



Basic Study for Machine Learning Application to Collider Experiments

Character Classification Using Dimensionality Reduction

Osaka-City Univ. Chikayo Kato

1. Introduction
2. Character classification
3. Summary

1-1. Introduction

High Energy Physics Collider experiment

Challenges

- Huge amount of Experimental data
e.g. **~30 Peta Byte / year** in Belle II experiment
- Huge amount of Detector sensors
number of sensors **$O(10^3) \sim O(10^4)$ or more**
→Big data size

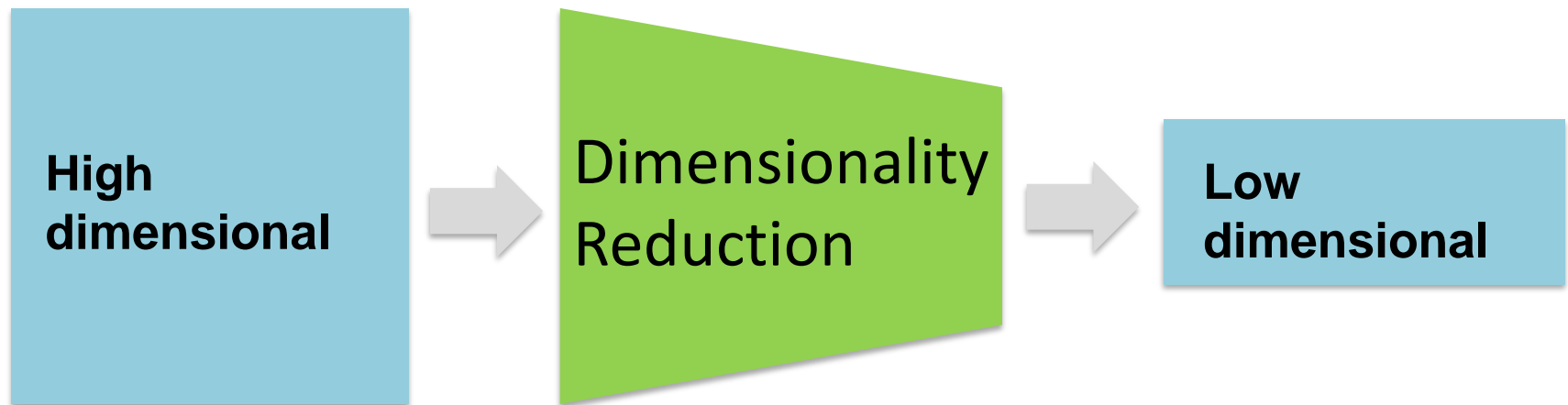
Novel technology to treat the big data is required.



**We have developed Dimensionality Reduction
with machine learning**

1-1. Introduction

- Dimensionality Reduction is the transformation of data from a high-dimensional space into a low-dimensional space.



Dimensionality Reduction may be possible to reduce the data size without degrading the data.

1-2 Character Classification Using Dimensionality Reduction

This time, we have studied the Character Classification using Dimensionality Reduction as a basic study for machine learning application.

- For the study we use MNIST data (image data of handwritten numbers) (784parameters)





Image data				
Label data	 5	 0	 4	 1

Image data of 0~9 Total:70000

Train data :60000
Test data :10000

1-2 Character Classification Using Dimensionality Reduction

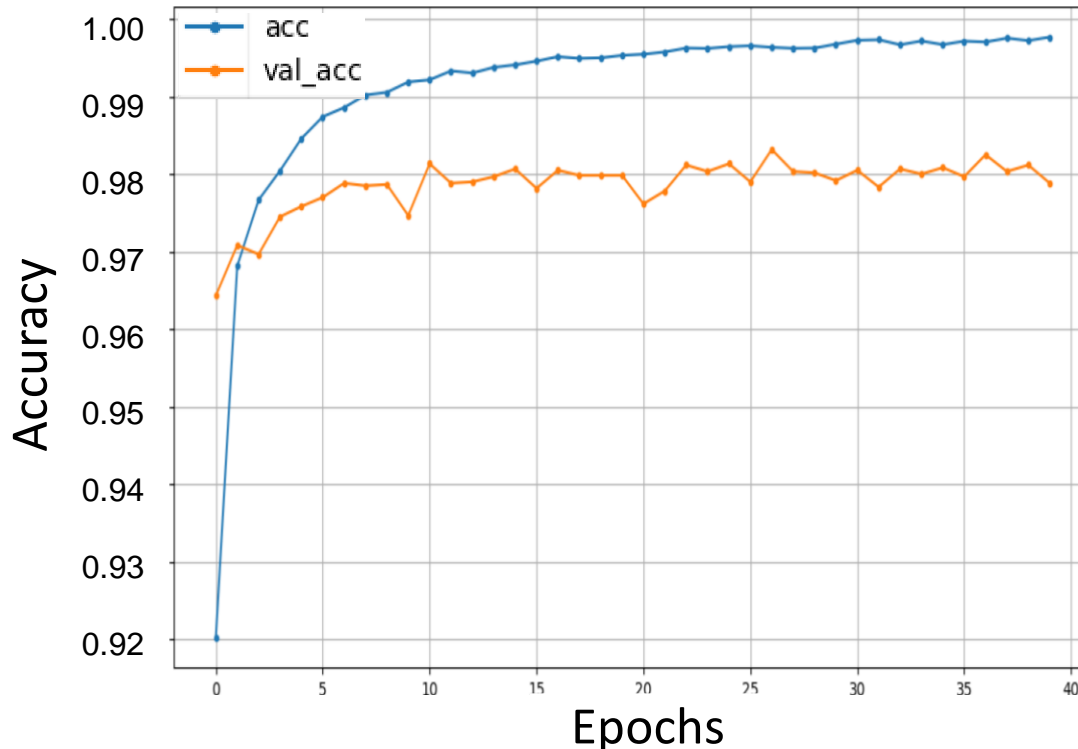


- Dimensionality reduction methods

- ① VAE (Variational Auto Encoder)
- ② Random down sampling by hand

In this study we have applied dimensionality reduction to the MNIST data and evaluate the character classification performance.

2-1. Character classification without dimensionality reduction



Middle layer : 2
Batch size : 128

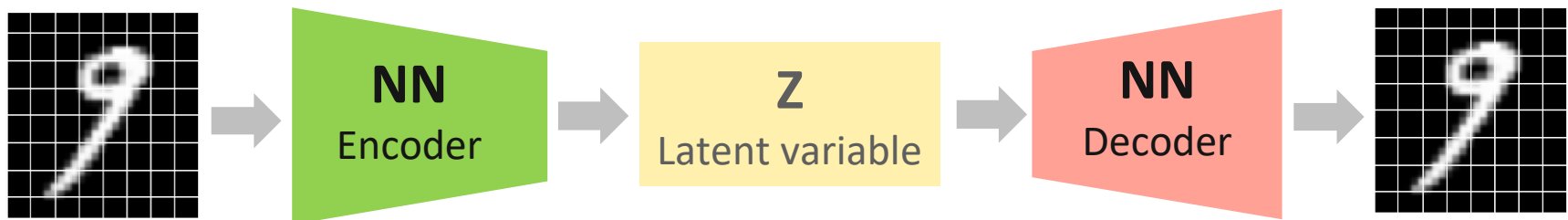
- **Optimizer** : RMSprop
- **Loss function** : categorical_crossentropy
- **Activation function** : ReLU

Without dimensionality reduction accuracy : ~98%

2-2. Character classification with dimensionality reduction using VAE

- VAE(Variational Auto Encoder)

VAE = Encoder + Decoder



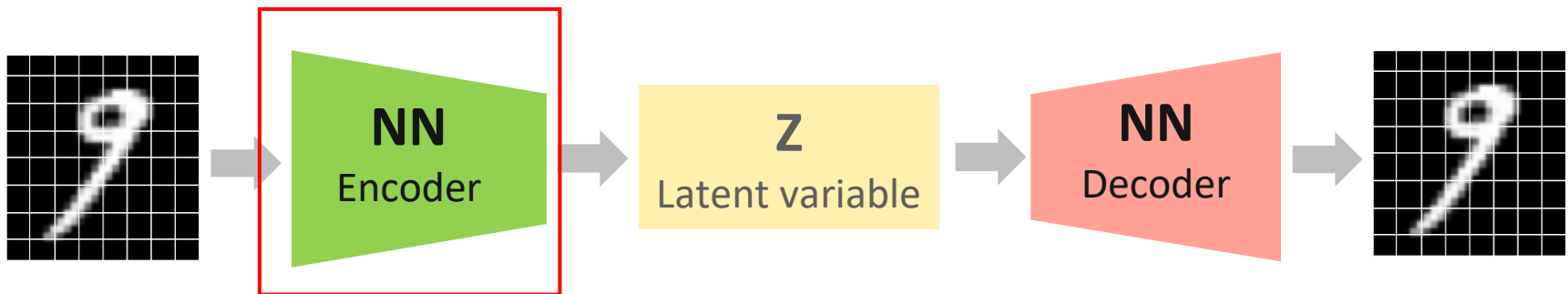
▪ The encoder is a NN that obtains the latent variables that characterize the input data.

▪ The decoder is a NN that reconfigures the original image from latent variables.

2-2. Character classification with dimensionality reduction using VAE

- VAE (Variational Auto Encoder)

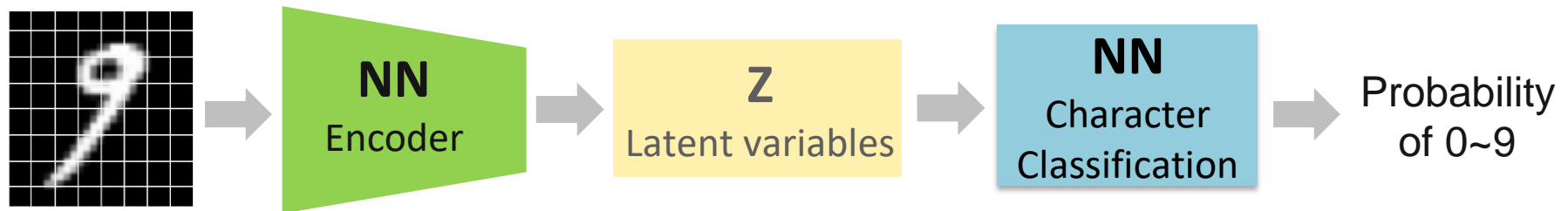
VAE = Encoder + Decoder



In this case, only the encoder was used for dimensionality reduction.

2-2. Character classification with dimensionality reduction using VAE

- Character Classification by learning latent variables

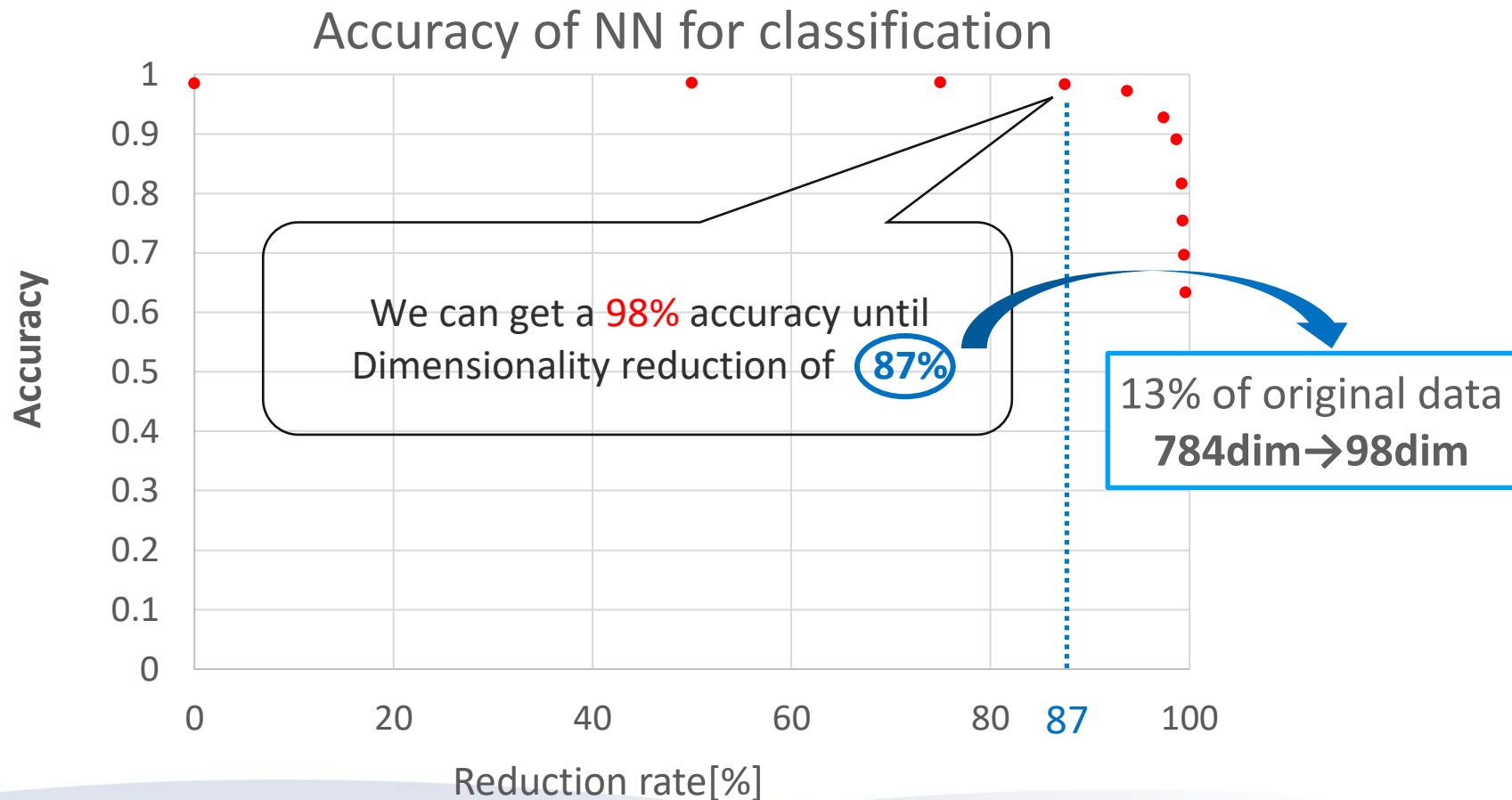


- Output the latent variables in dimensions 2, 3, 4, 5, 6, 10, 20, 49, 98, 196, and 392, and these are trained on the NN.

→ We evaluate that character classification performance after applying the dimensionality reduction with VAE.

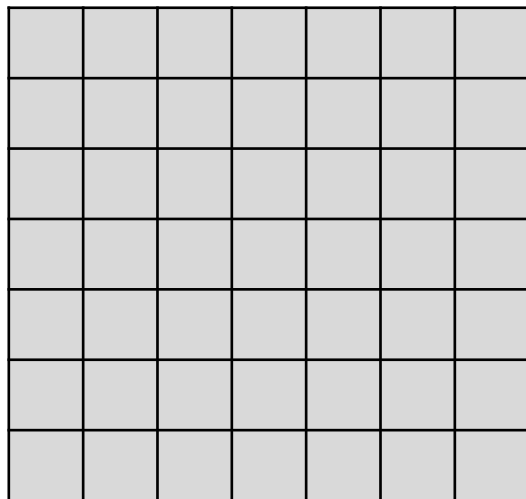
2-2. Character classification with dimensionality reduction using VAE

- Character Classification by learning latent variables



2-3. Character classification with dimensionality reduction using random down sampling

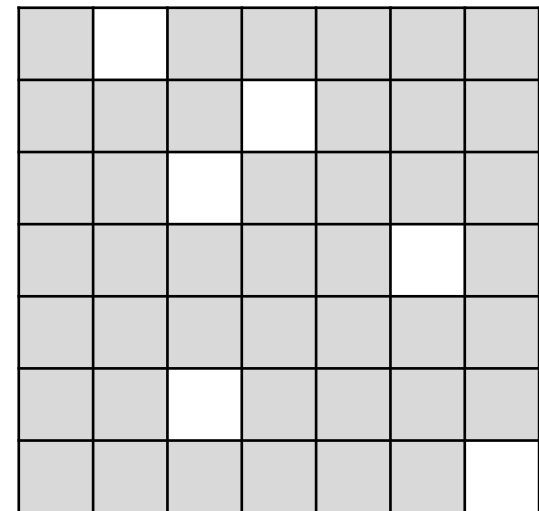
- Random down sampling of data



Input image



784parameters



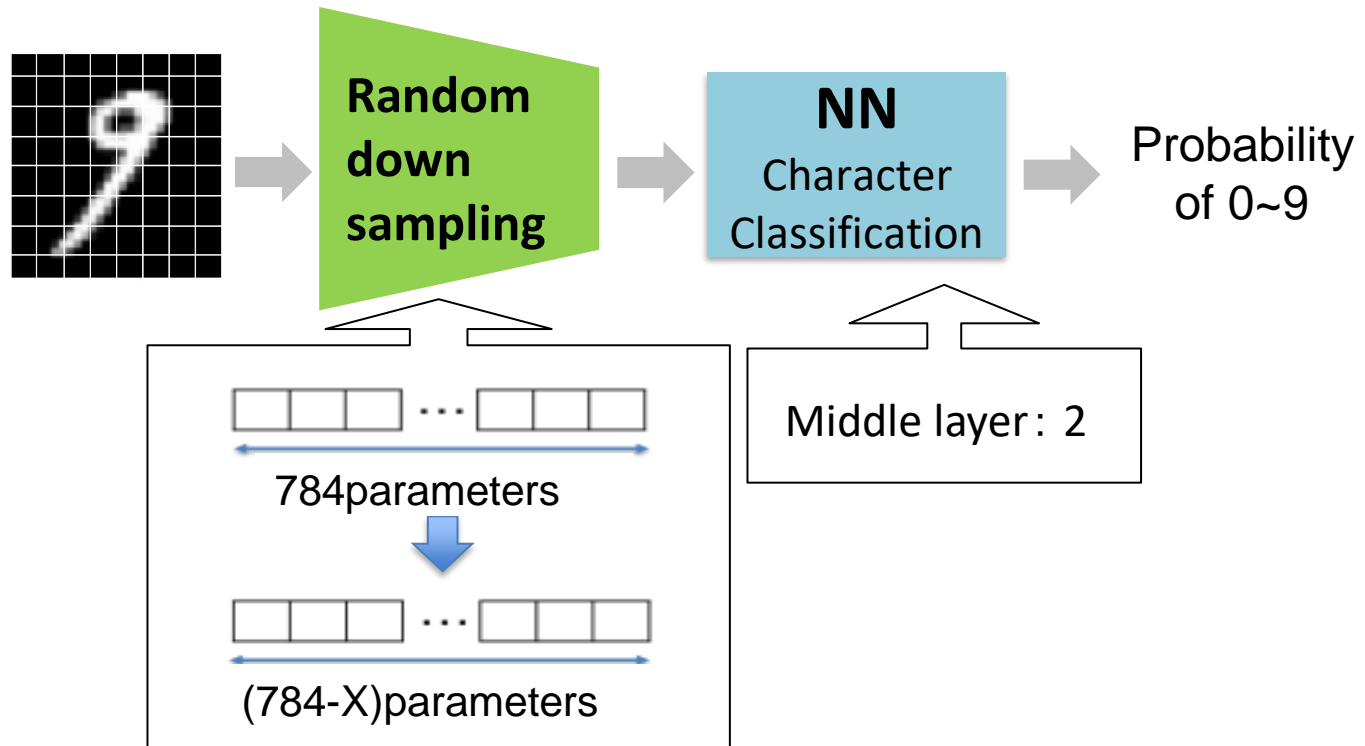
Erase information from random locations.



(784-X)parameters

2-3. Character classification with dimensionality reduction using random down sampling

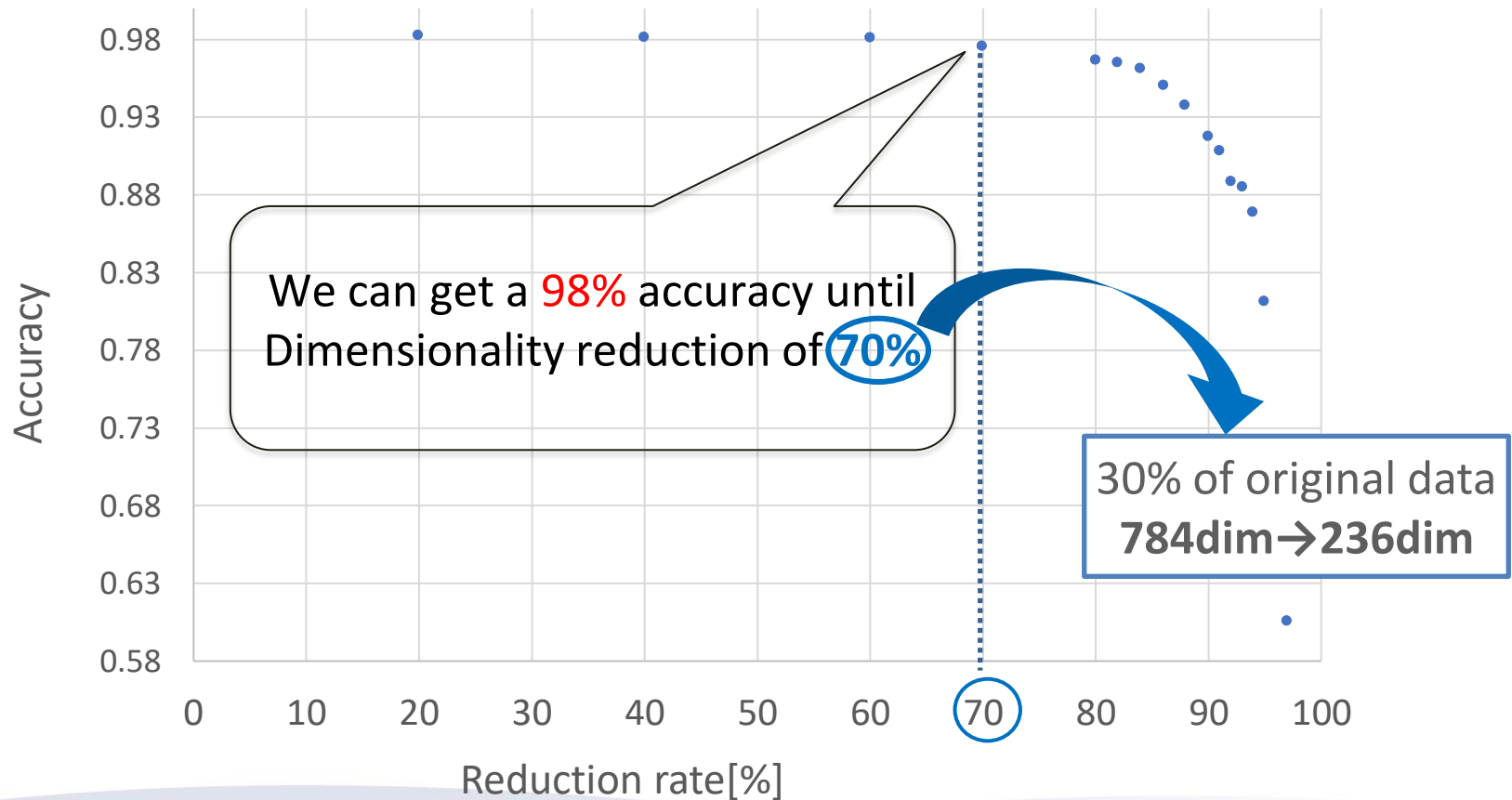
- Random down sampling by hand.



→ We evaluate that character classification performance after applying the dimensionality reduction with Random down sampling by hand.

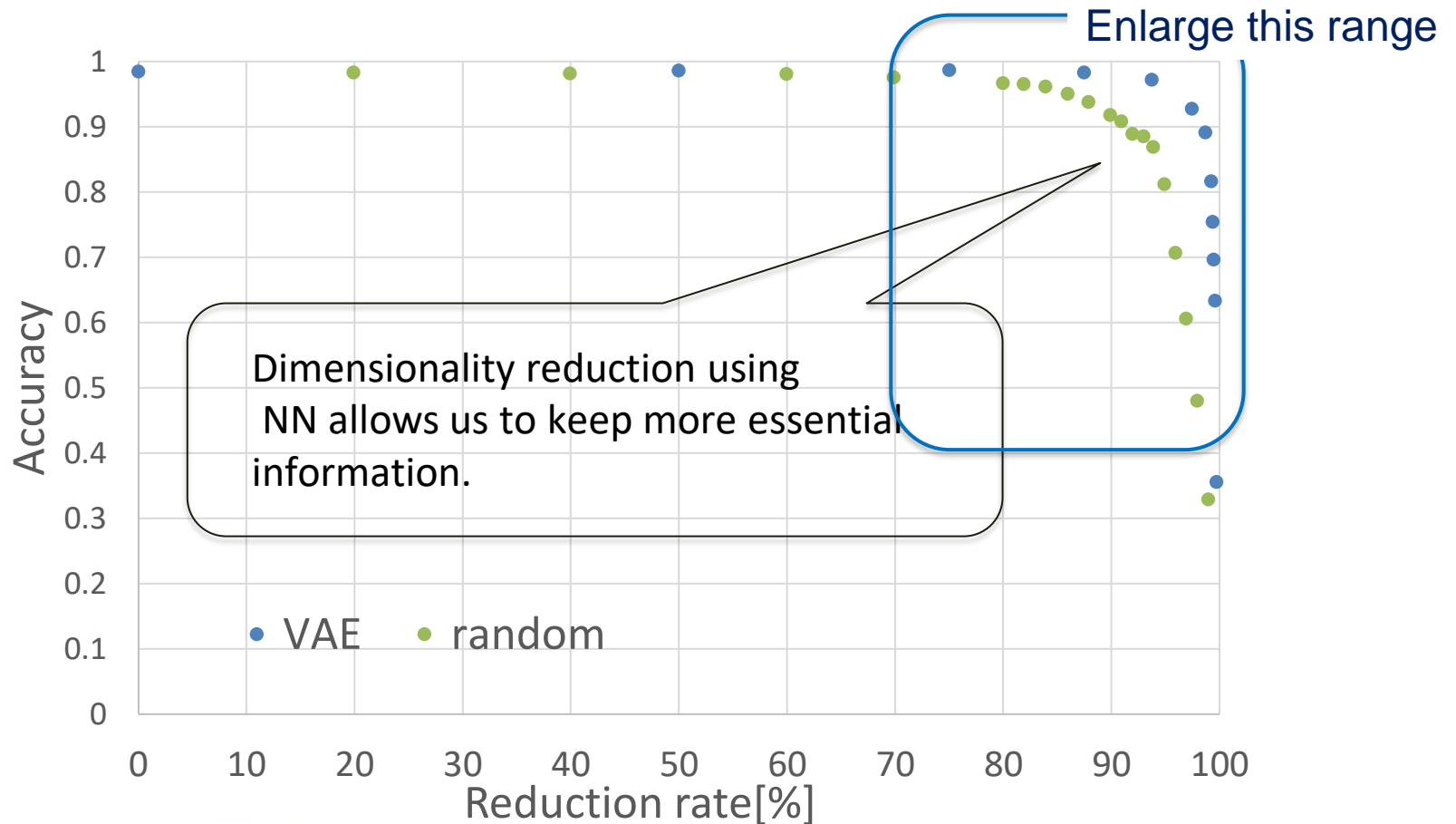
2-3. Character classification with dimensionality reduction using random down sampling

- Result : (Character classification using random down sampling)



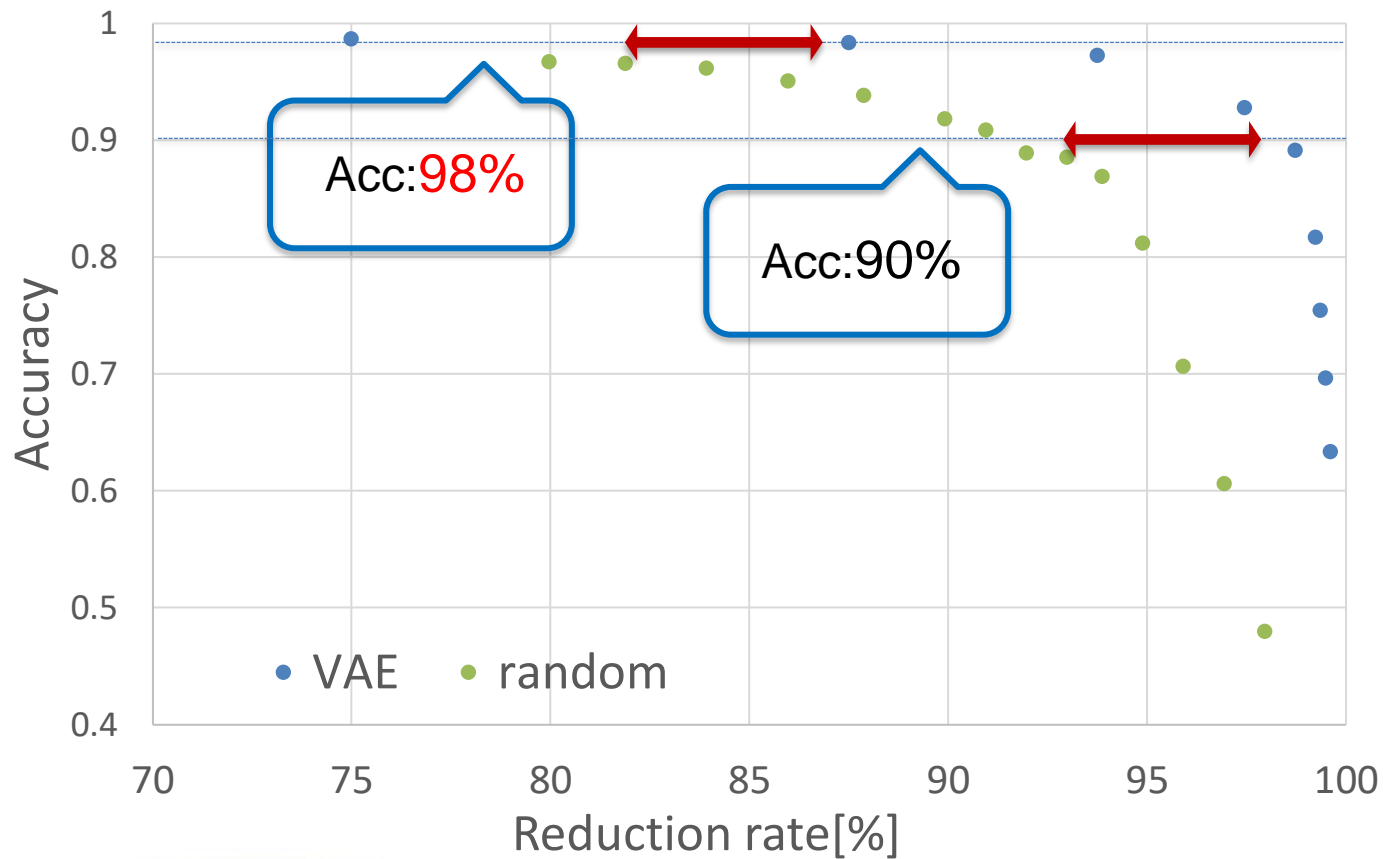
2-4. VAE vs Random down sampling by hand

- Result : VAE vs Random down sampling by hand



2-4. VAE vs Random down sampling by hand

- Result : VAE vs Random down sampling by hand



2-4. VAE vs Random down sampling by hand

- Result : VAE vs Random down sampling by hand

Accuracy	Reduction rate 3 times	
	Random down sampling	VAE
	30%	13%
98%	70%(Compression up to 1/3)	87%(Compression up to 1/8)
95%	85%(Compression up to 1/7)	95%(Compression up to 1/20)
90%	90%(Compression up to 1/10)	98%(Compression up to 1/50)

Without Dimensionality reduction accuracy : **~98%**

- ① VAE : Reduction rate **87%**
- ② Random down sampling by hand : Reduction rate **70%**
→ Keep classification performance.

2-4. VAE vs Random down sampling by hand

- Result : VAE vs Random down sampling by hand

Accuracy	Reduction rate 3 times	
	Random down sampling	VAE
	30%	13%
98%	70%(Compression up to 1/3)	87%(Compression up to 1/8)
95%	85%(Compression up to 1/7)	95%(Compression up to 1/20)
90%	90%(Compression up to 1/10)	98%(Compression up to 1/50)

Without Dimensionality reduction accuracy :

Down Sampling
with ML??

- ① VAE : Reduction rate **87%**
- ② Random down sampling by hand : Reduction rate **70%**
→ Keep classification performance.

3. Summary

- We have studied the Character Classification using Dimensionality Reduction as a basic study for machine learning application.

- Dimensionality reduction methods

- ① VAE (Variational Auto Encoder)

- ② Random down sampling by hand

- **Dimensionality Reduction**

- We can get a **98%** accuracy until Dimensionality reduction of

- ① VAE (encoder)

- Reduction rate: 87% (Compression up to 13%)

- ② Random down sampling by hand

- Reduction rate: 70% (Compression up to 30%)

- **Future Plan**

- To aim to achieve an accuracy closer to that of VAE, we will study the down sampling method using machine learning.

Thank you for listening.

