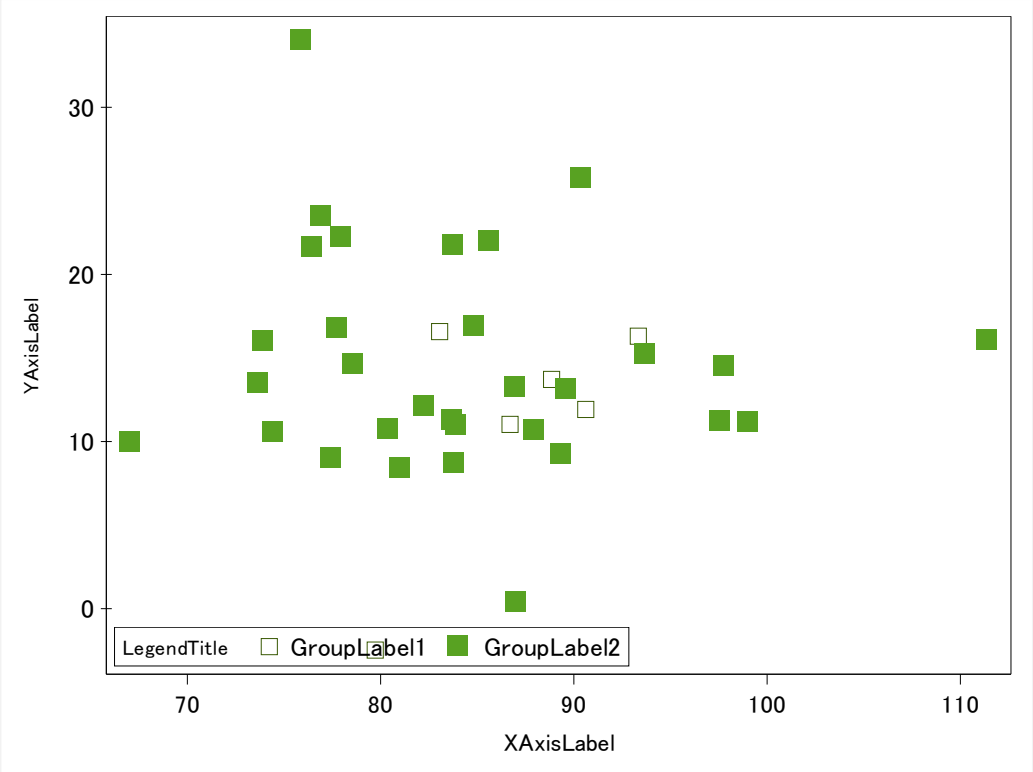
Automating Graph Verification with Neural Networks

Fukushima, Ryosuke

Graph Verification



Graphs in Clinical Trials

- Graphs serve as visual summaries of data
- The accuracy directly impacts decision-making

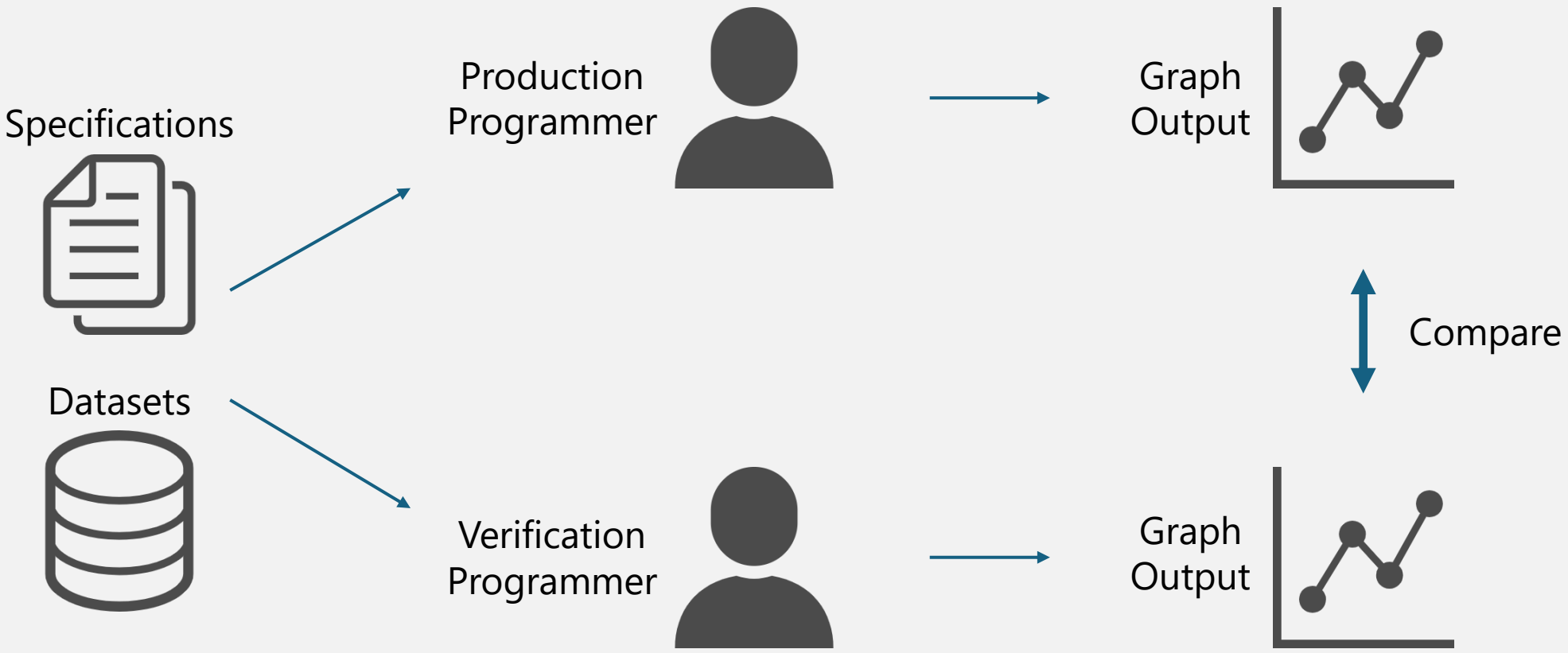
Possible Mistakes

- Wrong analysis population is shown
- Some plots are out of graph range
- Wrong labels for markers

To control the mistakes, double programing is used.

Verification with Double Programming

If independent outputs are matched, the graph would be reliable



Candidates for Comparison Method

(1). Visually Check:

Visually read out plotted points and compare

(2). Comparison of Intermediate Datasets:

Compare intermediate datasets just before creating graphs

(3). Image Comparison:

Compare graphs as images

(4). Image Recognition:

Recognize graph components and read out plotted points

Summary

Question 1:

Which method is better for verification, and what are the characteristics of each?

This presentation will explain and compare each method.
It will help you in choosing a method.

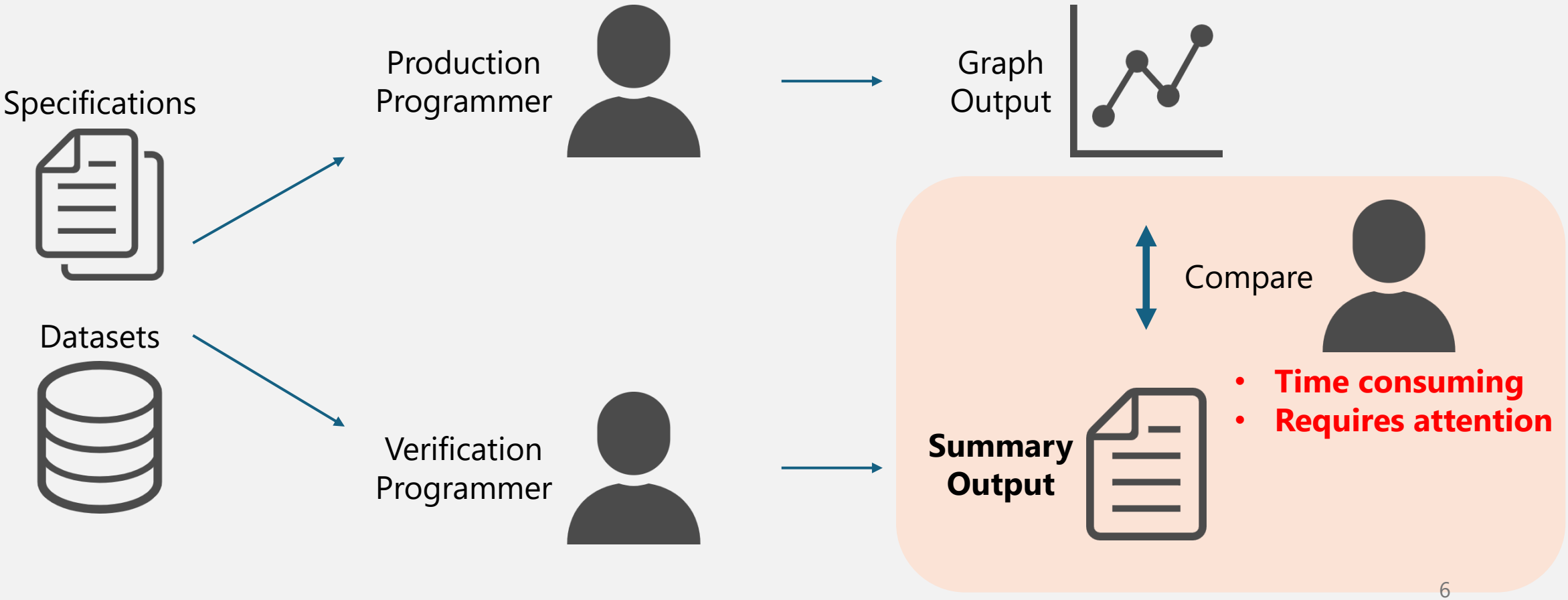
Question 2:

Image recognition is useful for verification, and available?

To my knowledge, no specialized tools.
This presentation introduce an image recognition method based on EMF image generation and neural network analysis

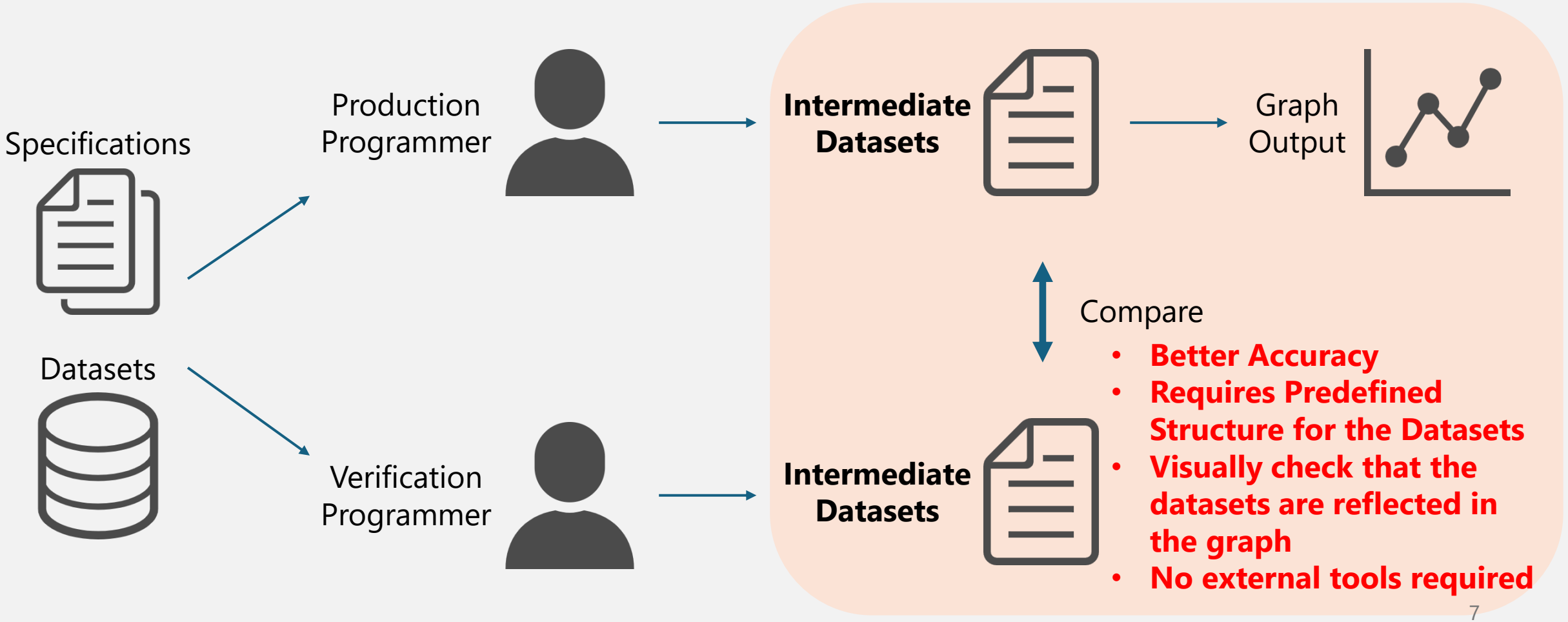
(1). Visually Check

Visually read out plotted points and compare



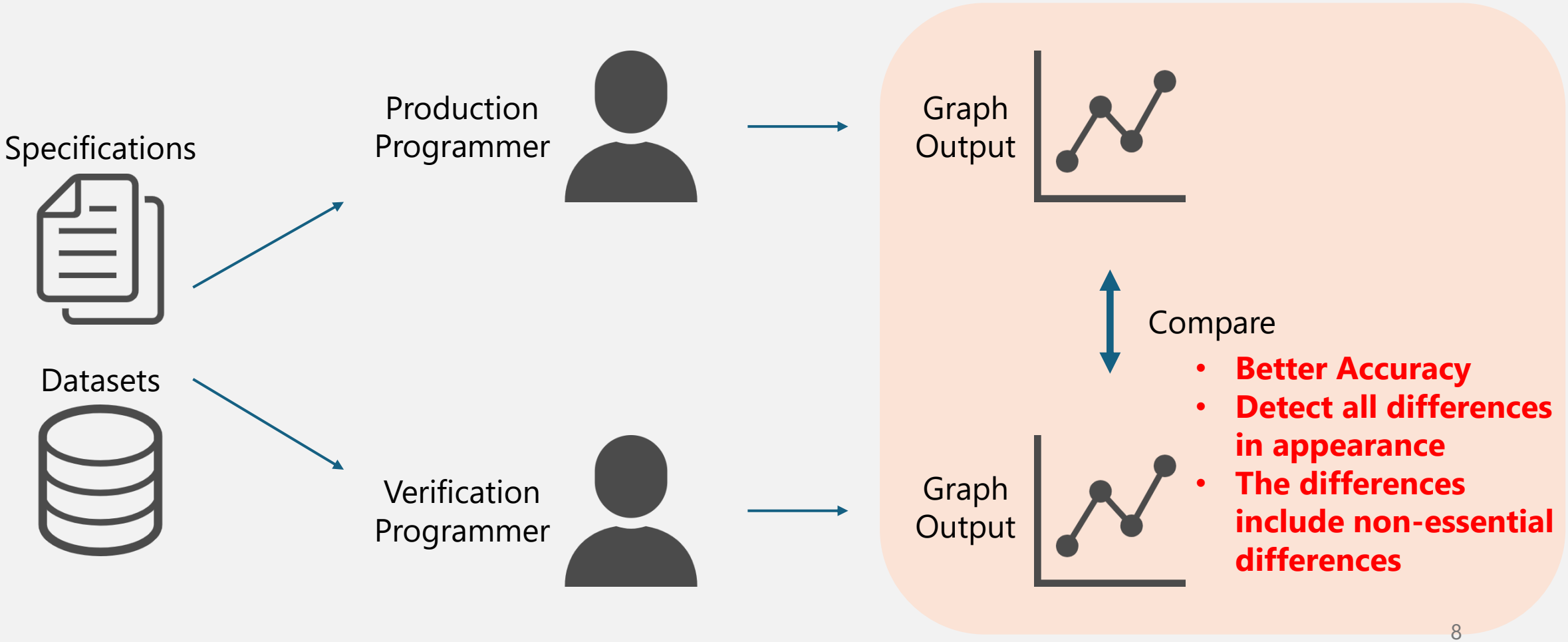
(2). Comparison of Intermediate Datasets

Compare intermediate datasets just before creating graphs



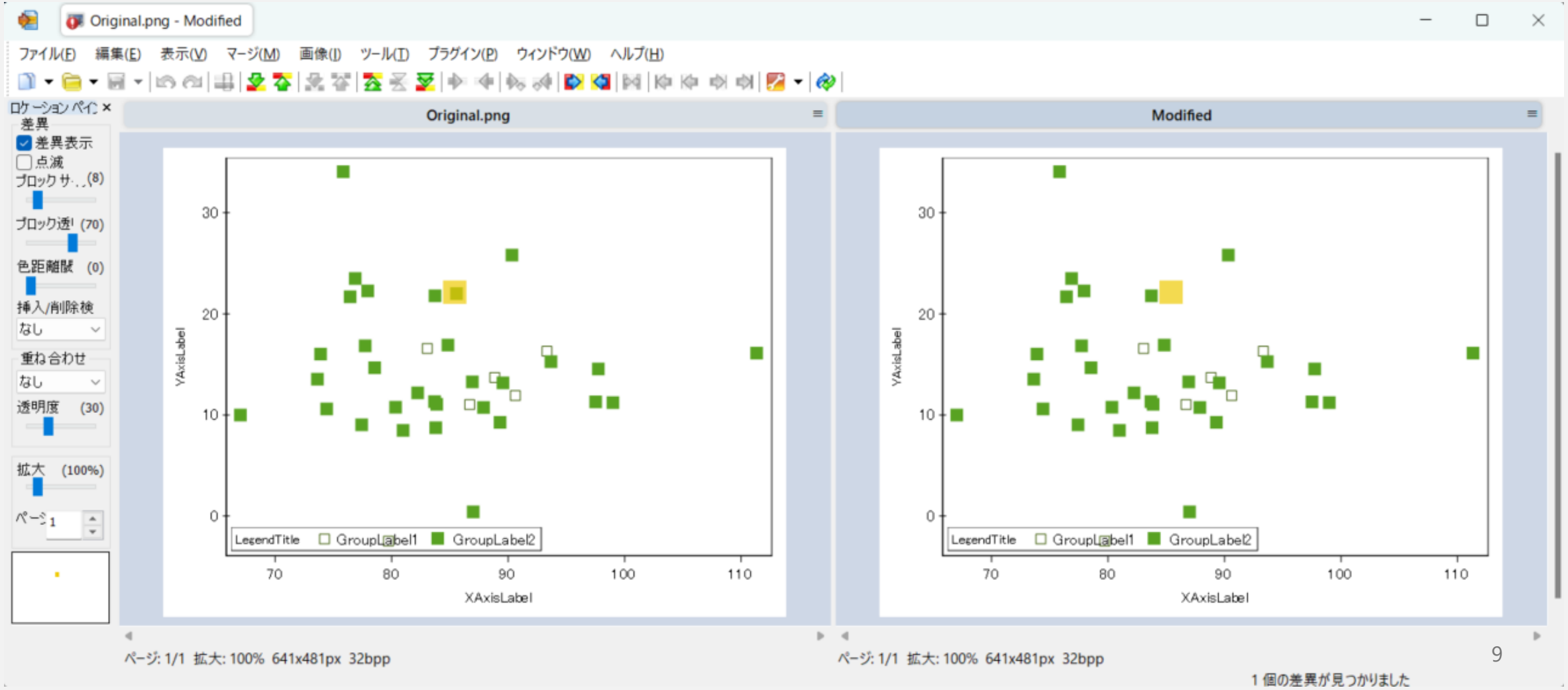
(3). Image Comparison

Compare graphs as images



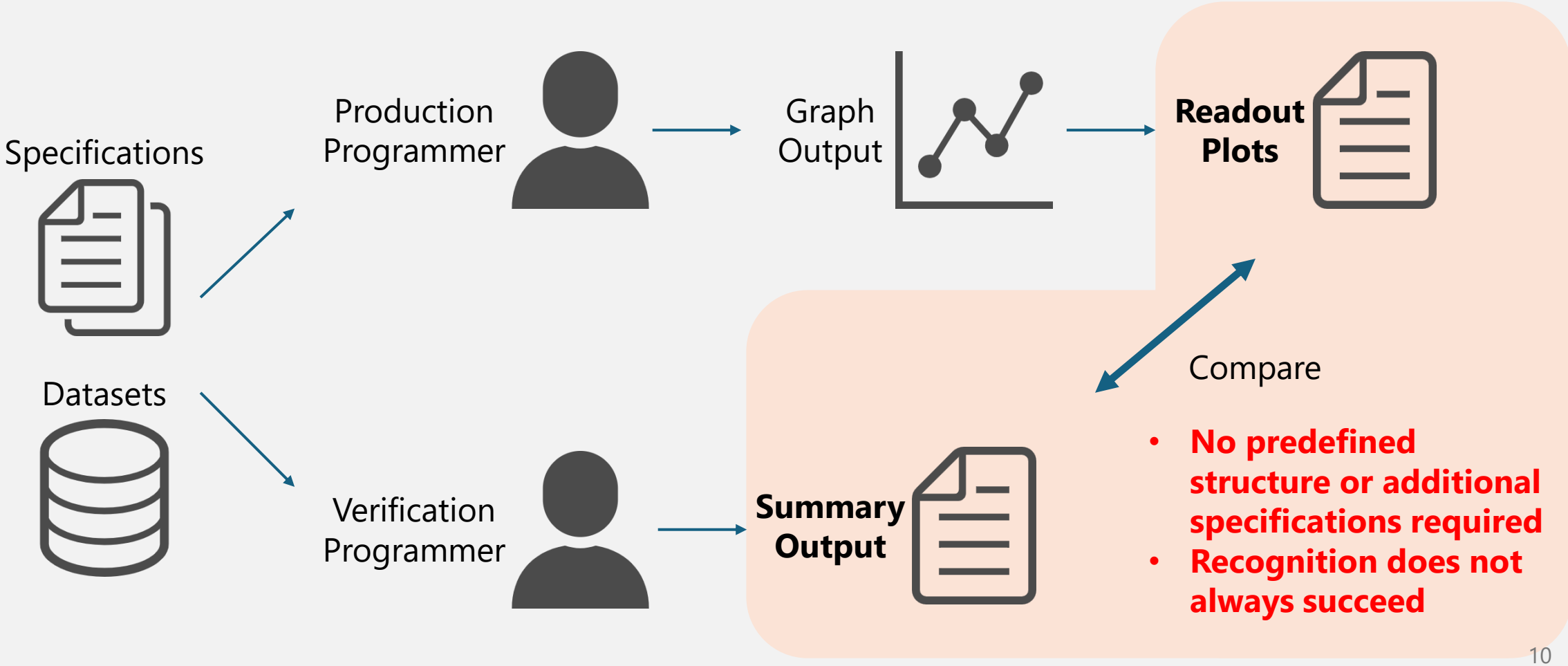
(3). Image Comparison

WinMerge:
Differencing and merging tool for Windows. The differences are highlighted in yellow as follows.



(4). Image Recognition

Recognize graph components and read out plotted points



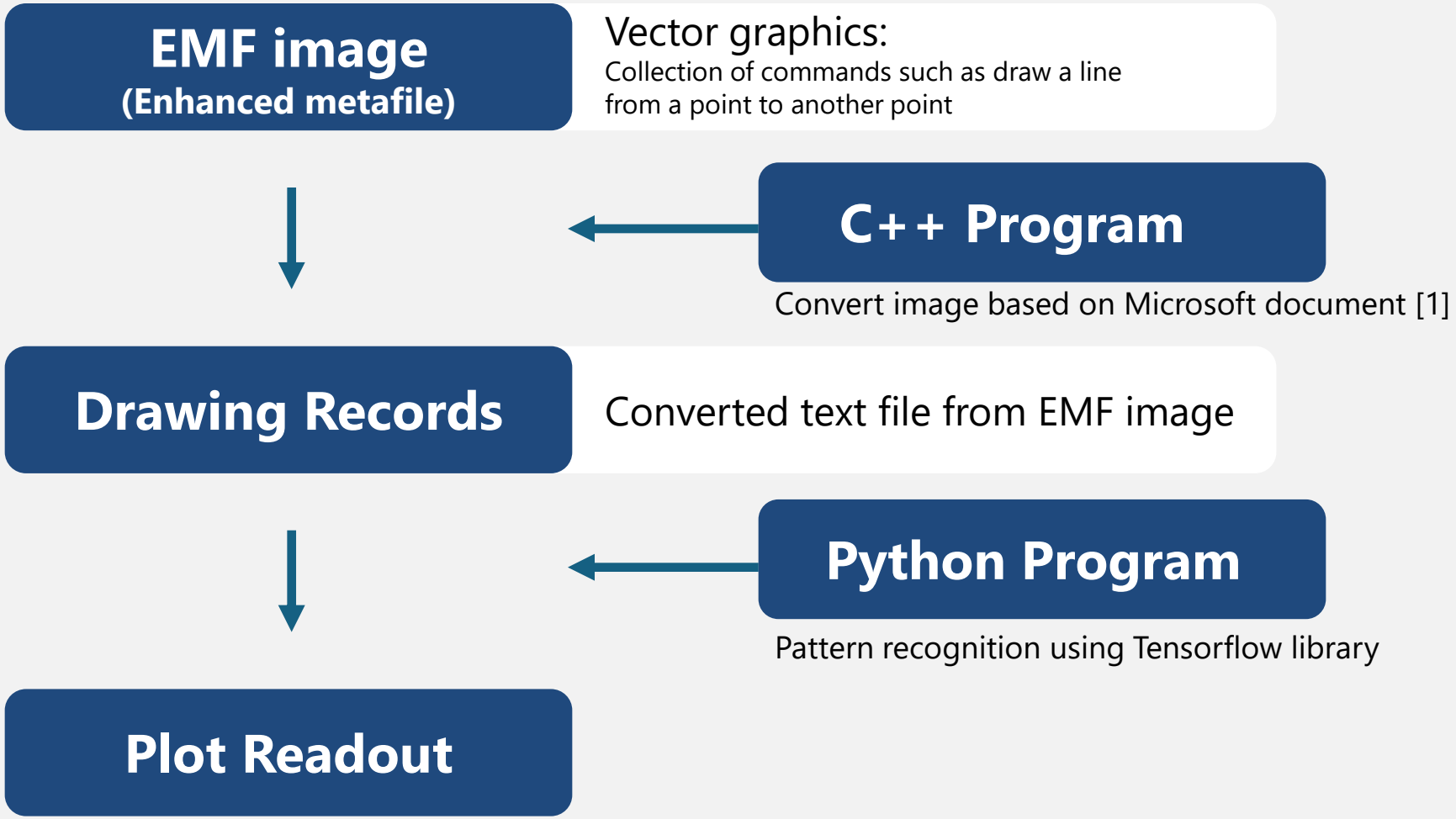
Advantages and Disadvantages

Easy to work on

	Method	Advantage	Disadvantage
(1)	Visually Check	• Requires only a few preliminary steps	• Time consuming • Requires attention • Unclear accuracy
(2)	Comparison of Intermediate Datasets	Compared to (1) • Less time consuming • Better accuracy	• Not verification to graph outputs • Requires predefined structure for intermediate datasets
(3)	Image Comparison	Compared to (2) • Verification to graph outputs	• Requires detailed additional specifications for graph images
(4)	Image Recognition	Compared to (2), (3) • No predefined structure or additional specifications required	• Accuracy depends on training data • Requires human review • No specialized tools (to my knowledge)

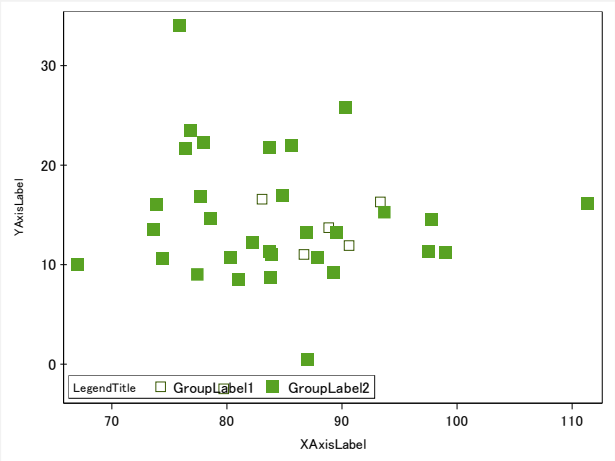
Requires specialized software

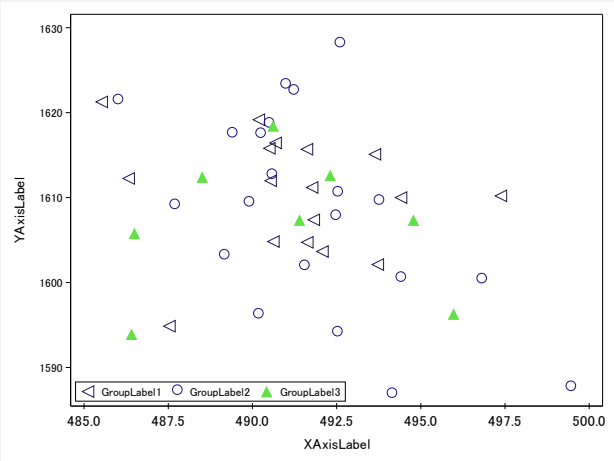
Method for (4). Deep Learning-based Recognition

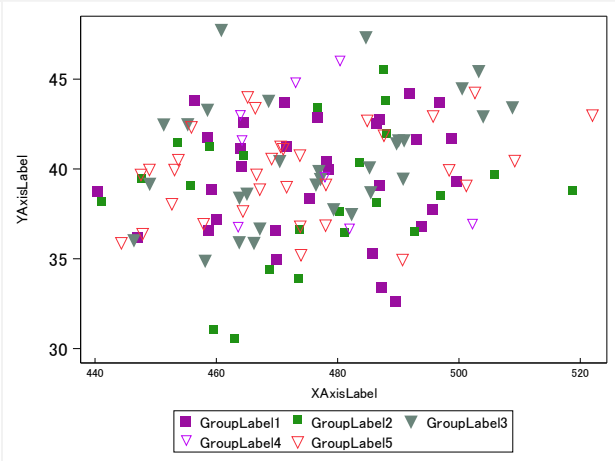


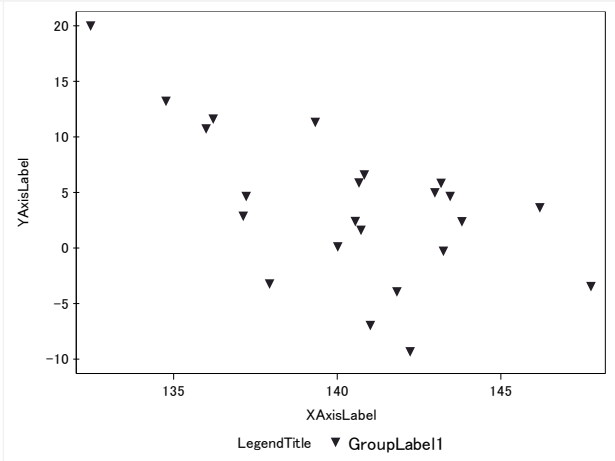
[1] https://learn.microsoft.com/ja-jp/openspecs/windows_protocols/ms-emfplus/5f92c789-64f2-46b5-9ed4-15a9bb0946c6

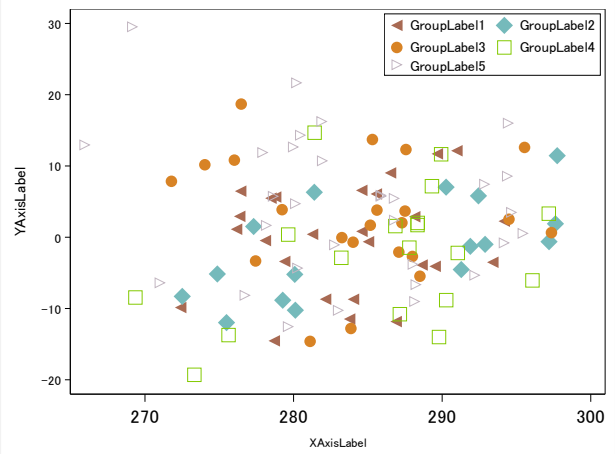
Simulated Examples of Scatter Plots

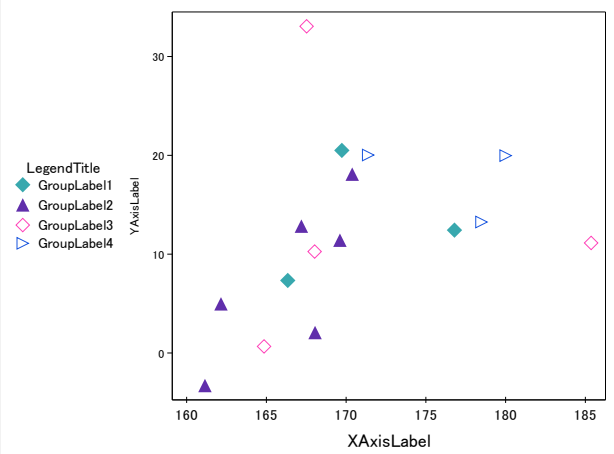


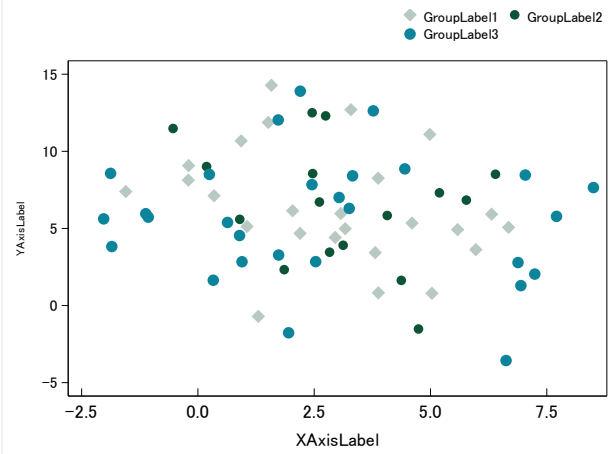


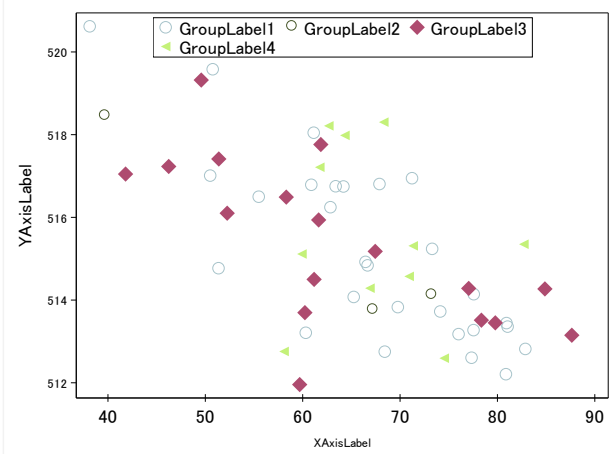












Simulation Condition

Scatter Plots

- Plot function: sgplot (SAS)
- Max group: 5
- Symbols: 14 types

(Shown on the right. Excluded from the choices were grayed out.)

- Data points: 2D normal distribution
- No title, No footnote
- Image size: 640x480
- Color: random RGB
- Size of symbol and label: random
- Legend position: random

Symbol Name	Plot Symbol	Symbol Name	Plot Symbol
ArrowDown	↓	StarFilled	★
Asterisk	✳	Tack	⊤
Circle	○	Tilde	∿
CircleFilled	●	Triangle	△
Diamond	◇	TriangleFilled	▲
DiamondFilled	◆	TriangleDown	▽
GreaterThan	>	TriangleDownFilled	▼
Hash	#	TriangleLeft	◁
HomeDown	⌂	TriangleLeftFilled	◀
HomeDownFilled	⌂	TriangleRight	▷
IBeam	I	TriangleRightFilled	▶
LessThan	<	Union	∪
Plus	+	X	×
Square	□	Y	Y
SquareFilled	■	Z	Z
Star	☆		

https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.5/grstatproc/p0i3rles1y5mvsn1hrq3i2271rmi.htm

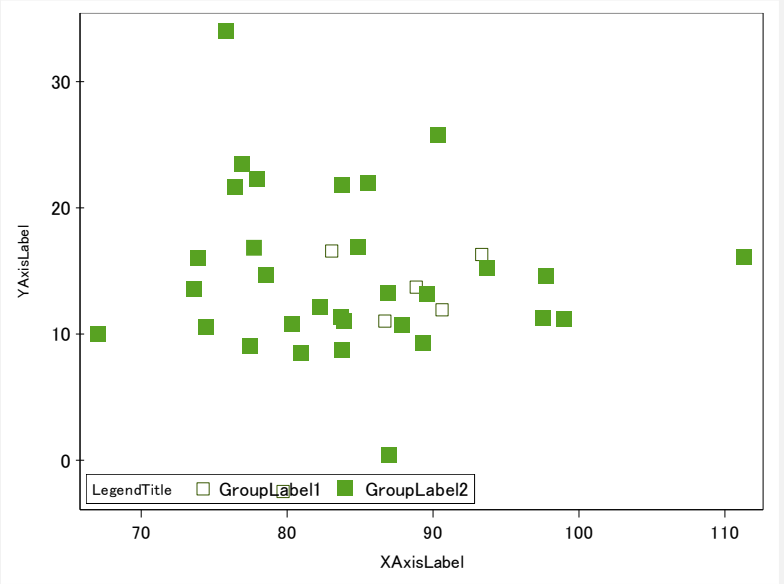
Extracted EMF Records

Treat the drawing positions as straight lines

Path Record, Filled Path Record

X1	Y1	X2	Y2	RGBA (hex)
191.1	127.3	204.1	127.3	58A222FF
204.1	127.3	204.1	140.3	58A222FF
204.1	140.3	191.1	140.3	58A222FF
191.1	140.3	191.1	127.3	58A222FF

(0, 0)



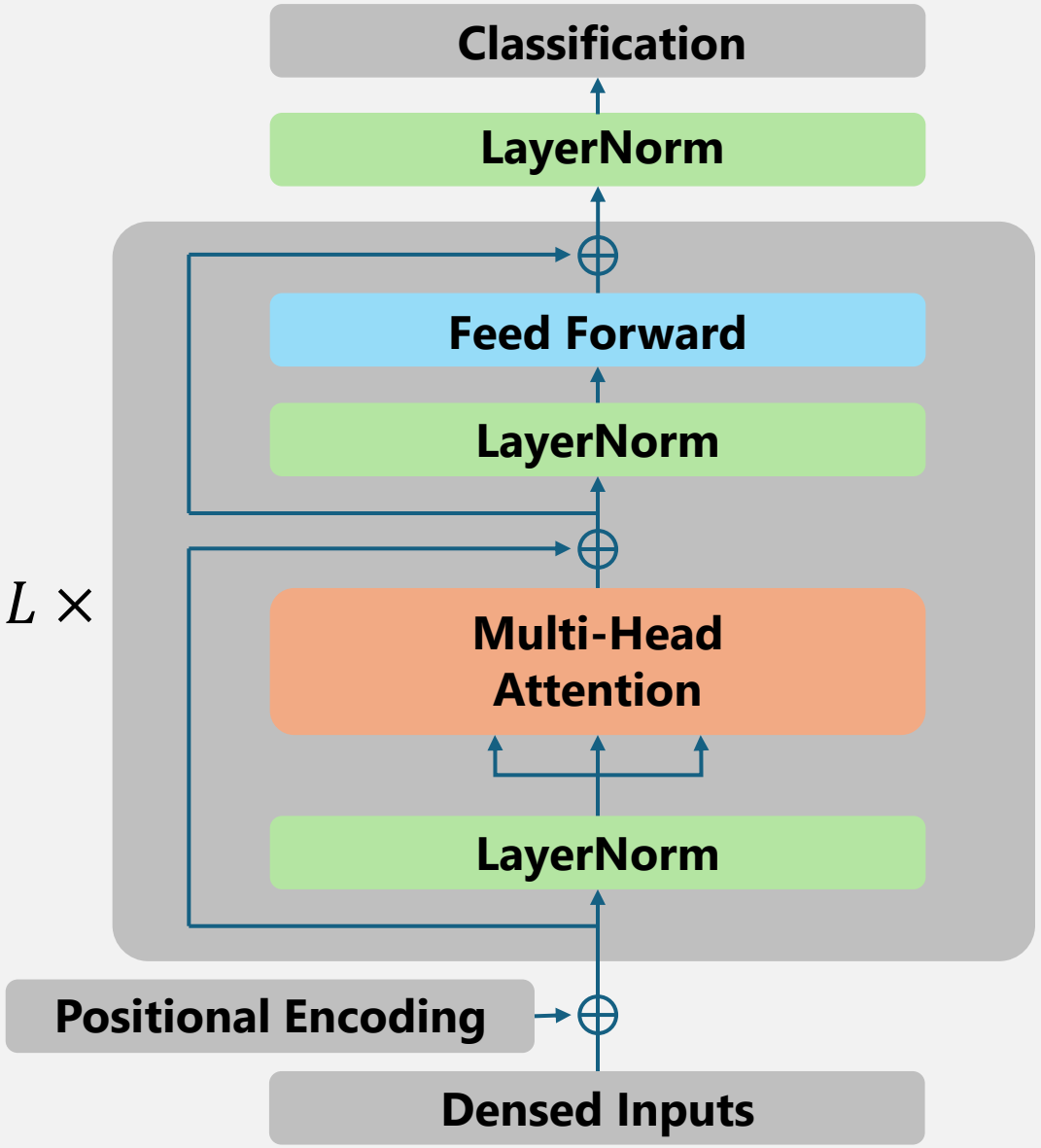
(640, 480)

String Record

X1	Y1	X2	Y2	RGBA (hex)	String	Font
311.6	466.0	377.5	466.0	000000FF	XAxisLabel	MS PGothic
22.7	244.1	22.7	187.1	000000FF	YAxisLabel	MS PGothic

No recognition required for strings shown in graphs

Graph Recognition



Transformer-Encoder:

A deep learning model
based on the multi-head attention

Parameters (2.5M)
Inputs: 23
Class: 27

Image Numbers
Training: 1900
Validation: 50
Test: 50

Accuracy on Test Data
99.88%

Label Assignment

A challenge in supervised learning:
Preparing answer labels is time consuming

Solution:

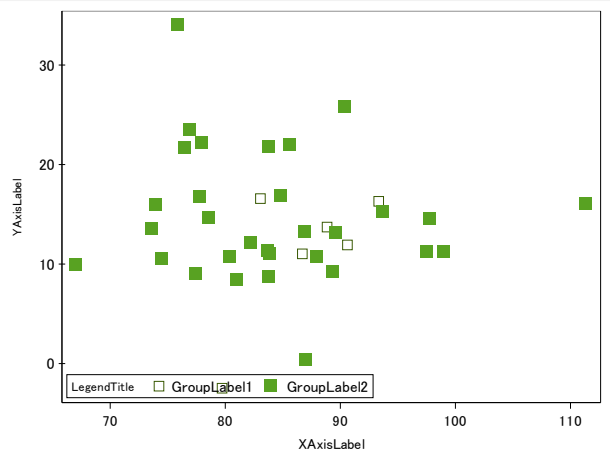
Creating Graphs with Masking

- Creating multiple images with different masks
- The masks make the color transparent or white except in certain areas
- Assigning answer label based on colored areas in masked images

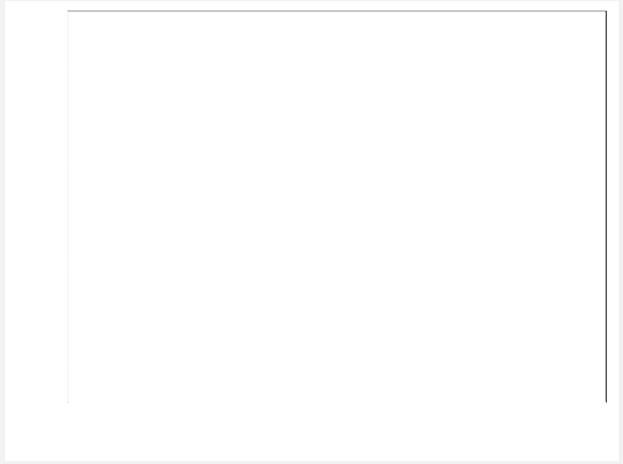
Label Assignment

Create images for each graph element

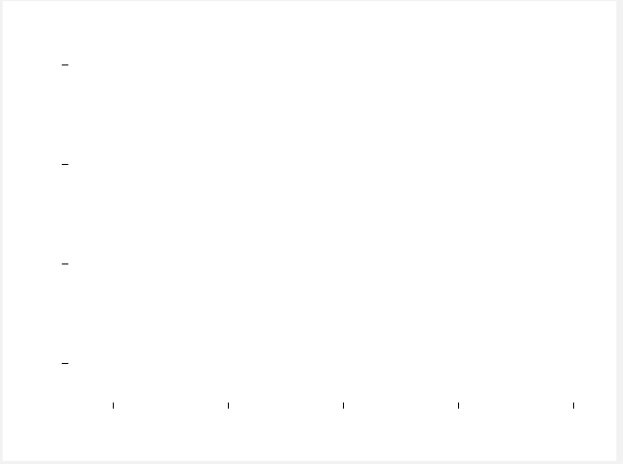
Original Graph



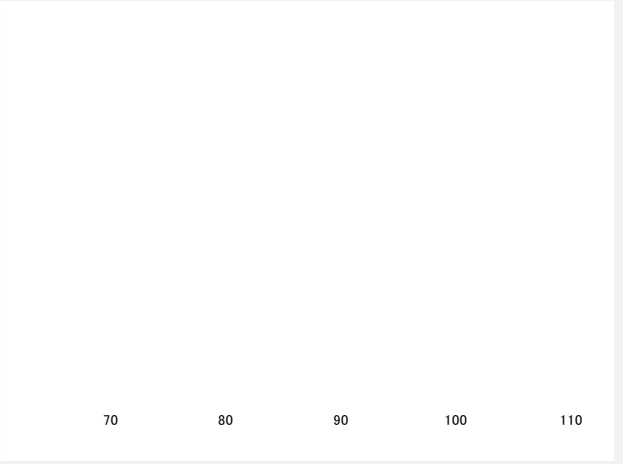
Axis



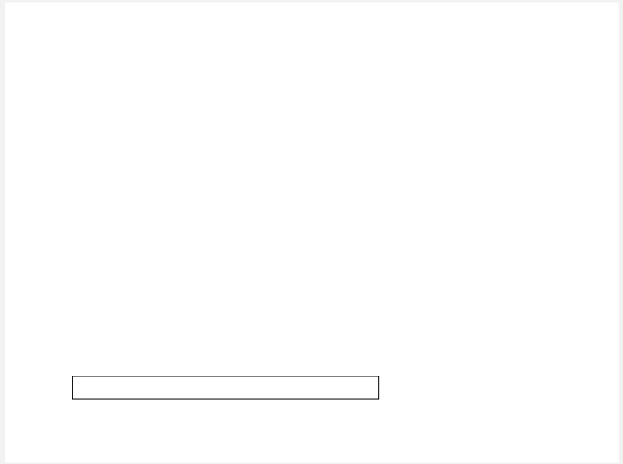
Axis Tick



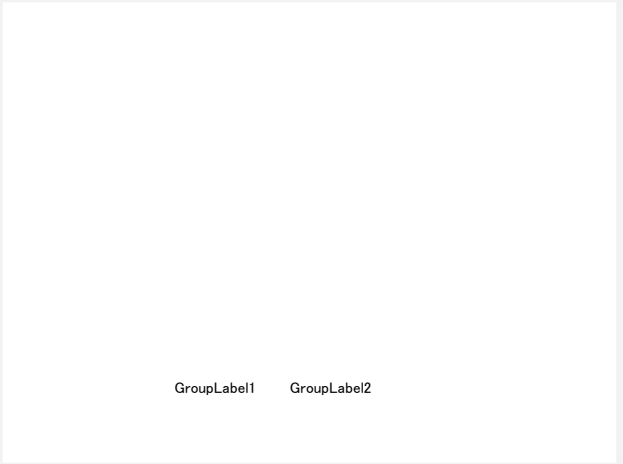
X1 Axis Tick Label



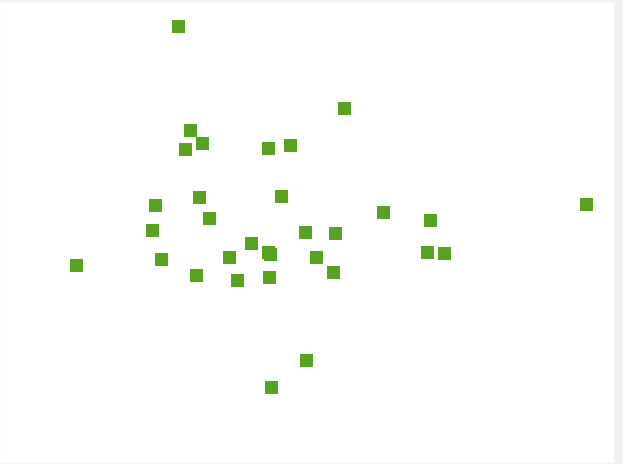
Legend Border



Legend Label



Square Filled



Label Category

Assign 27 labels to recognize lines, strings, and symbols

	Lines
1	X1 Axis
2	Y1 Axis
3	X2 Axis
4	Y2 Axis
5	X1 Axis Tick
6	Y1 Axis Tick
7	Legend Border

	Strings
8	X1 Axis Tick Label
9	Y1 Axis Tick Label
10	X1 Axis Title
11	Y1 Axis Title
12	Legend Title
13	Legend Label

	Symbols
14	Circle
15	Circle Filled
16	Diamond
17	Diamond Filled
18	Square
19	Square Filled
20	Triangle
21	Triangle Filled

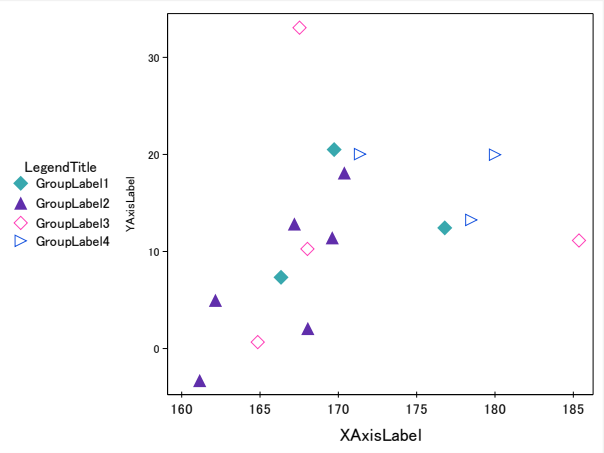
	Symbols
22	Triangle Down
23	Triangle Down Filled
24	Triangle Left
25	Triangle Left Filled
26	Triangle Right
27	Triangle Right Filled

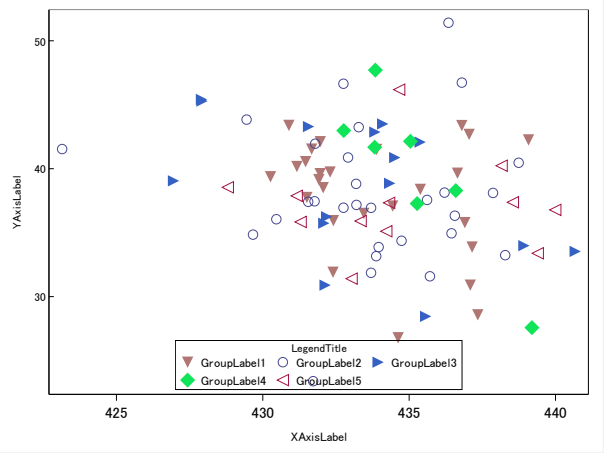
Recognition Mistakes

Mistakes Limited to “Legend Title” and “Legend Label”

	1st Prediction		2nd Prediction	
Answer Label	**Label**	**Probability**	**Label**	**Probability**
Legend Title	Legend Label	0.690	Legend Title	0.309
Legend Title	Legend Label	0.697	Legend Title	0.303
Legend Label	Legend Title	0.760	Legend Label	0.240
Legend Label	Legend Title	0.788	Legend Label	0.212

Even though the 1st prediction was wrong, the 2nd was right.
The mistakes have no effect on reading out plots.



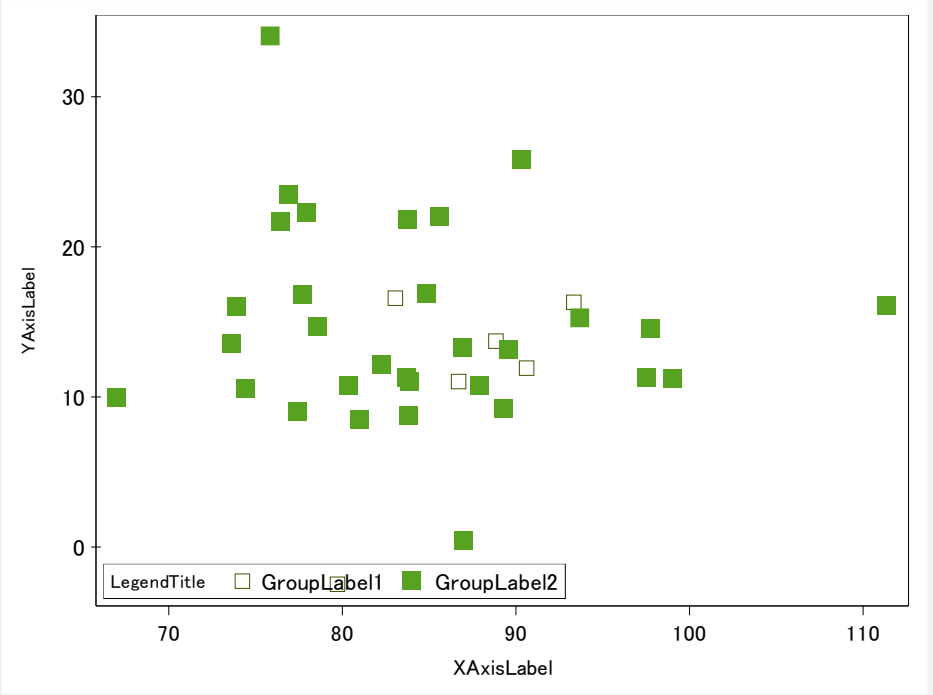


Readout Accuracy

Readout with an Absolute Error of about 1/100

	Absolute Error (X)		Absolute Error (Y)	
	Mean	SD	Mean	SD
Square Filled	4.17e-02	4.90e-06	-4.82e-02	4.03e-06
Square	-7.20e-06	4.40e-06	9.08e-06	4.79e-06

Absolute Error: original plot point – readout plot point (for X or Y)



Summary

1. **This presentation introduced following 4 methods for graph verification:**
   (1). Visually Check
   (2). Comparison of Intermediate Datasets
   (3). Image Comparison
   (4). Image Recognition

2. **The choice of method would depend on the quality required and the amount of time and effort spent**

3. **Although methods based on deep learning are not widely available at present. Once established, they could reduce the effort required.**

Reference

第10回大阪sas勉強会, 統計解析結果のバリデーションを考える
https://www.docswell.com/s/6484025/53G3GV-2024-10-21-172310#p1

WinMerge
https://winmerge.org/?lang=ja

EMF specification
https://learn.microsoft.com/ja-jp/openspecs/windows_protocols/ms-emfplus/5f92c789-64f2-46b5-9ed4-15a9bb0946c6

SAS, Marker Attributes and Symbols
https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.5/grstatproc/p0i3rles1y5mvsn1hrq3i2271rmi.htm

Automated Validation of Complex Clinical Trials Made Easy,
https://www.mwsug.org/proceedings/2017/PH/MWSUG-2017-PH05.pdf

Statistical Programmer’s Top Challenges in Clinical Trial Data Validation
https://beaconcure.com/resources/statistical-programmer-challenges/

Icon Images
https://icooon-mono.com/