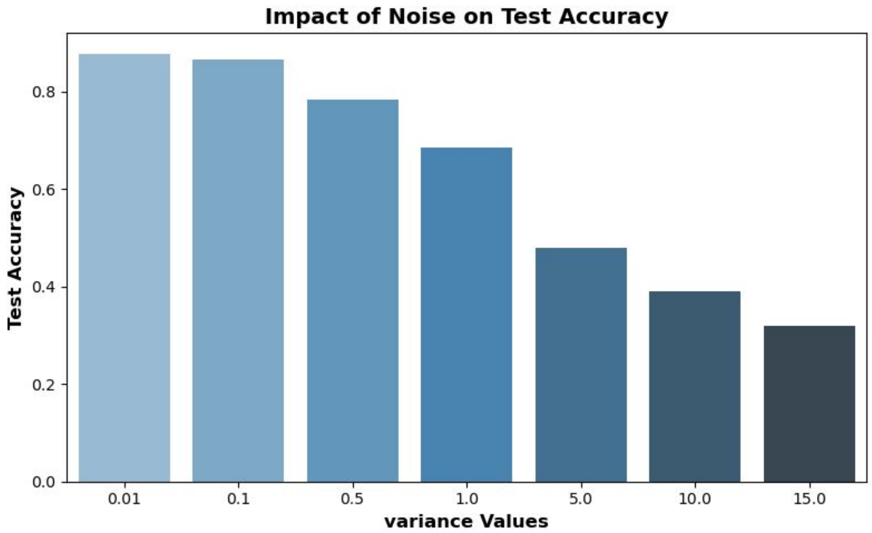


Impact of noise on prediction

CNN1D model : Training and prediction on noisy data



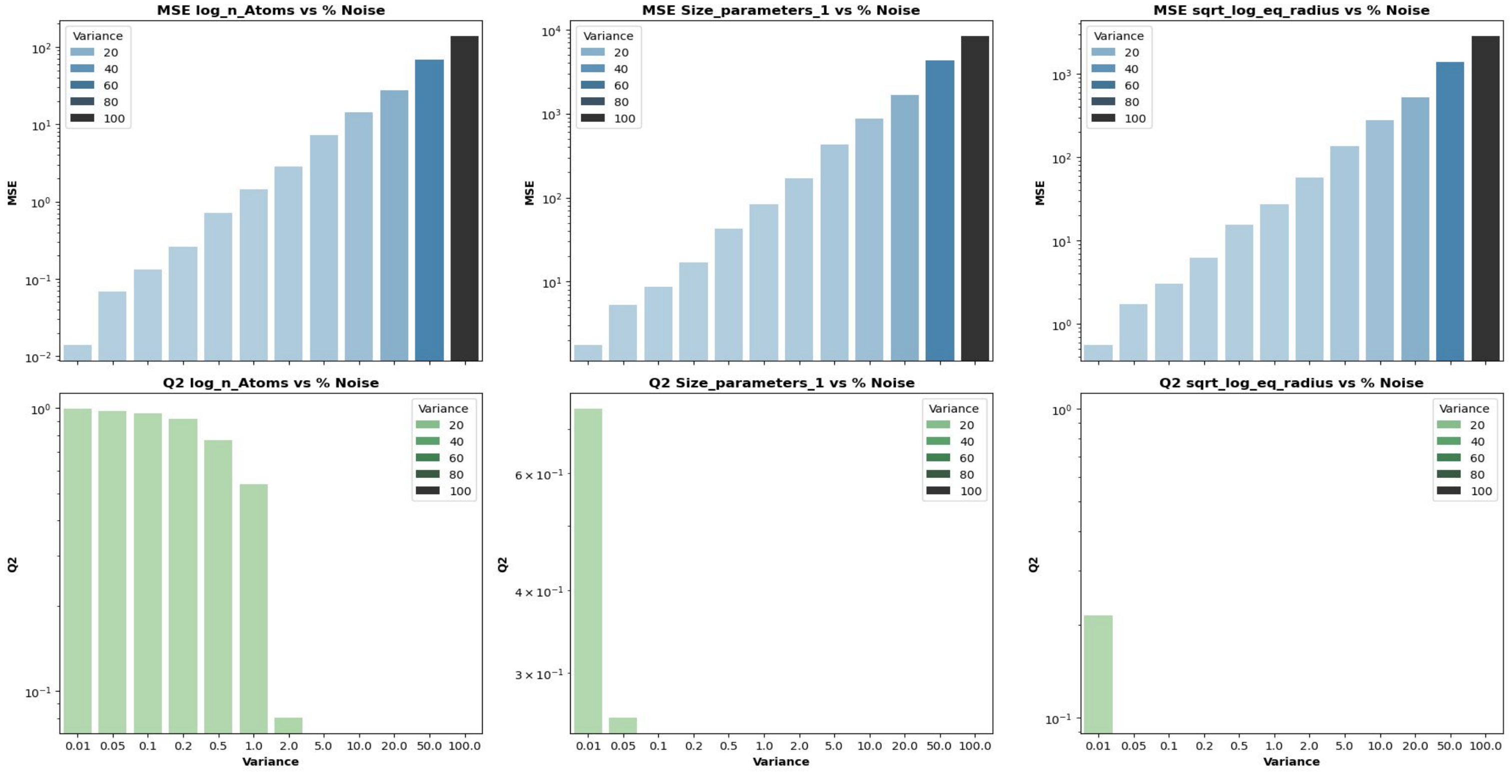
SIZE



SHAPE

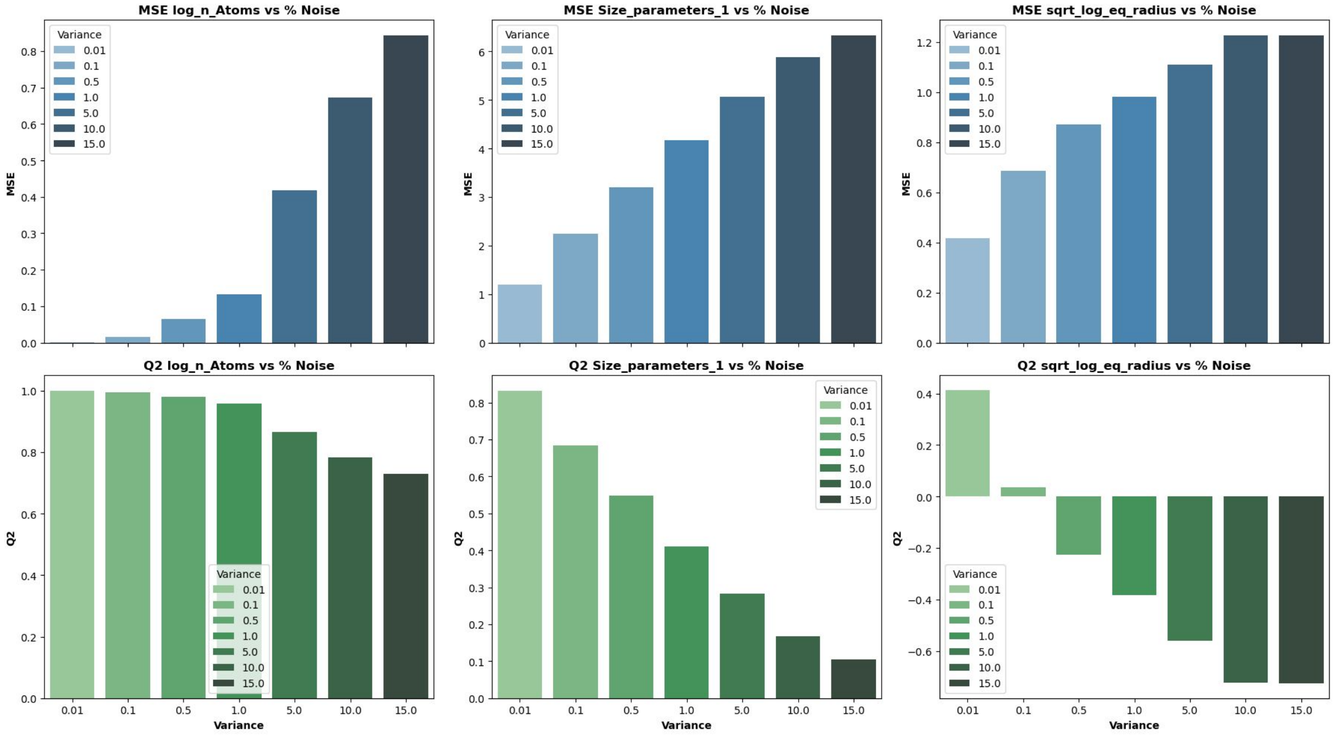
Impact of noise on prediction

Ridge Regression model: Simple predictions on noisy data



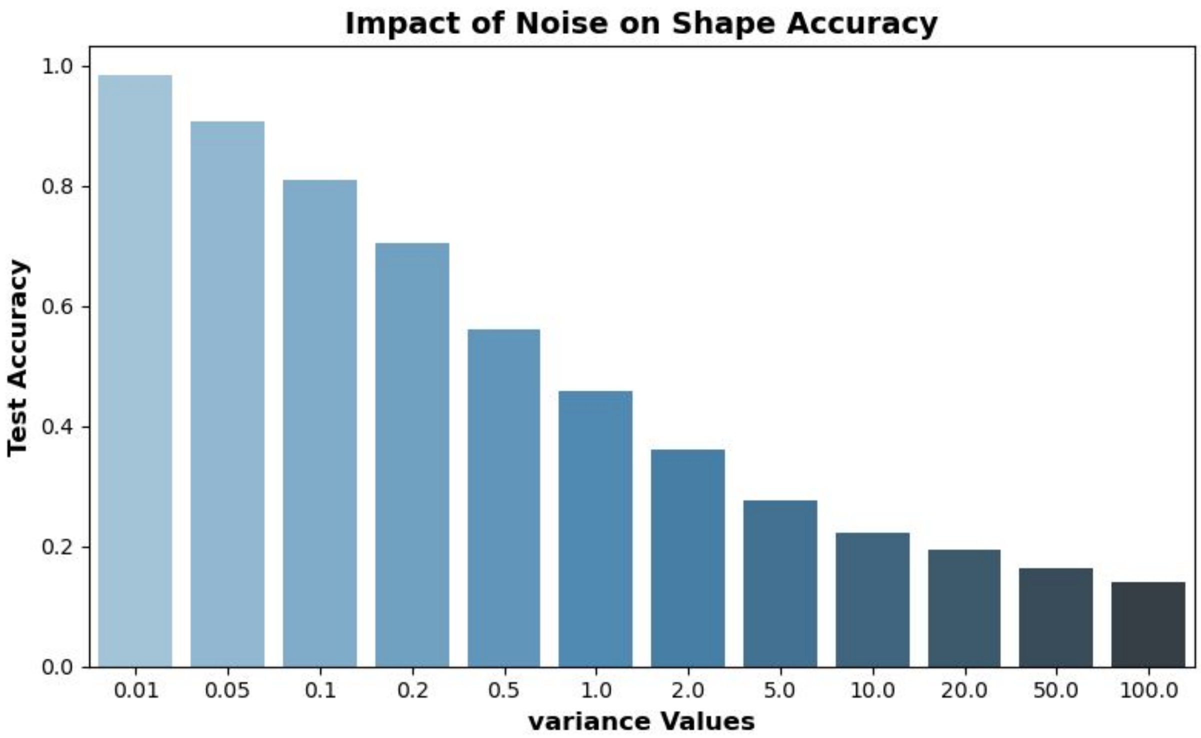
Impact of noise on prediction

Ridge Regression model: Training and predictions on noisy data



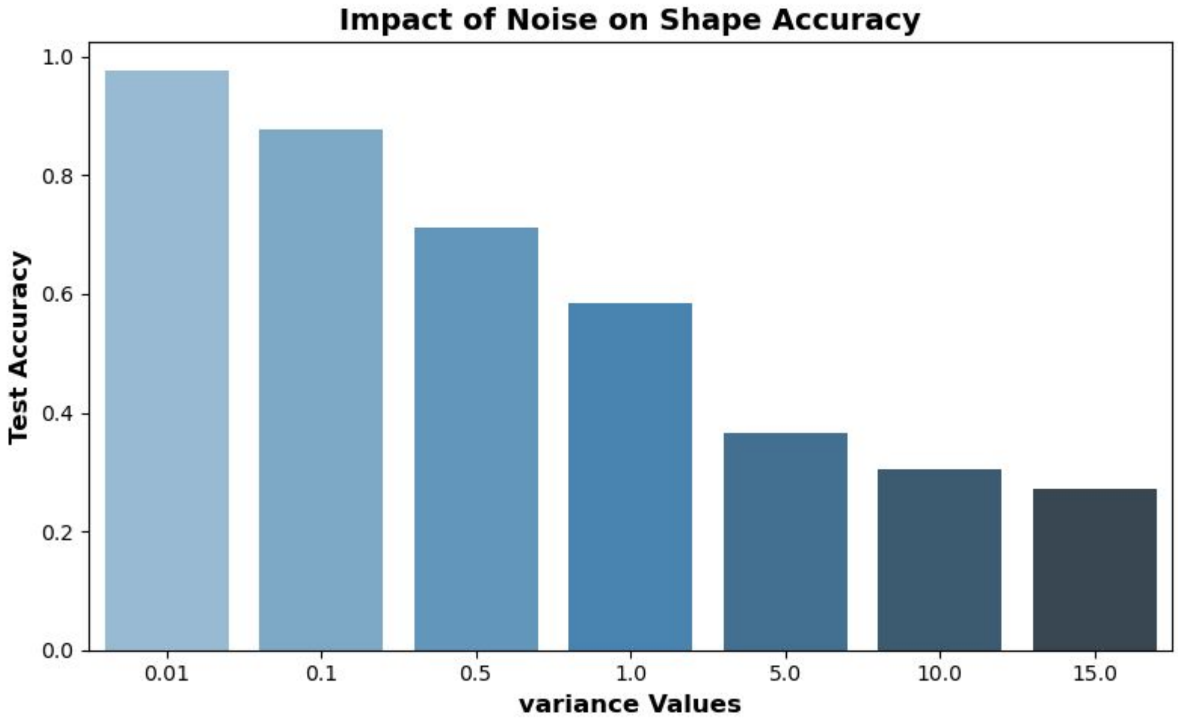
Impact of noise on prediction

Logistic Regression model: Simple predictions on noisy data

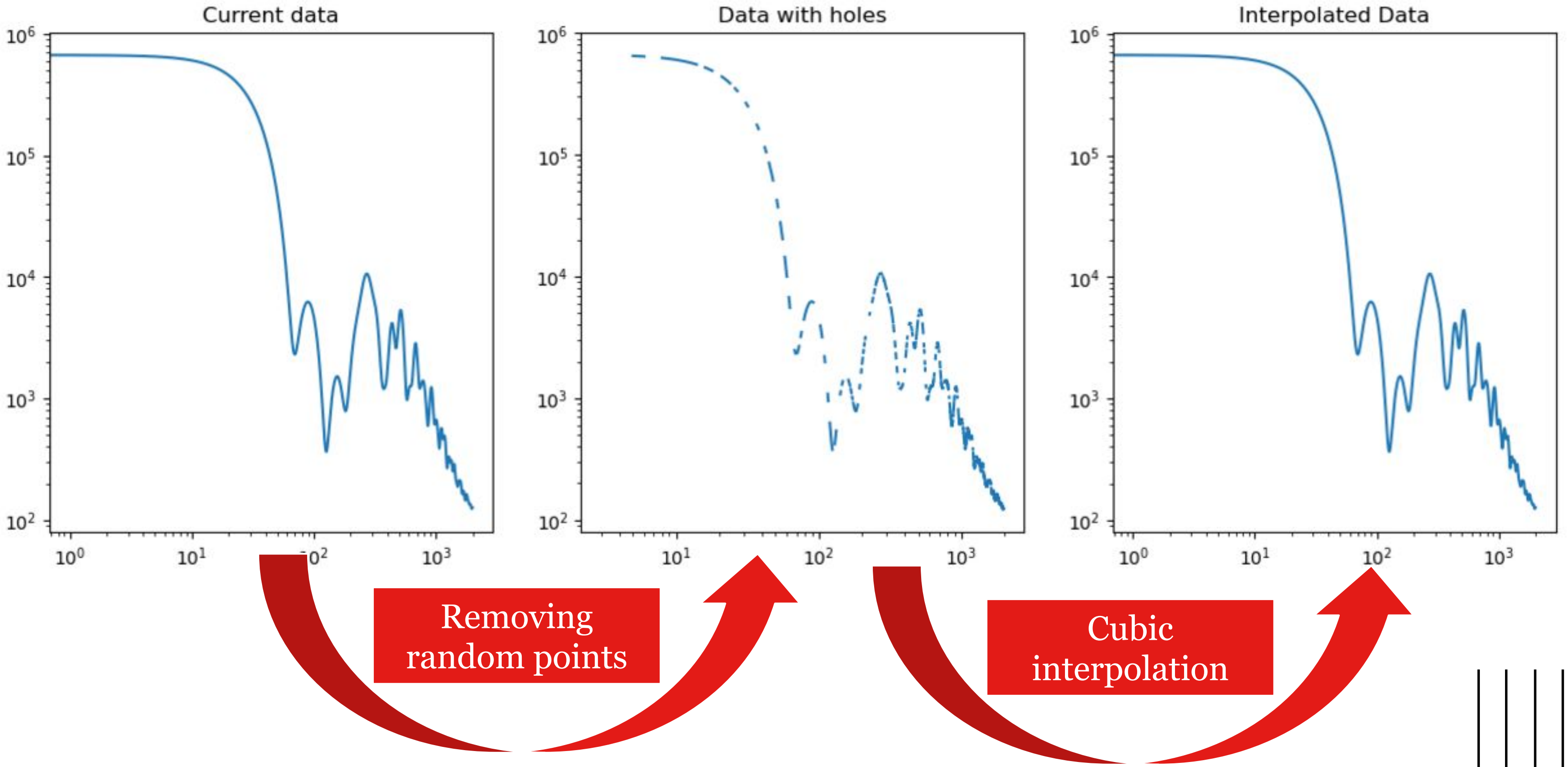


Impact of noise on prediction

Logistic Regression model: Training and predictions on noisy data

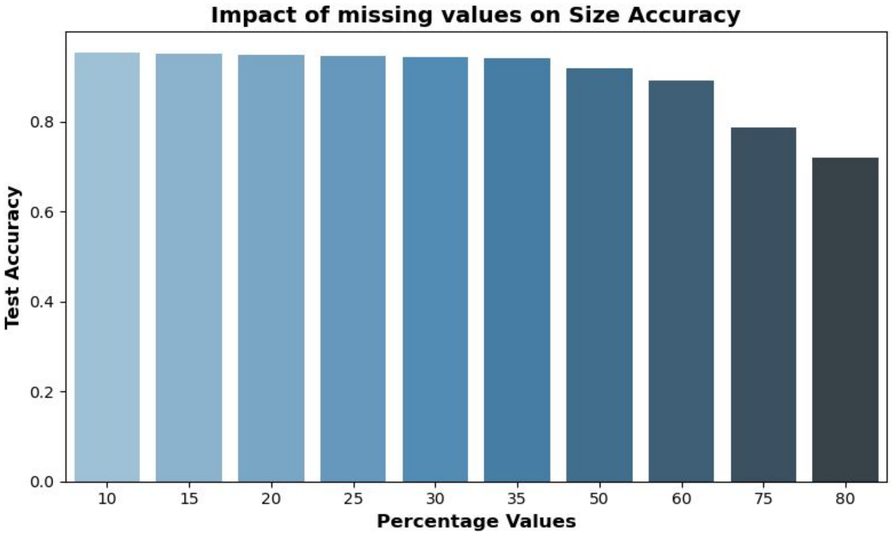


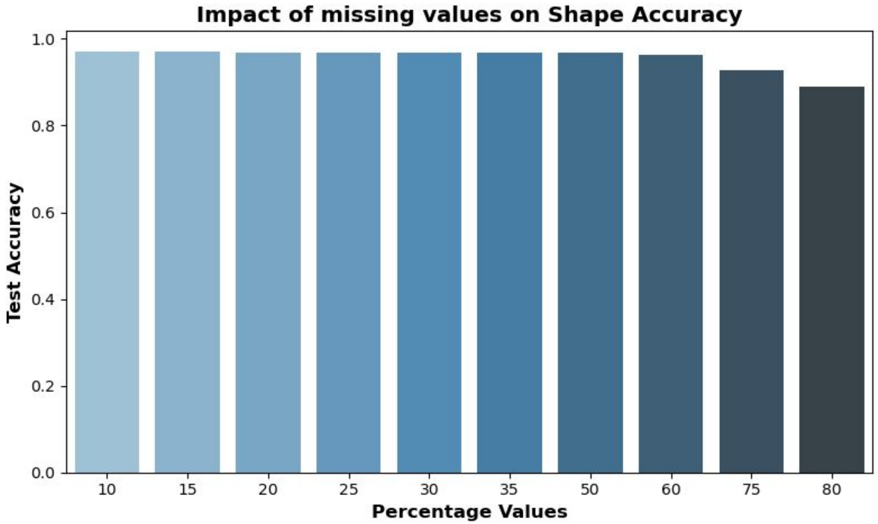
I.V.2 Adding Missing values



Impact of missing value + interpolation

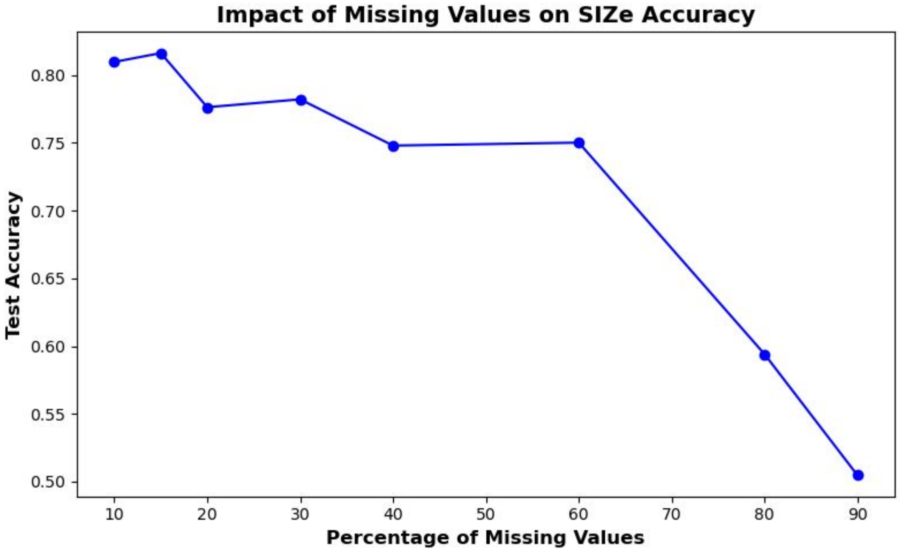
CNN model : Simple predictions on interpolated data



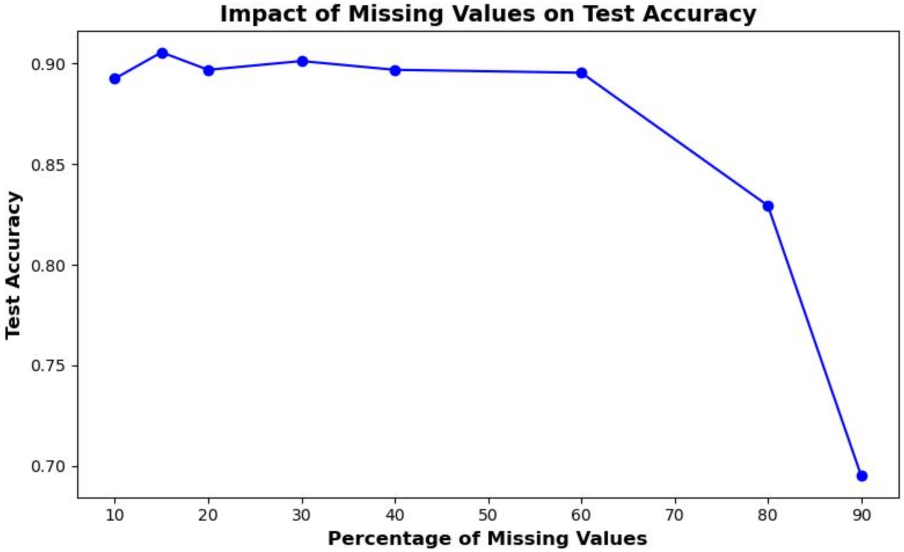


Impact of missing value + interpolation

CNN model : Training and prediction on interpolated data



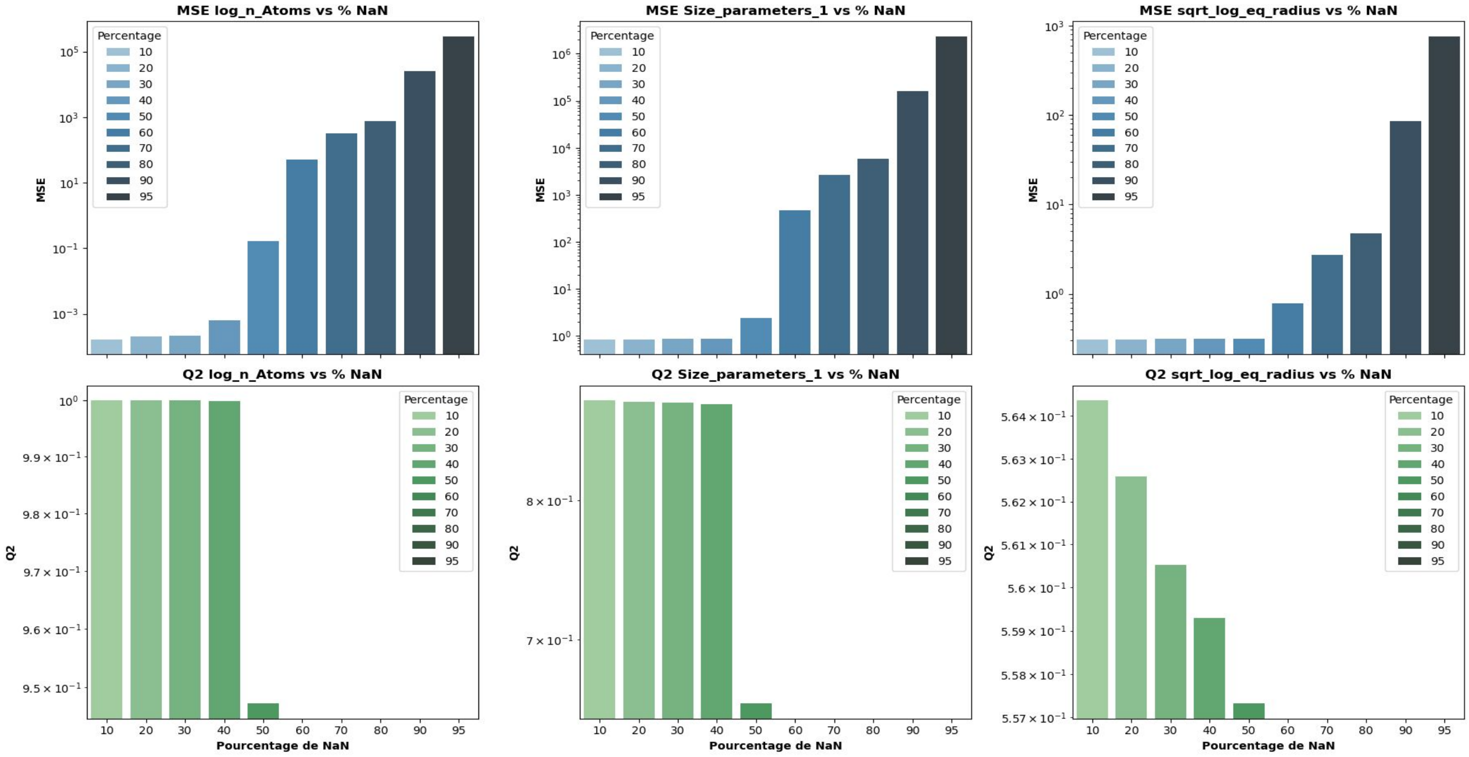
SIZE



SHAPE

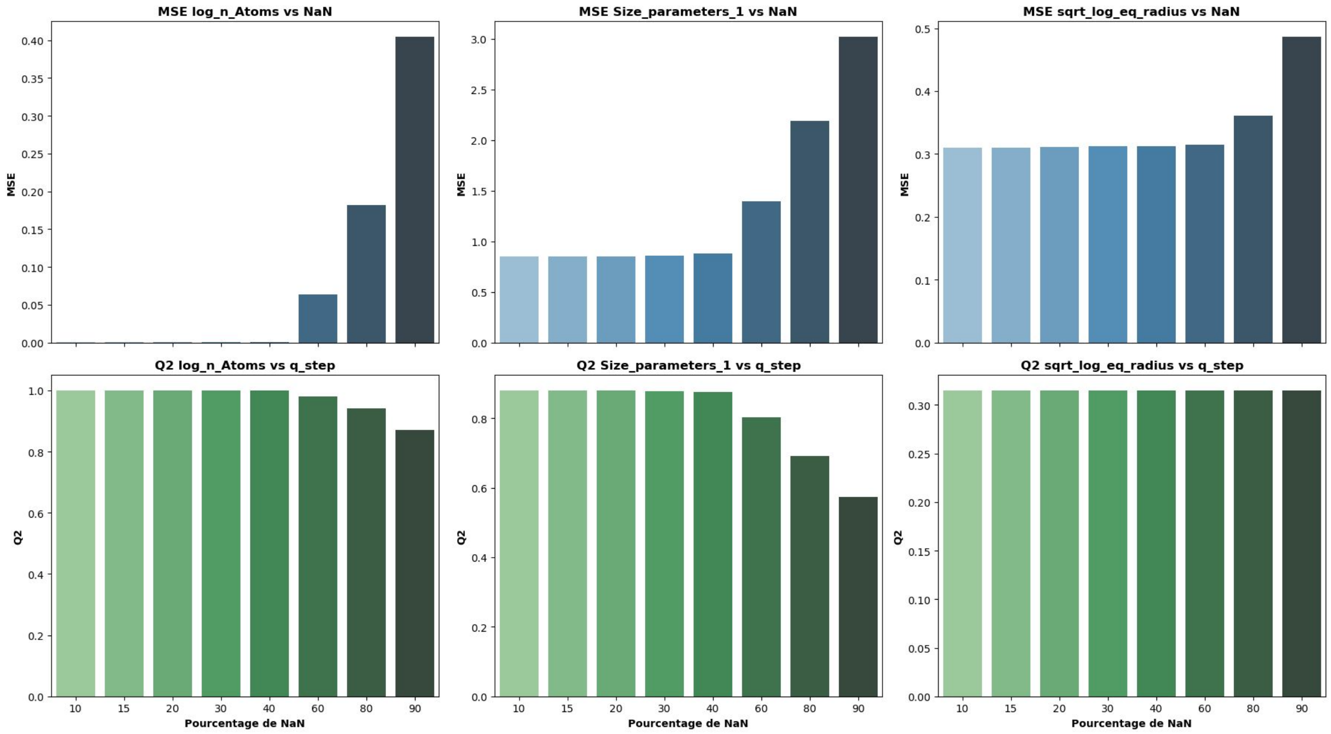
Impact of missing value + interpolation

Ridge Regression model: Simple predictions on interpolated data



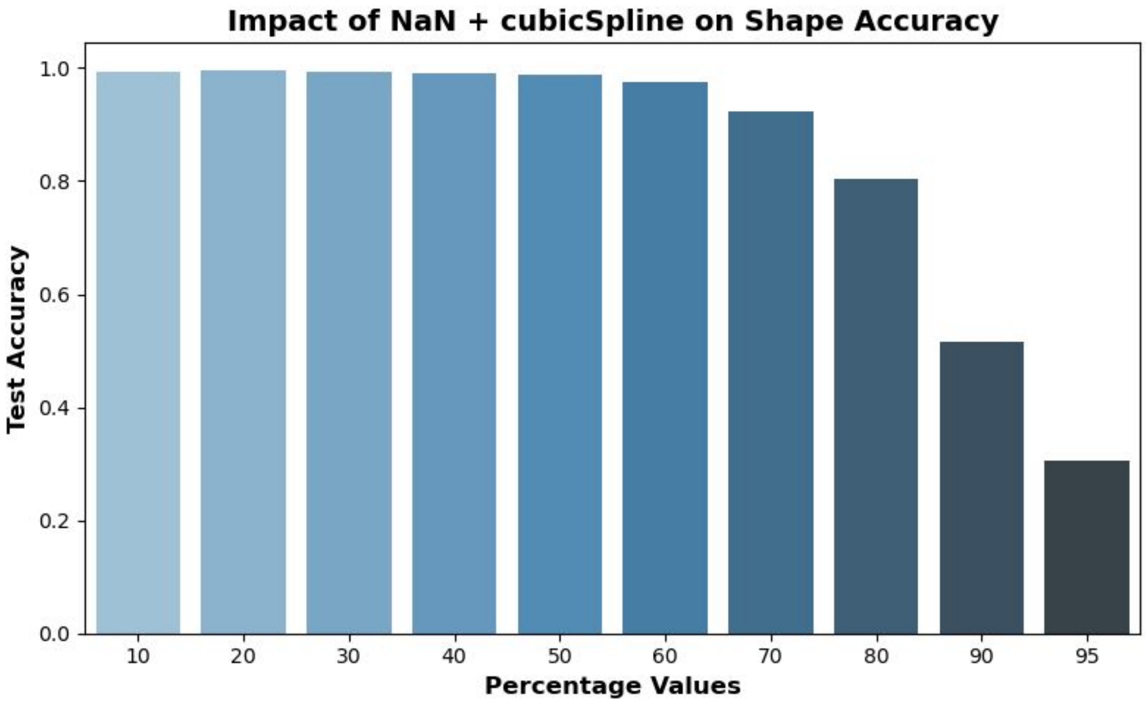
Impact of missing value + interpolation

Ridge Regression model: Training and predictions on interpolated data



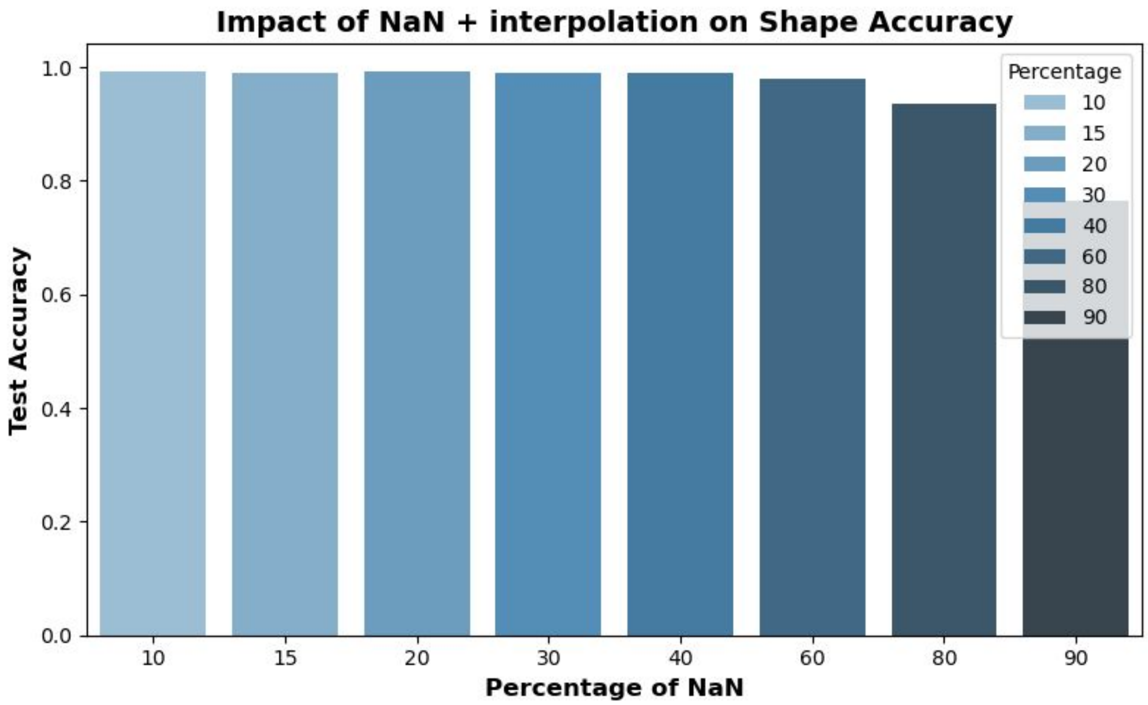
Impact of missing value + interpolation

Logistic regression model : Simple predictions on interpolated data

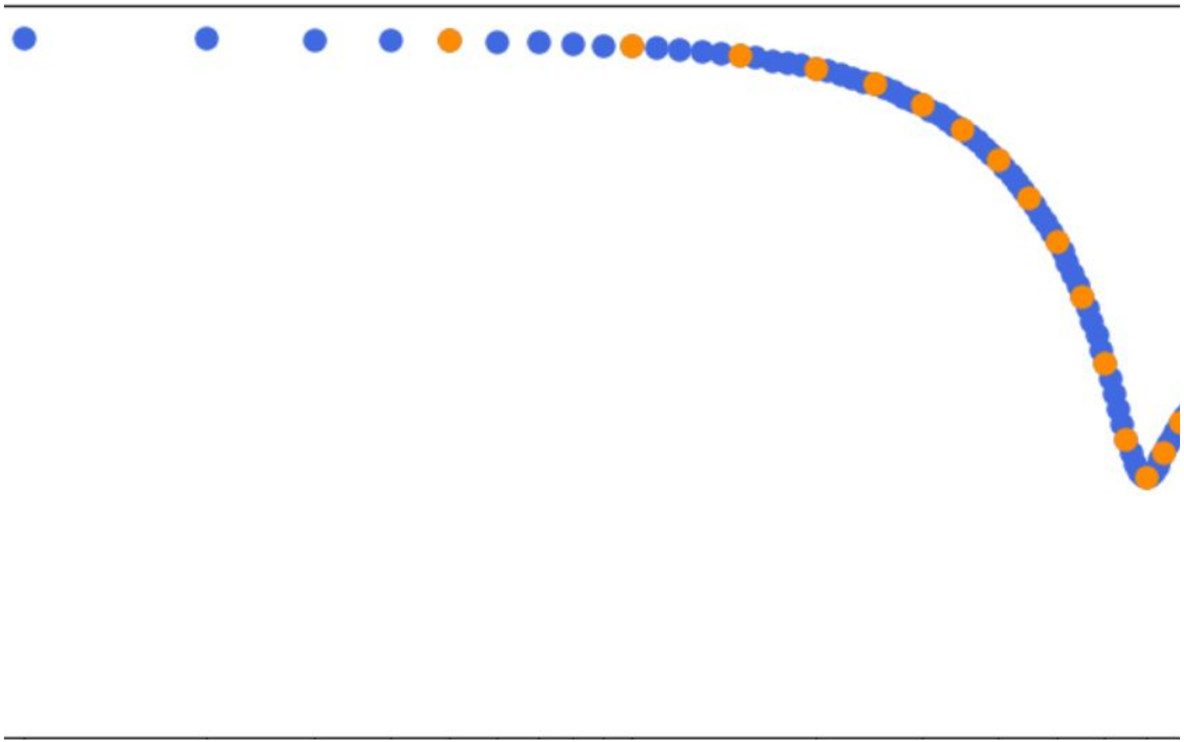


Impact of missing value + interpolation

Logistic regression model : Training and predictions on interpolated data



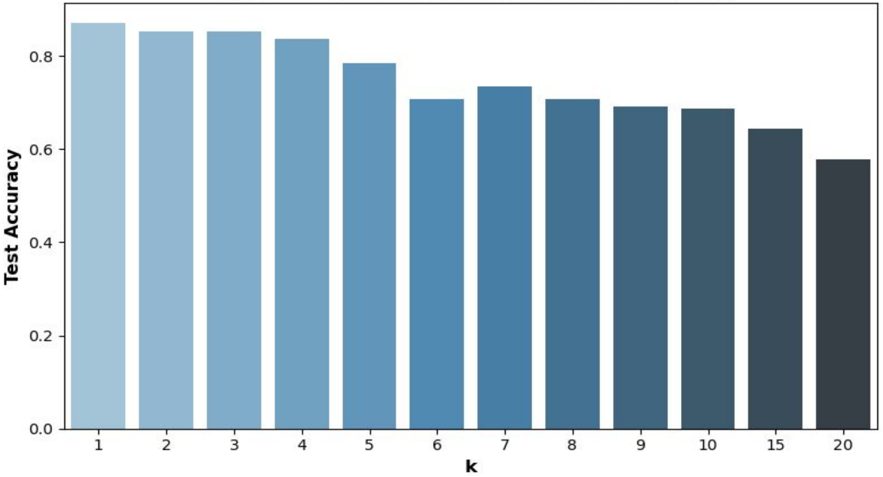
I.V.3 Step increase

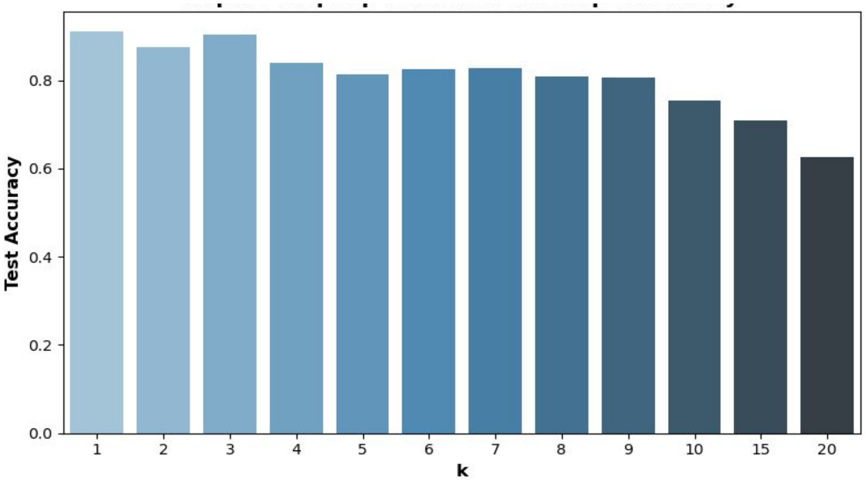


- original values
- Selected values

I.V.3 Step increase: Results

CNN1D model : Training and prediction on edited data





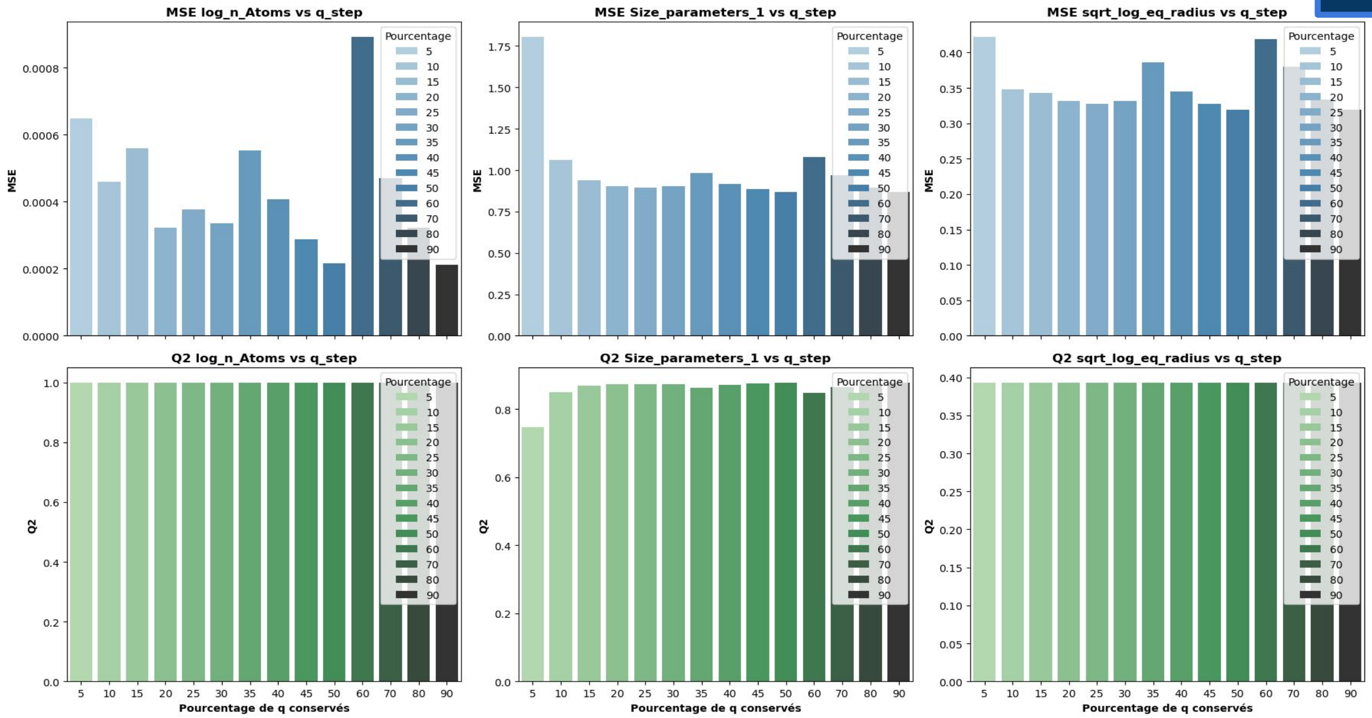
Impact of step increase on Size prediction

Impact of step increase on Shape prediction

I.V.3 Step increase: Results

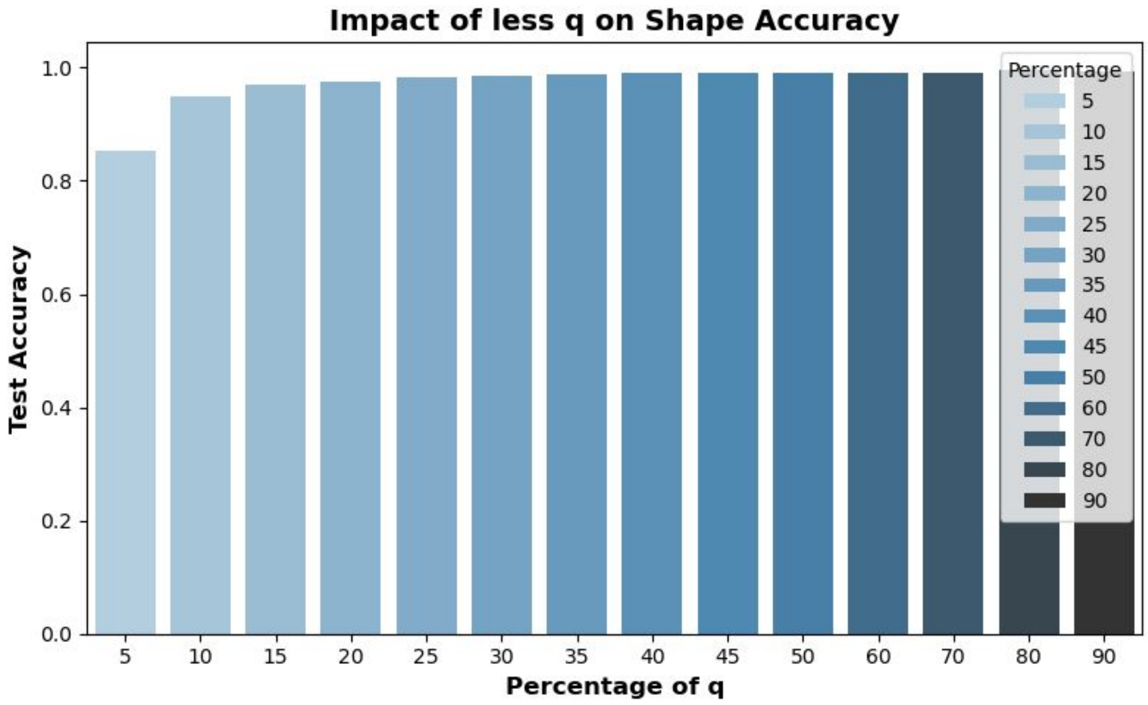
Ridge model: Training and prediction on edited data

Impact of step increase on size prediction



I.V.3 Step increase: Results

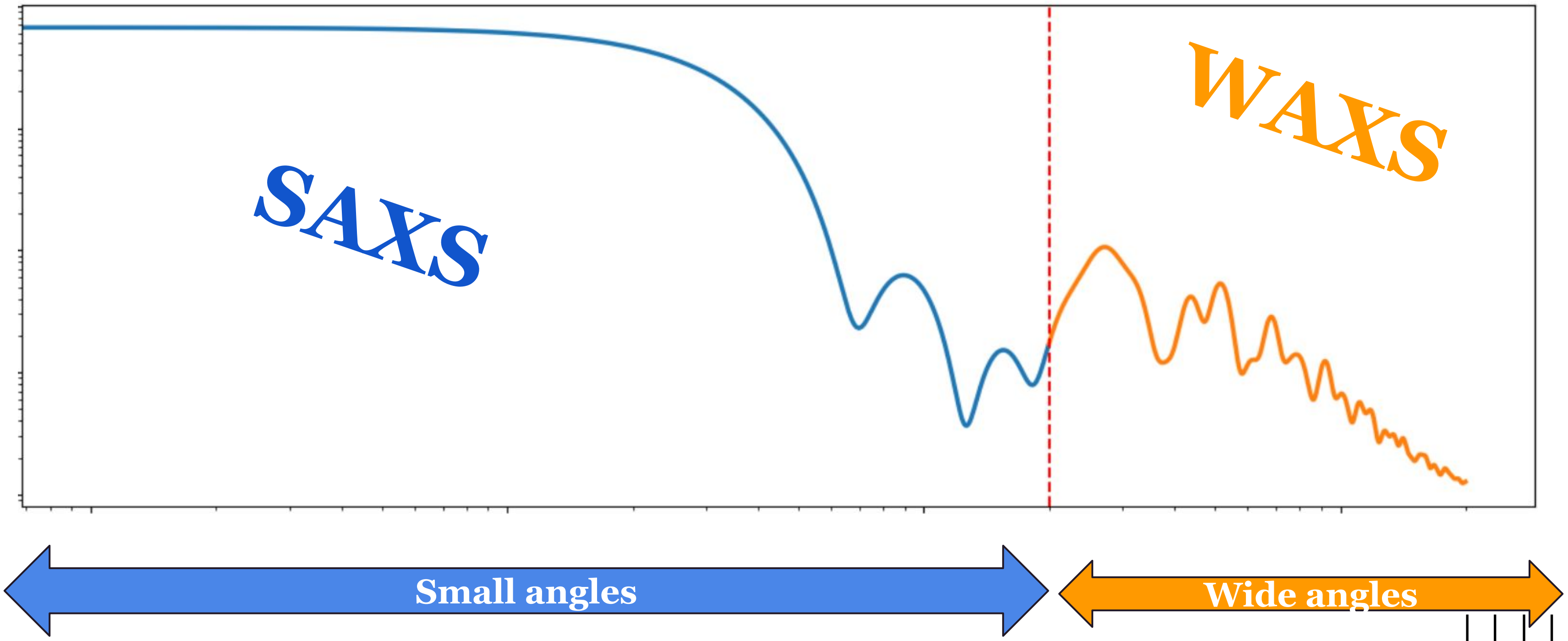
Logistic model: Training and prediction on edited data



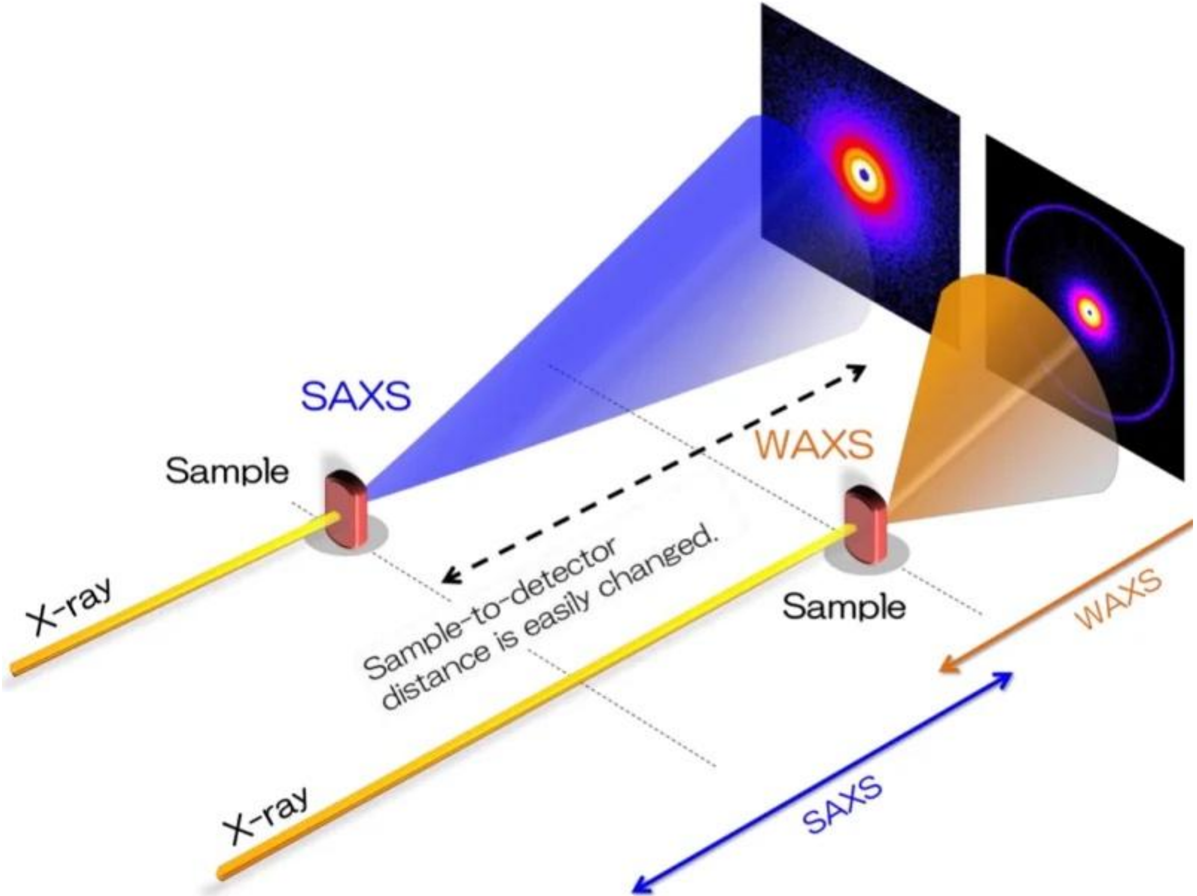
Impact of step increase on shape prediction

V. USING Only SAXS or WAXS

V. USING Only SAXS or WAXS

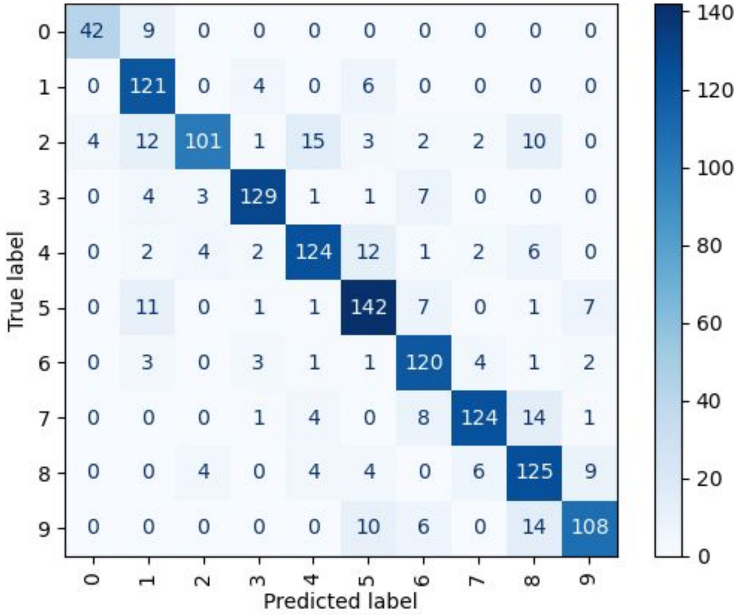


V. USING Only SAXS or WAXS



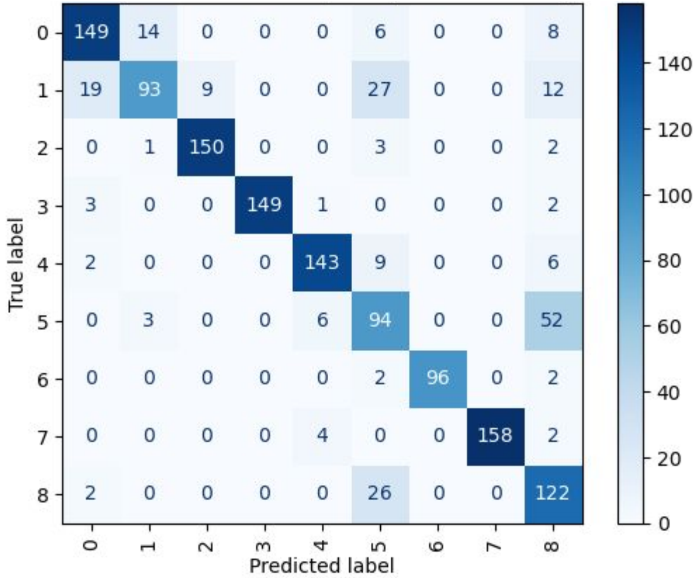
V.1 Prediction with WAXS

CNN1D : Size prediction



Accuracy=82%

CNN1D : Shape prediction



Accuracy=84%

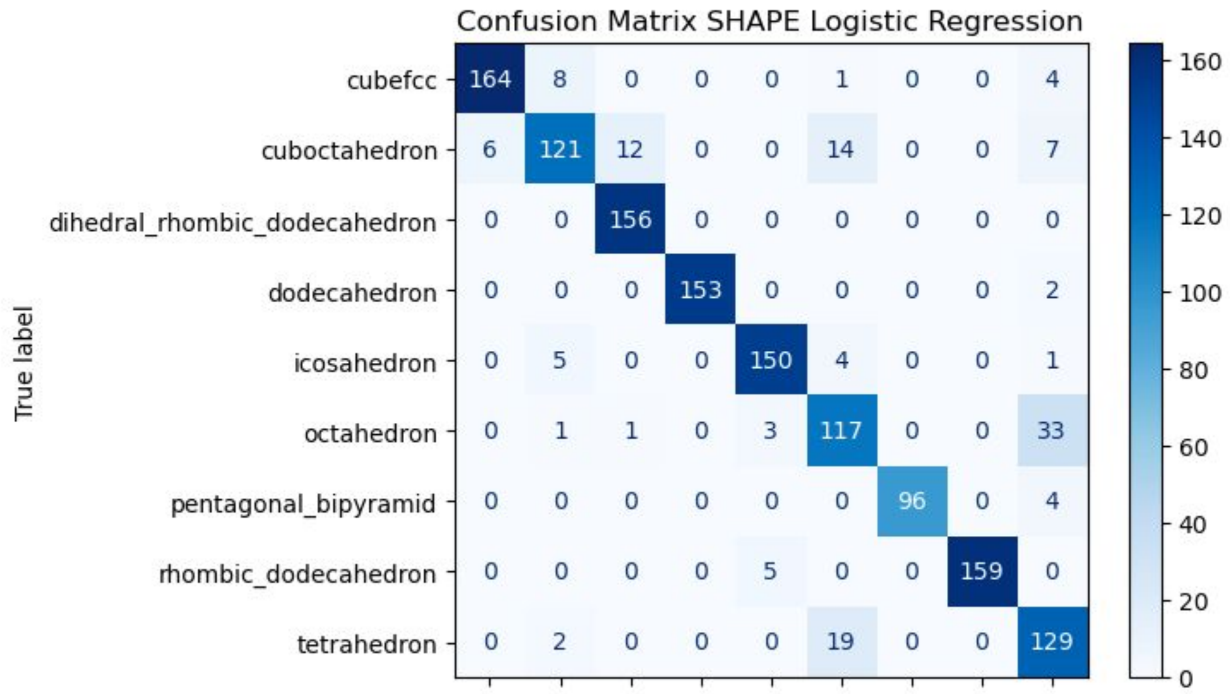
V.1 Prediction with WAXS

Ridge model : Size prediction

Features	MSE	Q2
log_n_Atoms	0.1152	0.9629
Size_parameter_1	2.0095	0.7158
sqrt_log_eq_radius	0.3200	0.5496

V.1 Prediction with WAXS

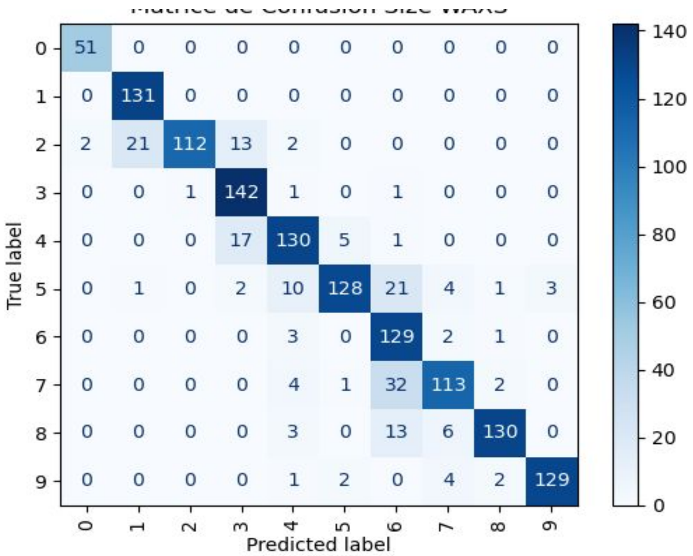
Logistic model : Shape prediction



Accuracy = 90%

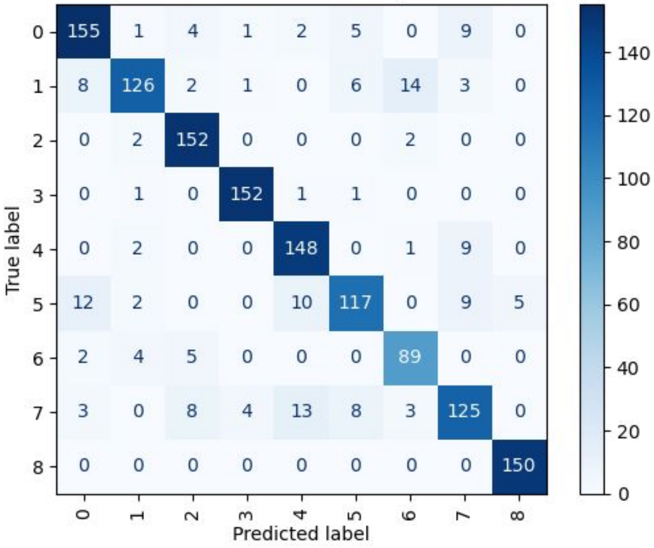
V.2 Prediction with SAXS

CNN1D : Size prediction



Accuracy=86%

CNN1D : Shape prediction



Accuracy=88%

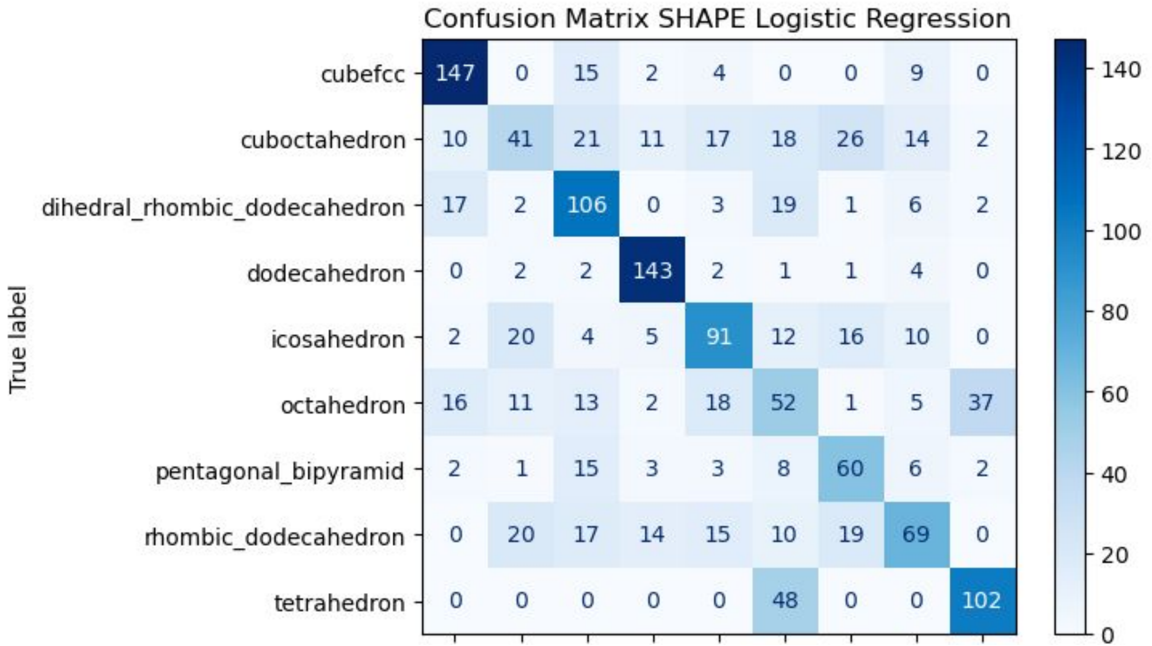
V.2 Prediction with SAXS

Ridge model : Size prediction

Features	MSE	Q2
log_n_Atoms	0.0300	0.9904
Size_parameter_1	2.1451	0.6966
sqrt_log_eq_radius	0.6271	0.1170

V.2 Prediction with SAXS

Logistic model : Shape prediction

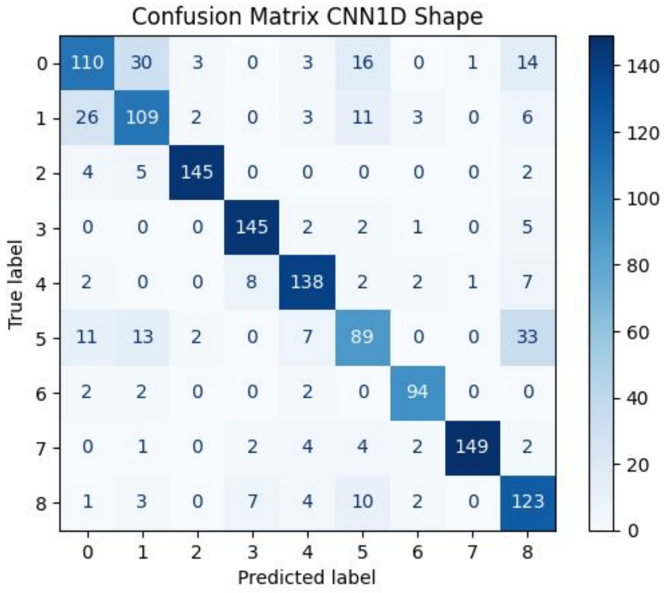


Accuracy = 59%

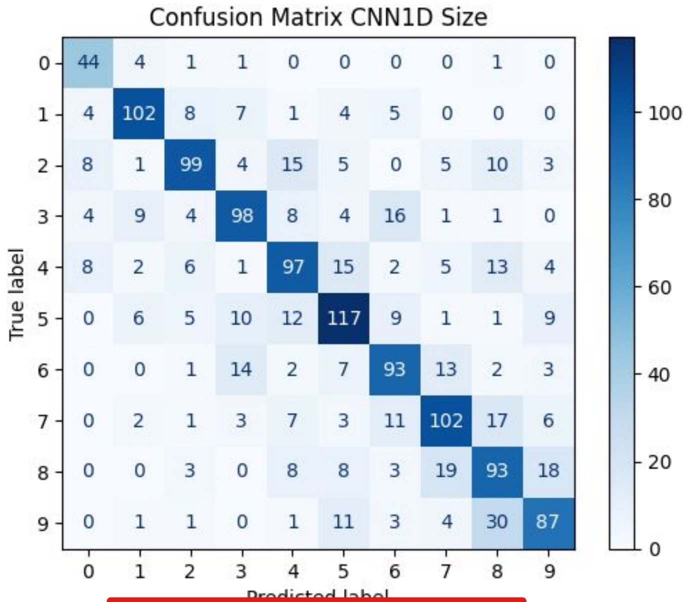
VI.Impact of Modifying Signal Bounds

CNN model

Training and prediction



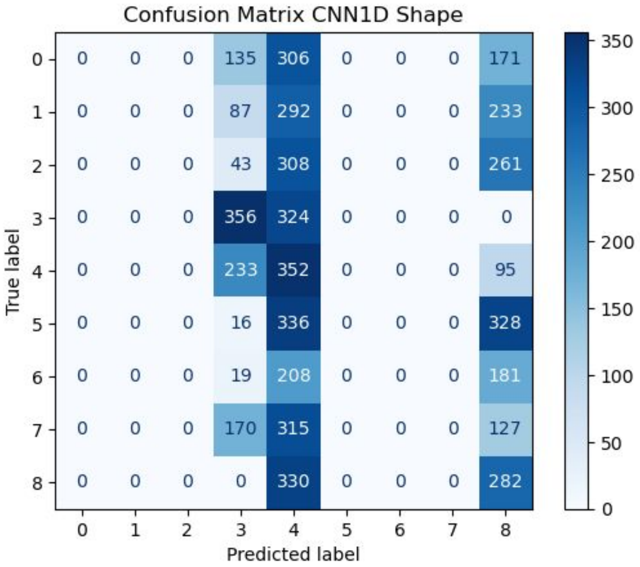
Shape Accuracy = 80%

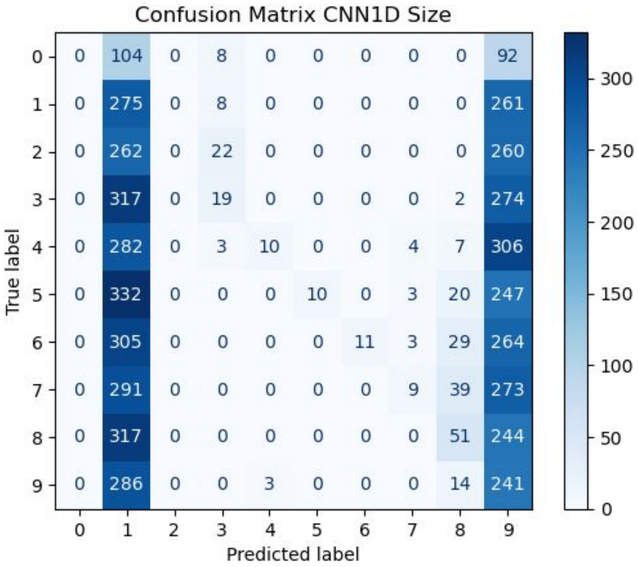


Size Accuracy = 66%

CNN model

Simple predictions

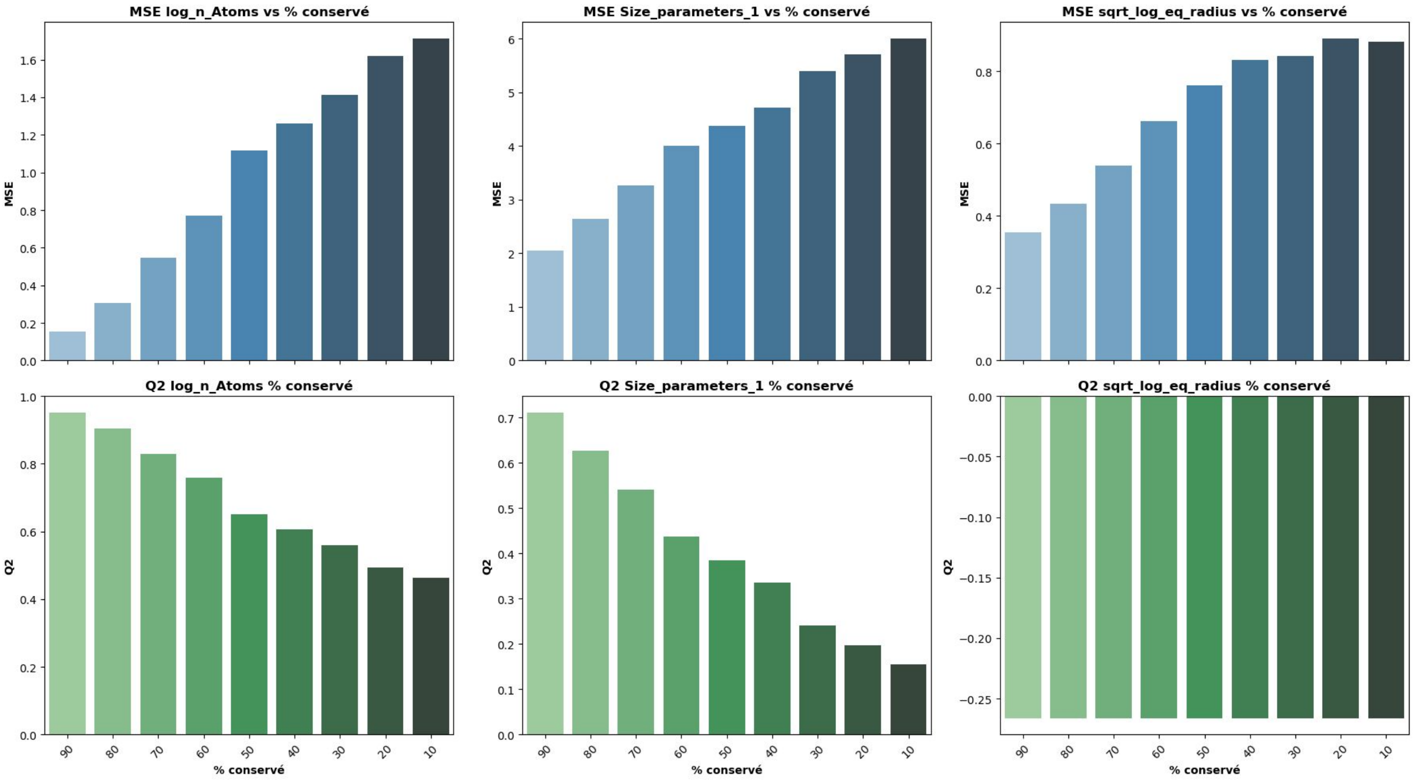




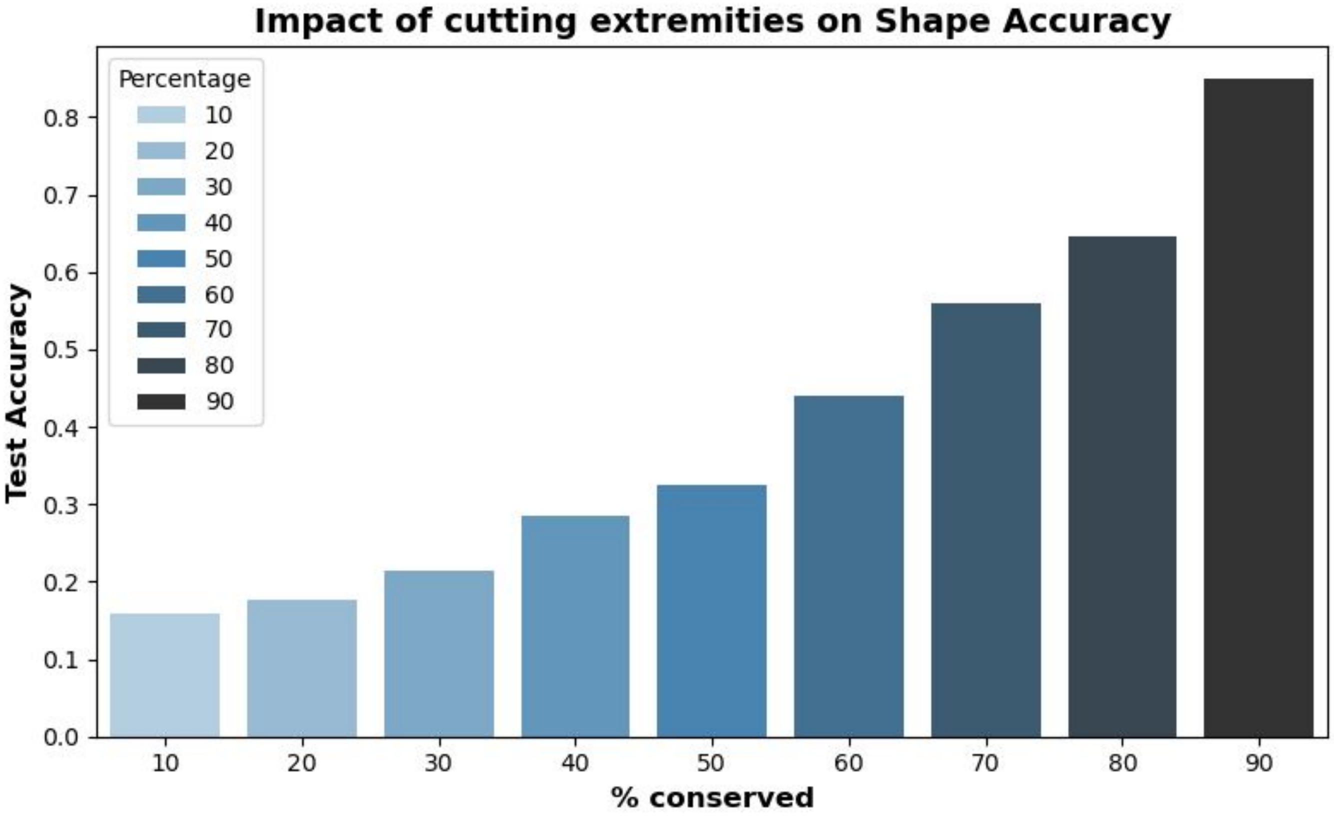
Shape Accuracy = 17%

Size Accuracy = 11%

Ridge model


MSE log_n_Atoms vs % conservé
MSE Size_parameters_1 vs % conservé
MSE sqrt_log_eq_radius vs % conservé
Q2 log_n_Atoms % conservé
Q2 Size_parameters_1 % conservé
Q2 sqrt_log_eq_radius % conservé
MSE
Q2
% conservé
90
80
70
60
50
40
30
20
10

Logistic model



CONCLUSION

- ★ **Work completed: Developing high-performance models using our simulated data**

  - ★ **Outlook:**
    - **working with more realistic data**
    - **Explainable Artificial Intelligence (XAI) methods to identify the specific q-range regions contributing to size and shape predictions**