DATA IS POTENTIAL

Identifying Defect Patterns in Hard Disk Drive Magnetic Media Manufacturing Processes Using Real and Synthetic Data

NVIDIA GPU TECHNOLOGY CONFERENCE

Nicholas Propes | Seagate Analytics

San Jose, CA

March 29, 2018

SEAGATE

Outline

- **Seagate Technology**
- **Magnetic Media, Scanned Data and Defect Patterns**
- **Manual Feature Extraction**
- **Automated Feature Extraction**
- **Architecture / Implementation**
- **Results**

Seagate's Global Presence

- Beaverton, OR, USA
- Fremont, CA, USA
- Cupertino, CA, USA
- Valencia, CA, USA
- Longmont, CO, USA
- Colorado Springs, CO, USA
- Oklahoma City, OK, USA
- Shakopee, MN, USA
- Bloomington, MN, USA
- Rochester, MN, USA
- Houston, TX, USA
- Round Rock, TX, USA
- Guadalajara, Mexico
- São Paulo, Brazil

- Danderyd, Sweden
- Dublin, Ireland
- Springtown, N. Ireland
- Paris, France
- Havant, UK
- Maidenhead, UK
- Munich, Germany
- Amsterdam, Netherlands
- Moscow, Russia

- Korat, Thailand
- Teparuk, Thailand
- Johor, Malaysia
- Penang, Malaysia
- Shugart, Singapore
- Woodlands, Singapore
- Sydney, Australia
- New Delhi, India
- Mumbai, India
- Pune, India
- Bangalore, India

- Tokyo, Japan
- Taipei, Taiwan
- Hong Kong, China
- Wuxi, China
- Shenzen, China
- Chengdu, China
- Shanghai, China
- Tianjin, China
- Beijing, China

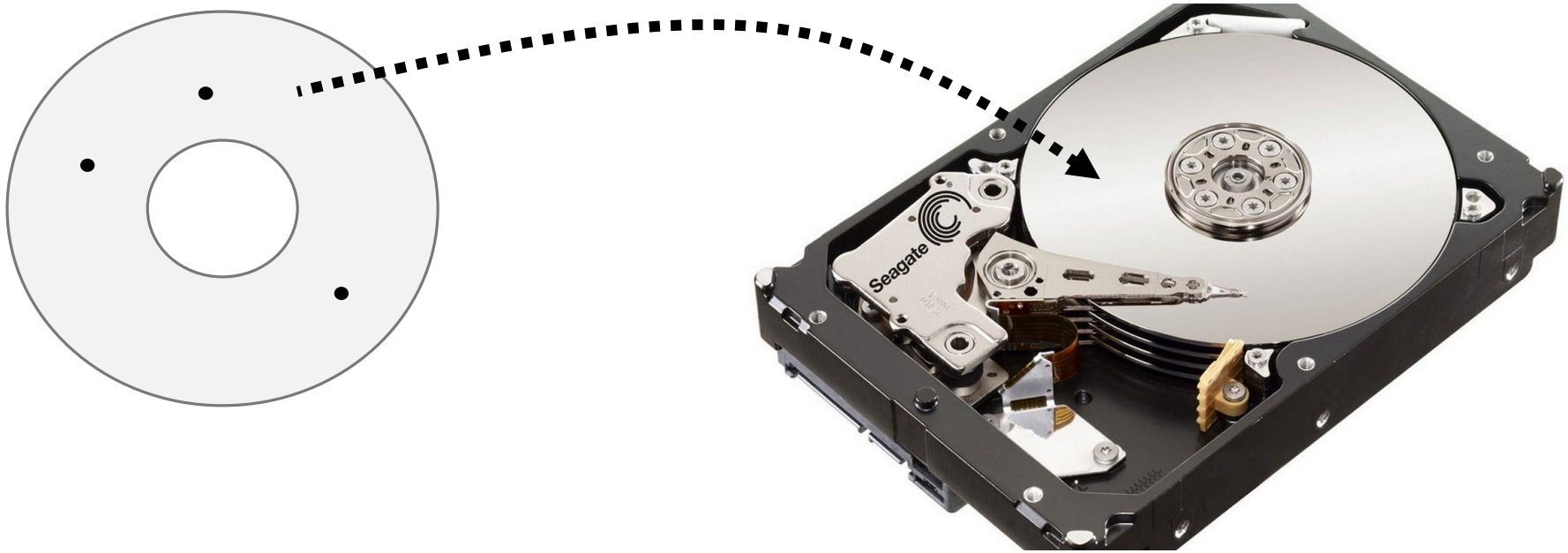
HQs, Admin/Sales	Design	Manufacturing	Customer Support

Hard Drive / Magnetic Media

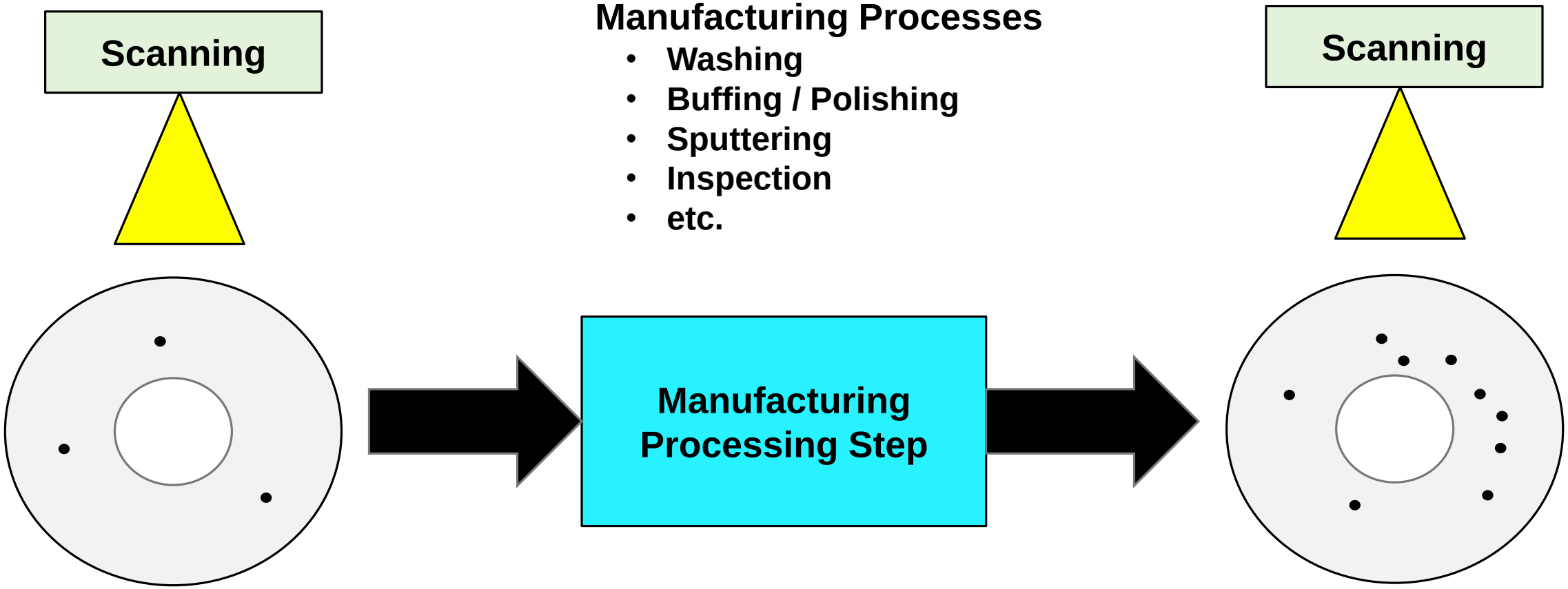
- **Complex System**
- **> 300,000 tracks per inch**
- **Read/write head fly height < 20 angstroms**
- **Rotation speed 4500-15000 RPM**
- **Control of read/write head**
- **Lots of testing for different parameters**
- **HAMR area density (2 TB / sq in)**

Problem Definition

Objective: Classify defect patterns that occur on scanned magnetic media for the purpose of identifying issues in manufacturing line.



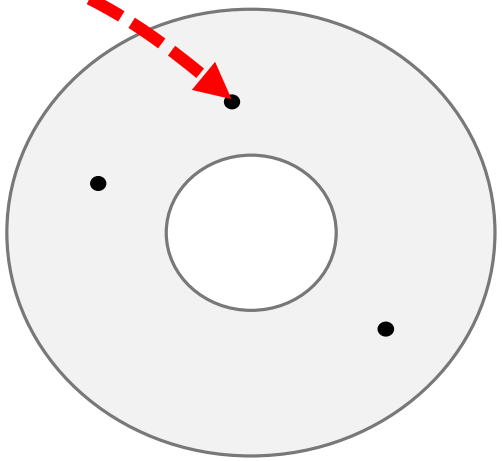
Scanning Magnetic Media Defects



Data

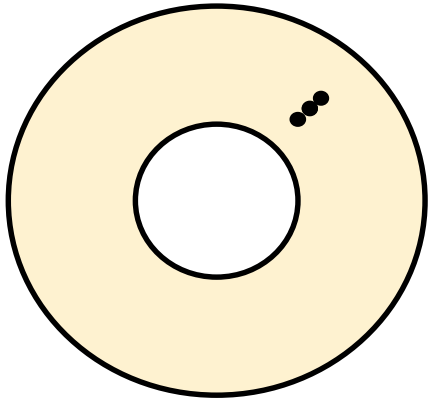
Defect Point Locations on Magnetic Media

ID	SIDE	Radius	Angle (Deg)
A1234	A	35000	20
A1234	A	64301	50
A1234	A	45000	185
A1234	A	21443	354
…	…	…	…
C3212	B	54531	124
C3212	B	34222	342
C3212	B	18888	351

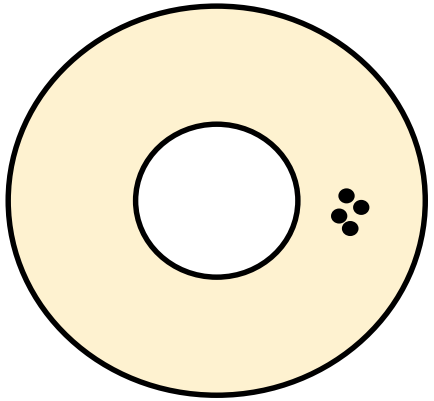


Defect Patterns

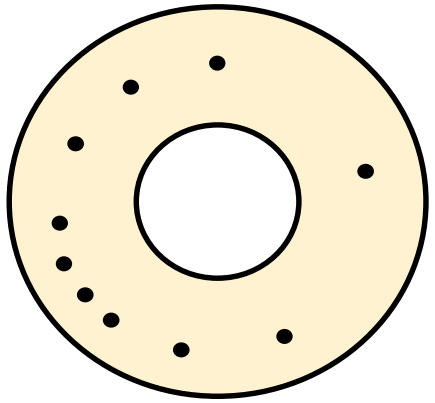
Pattern A



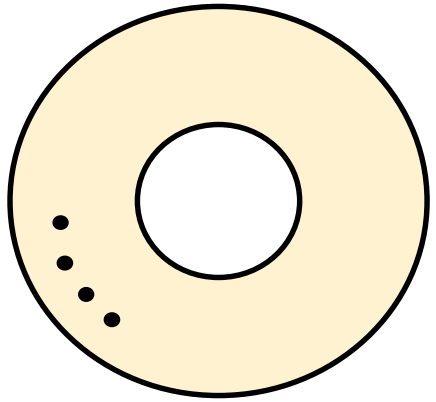
Pattern B



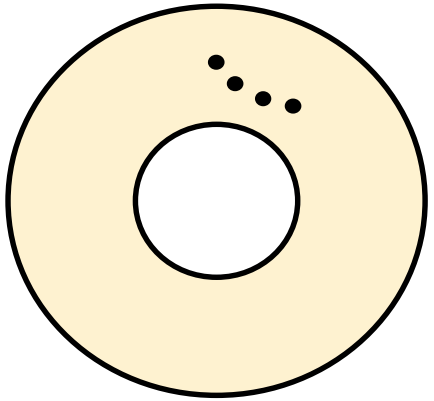
Pattern C



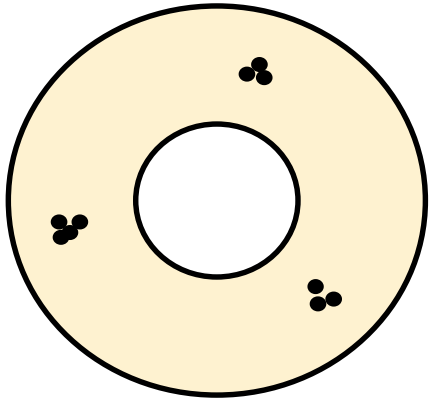
Pattern D



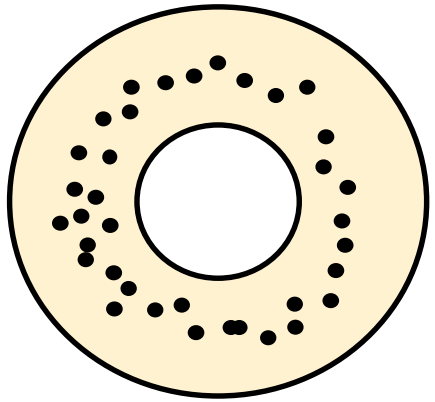
Pattern E



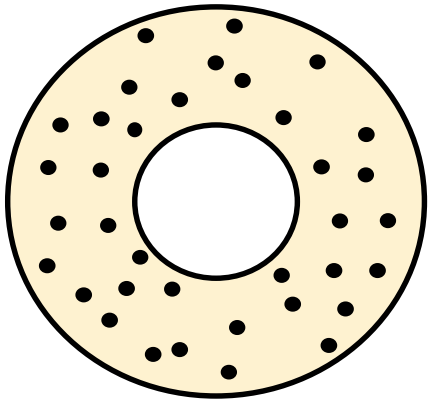
Pattern F



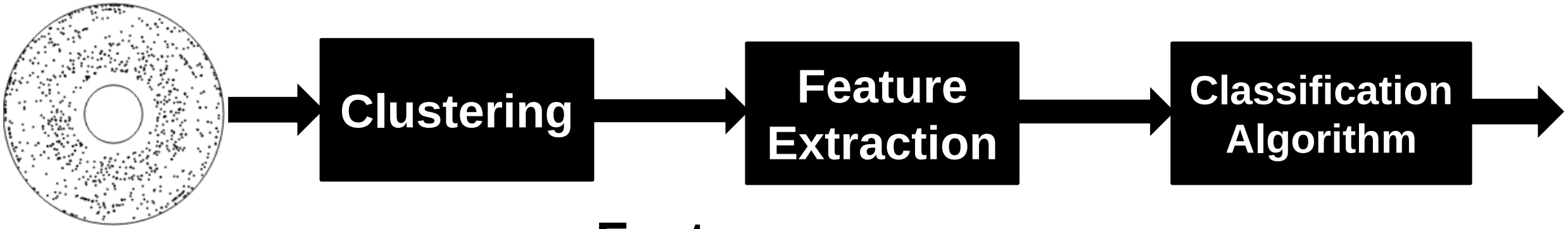
Pattern G

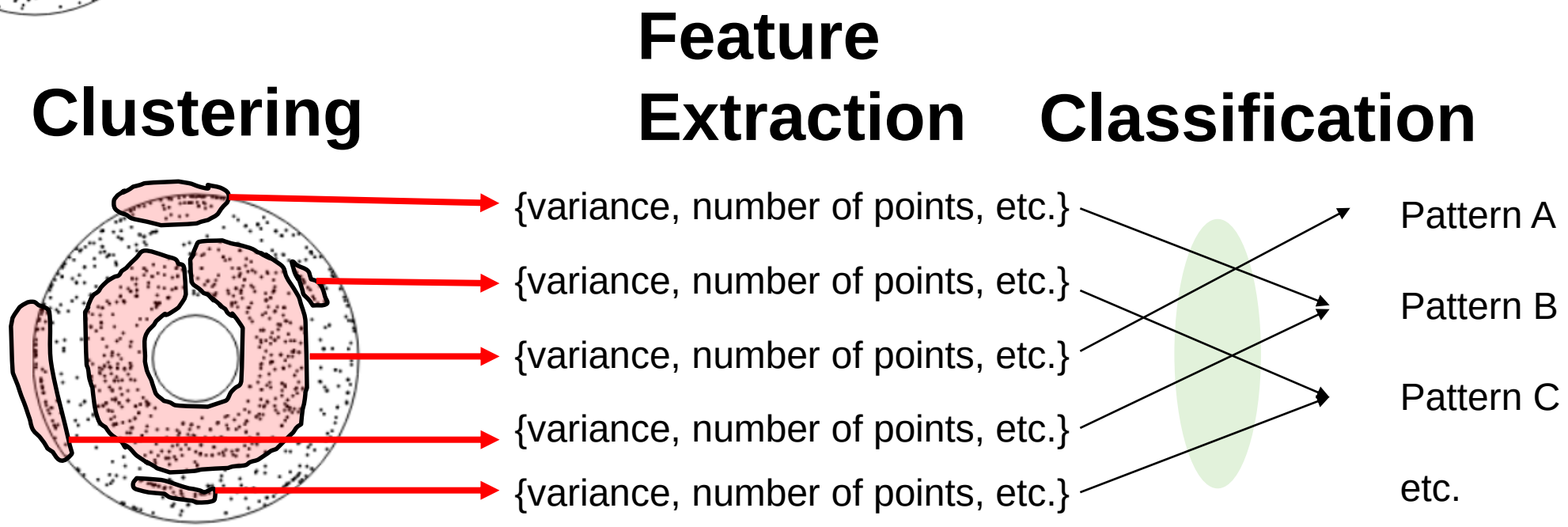


Pattern H



Method 1: Manual Feature Engineering

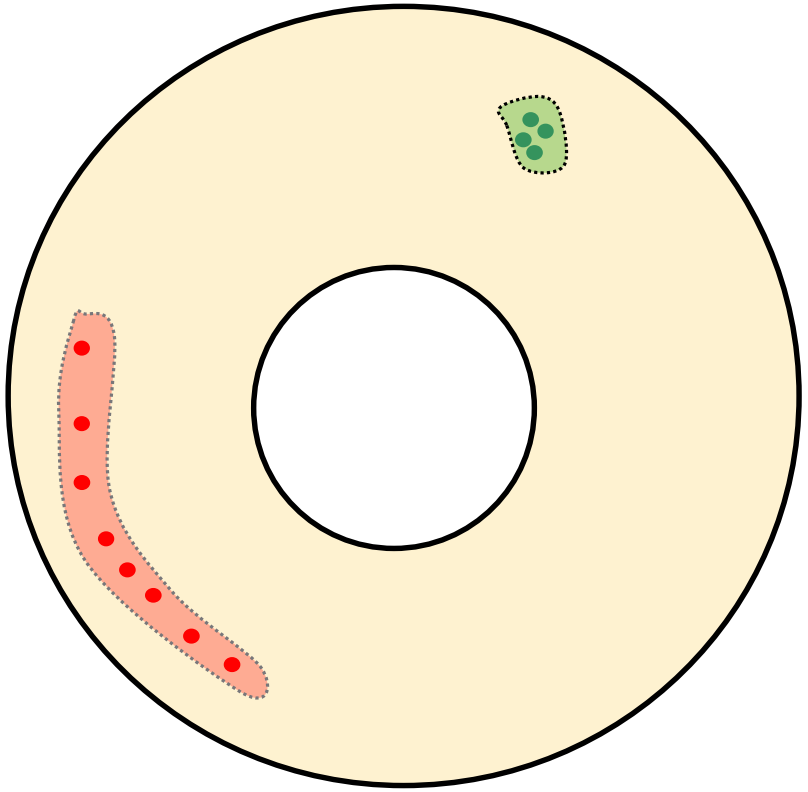


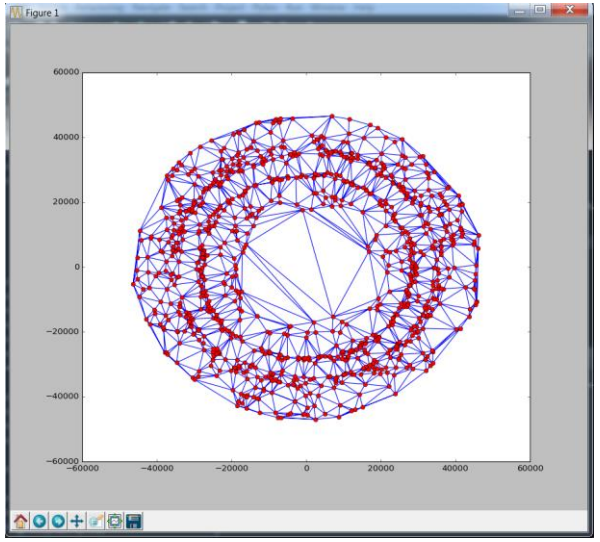


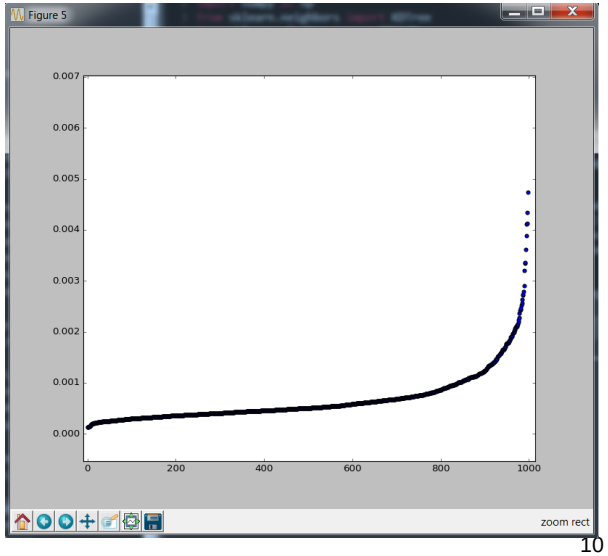
Method 1: Manual Feature Engineering

Clustering Algorithms

- Spatial Grouping
- KDClus
- Tesselation
- Band-pass Filtering / Downsampling Images
- Density-based Scan (DBSCAN)
- etc.



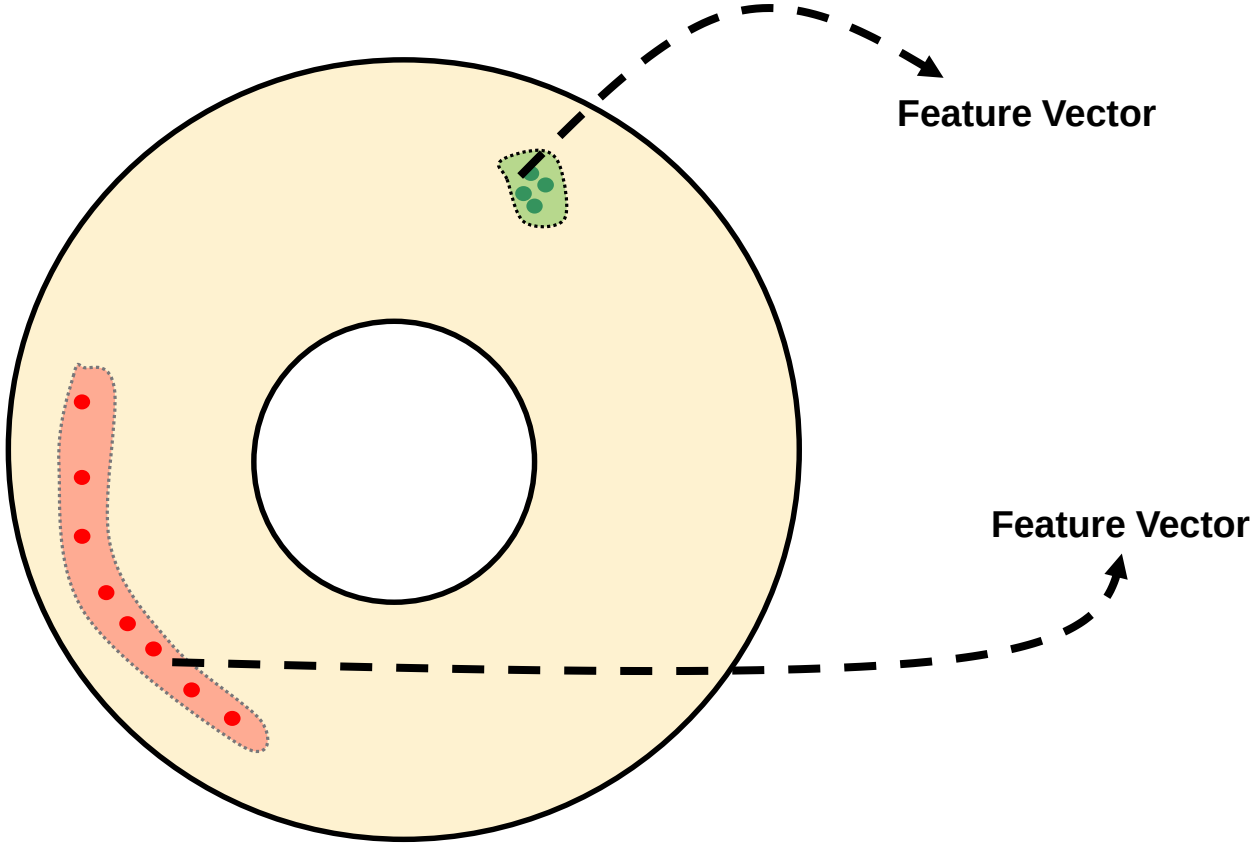




Method 1: Manual Feature Engineering

Feature Extraction

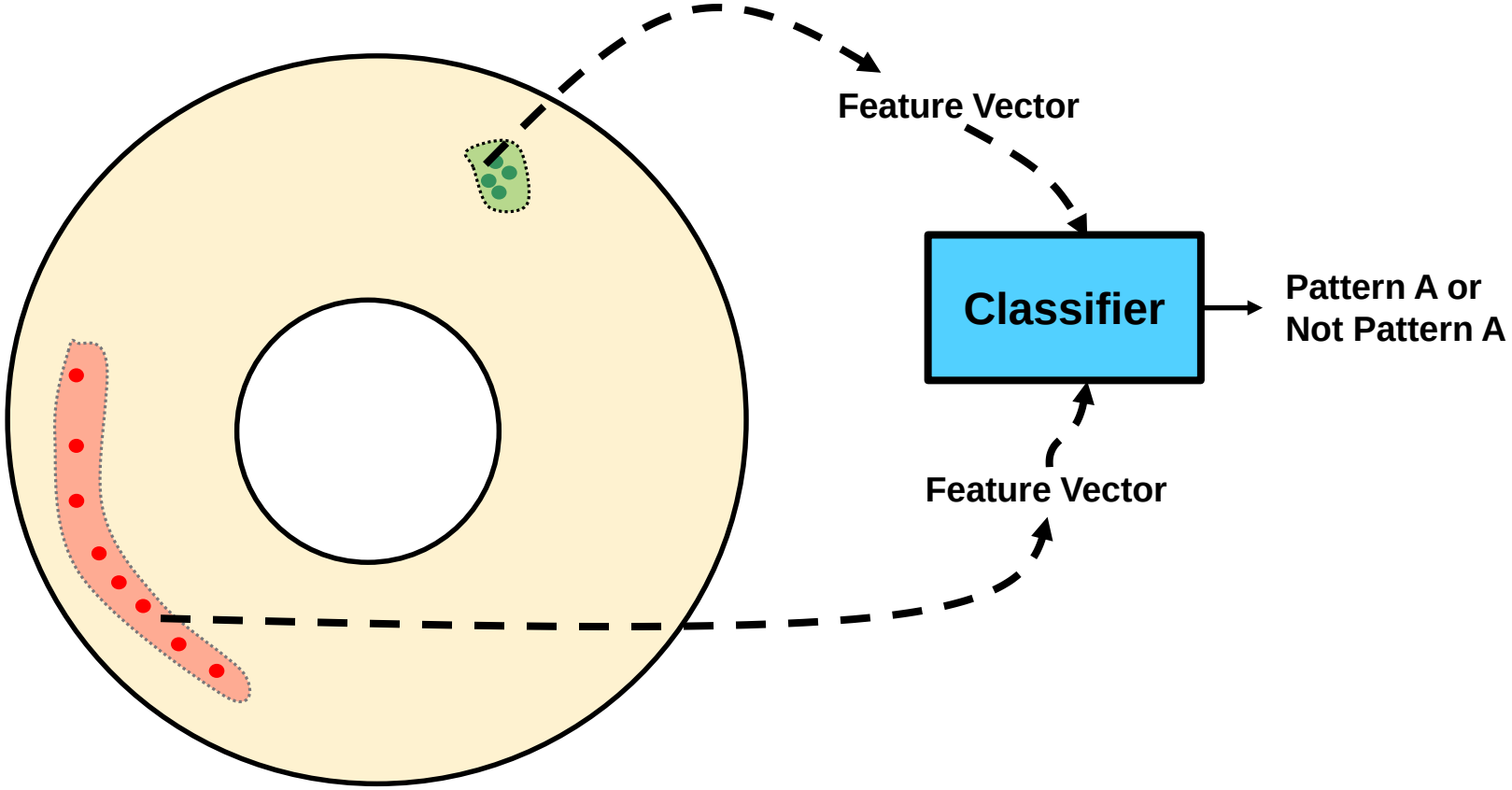
- cluster defect counts
- cluster lengths
- cluster widths
- cluster variances
- entropy
- etc.



Method 1: Manual Feature Engineering

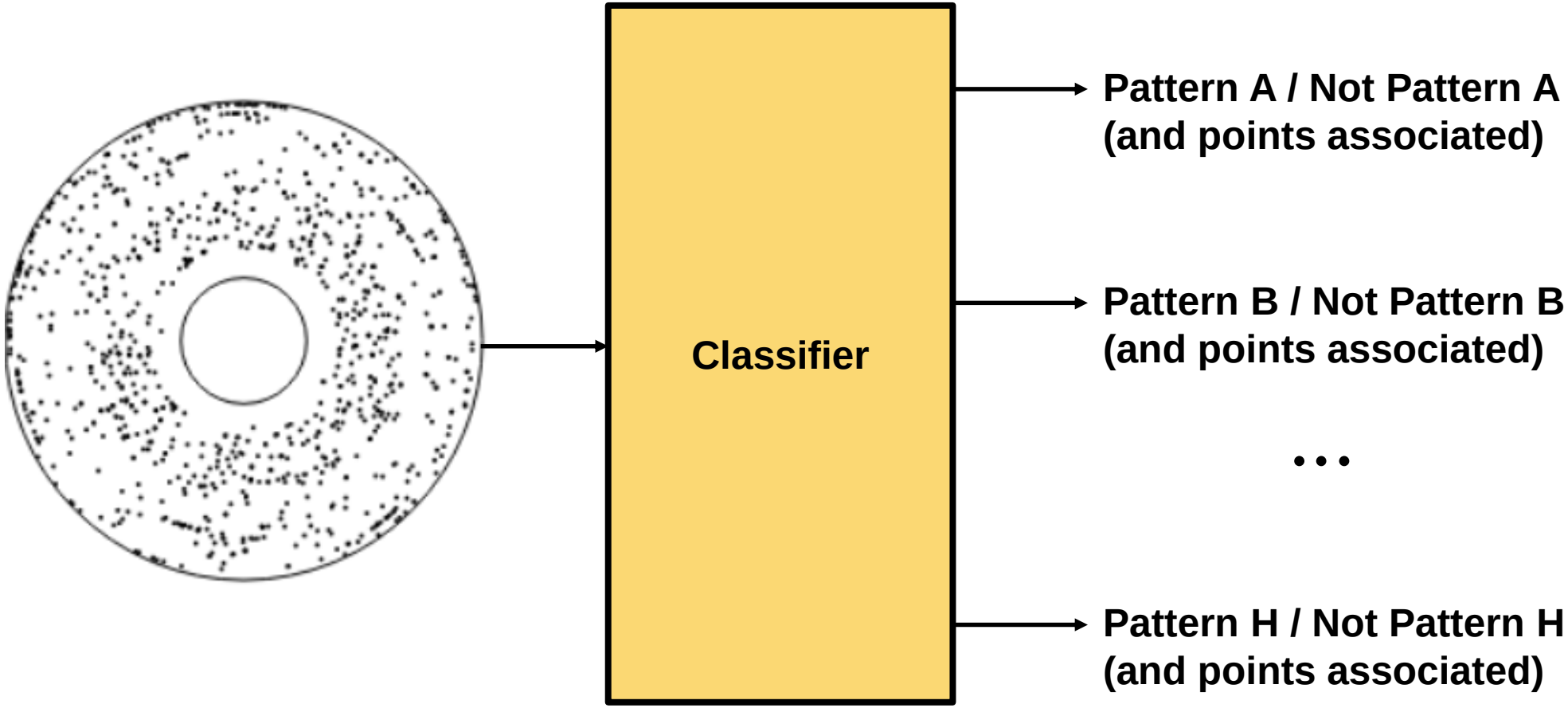
Classifiers

- decision trees
- fuzzy logic
- logistic regression



Method 1: Manual Feature Engineering

Classification Scheme



Method 1: Manual Feature Engineering

Classification Scheme

Pattern A Classifier → **Pattern A / Not Pattern A (and points associated)**

Pattern B Classifier → **Pattern B / Not Pattern B (and points associated)**

… **…**

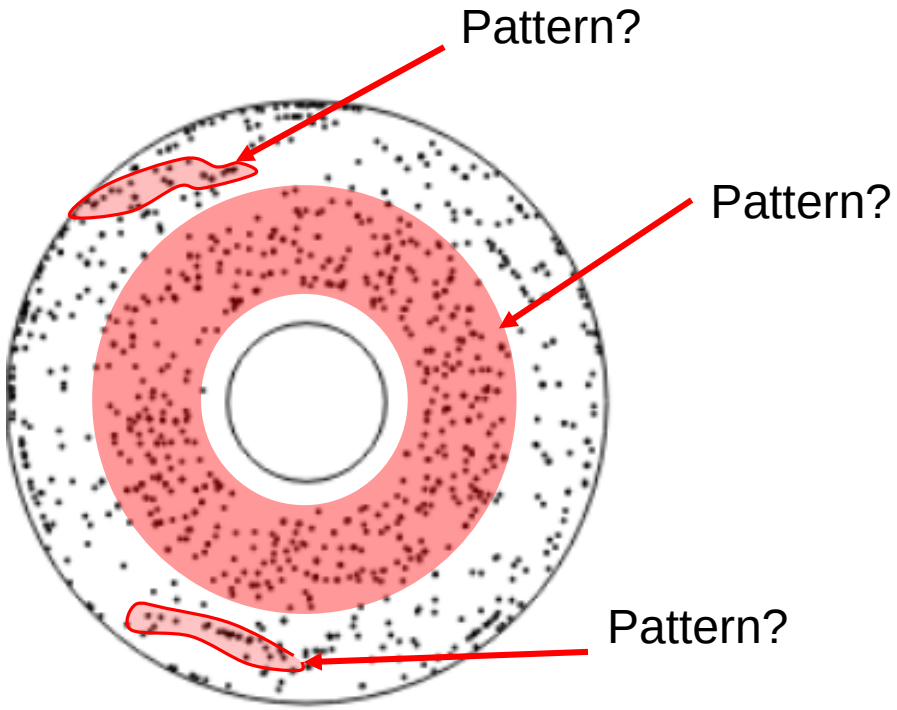
Pattern H Classifier → **Pattern H / Not Pattern H (and points associated)**

Method 1: Manual Feature Engineering

Issues

- Noisy patterns
- Density changes for defect patterns
- Overlapping patterns

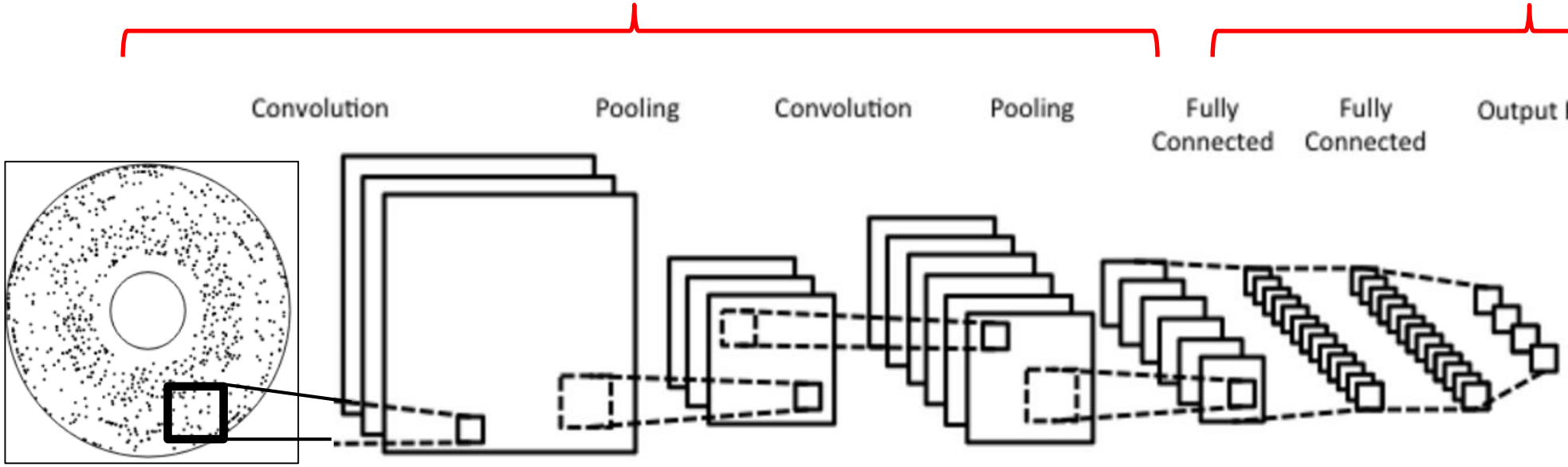
Makes clustering difficult to perform reliably!



Method 2: Automatic Feature Engineering

- **Multiple Image Processing Layers**
- **Image Processing Functions are Learned from Data**

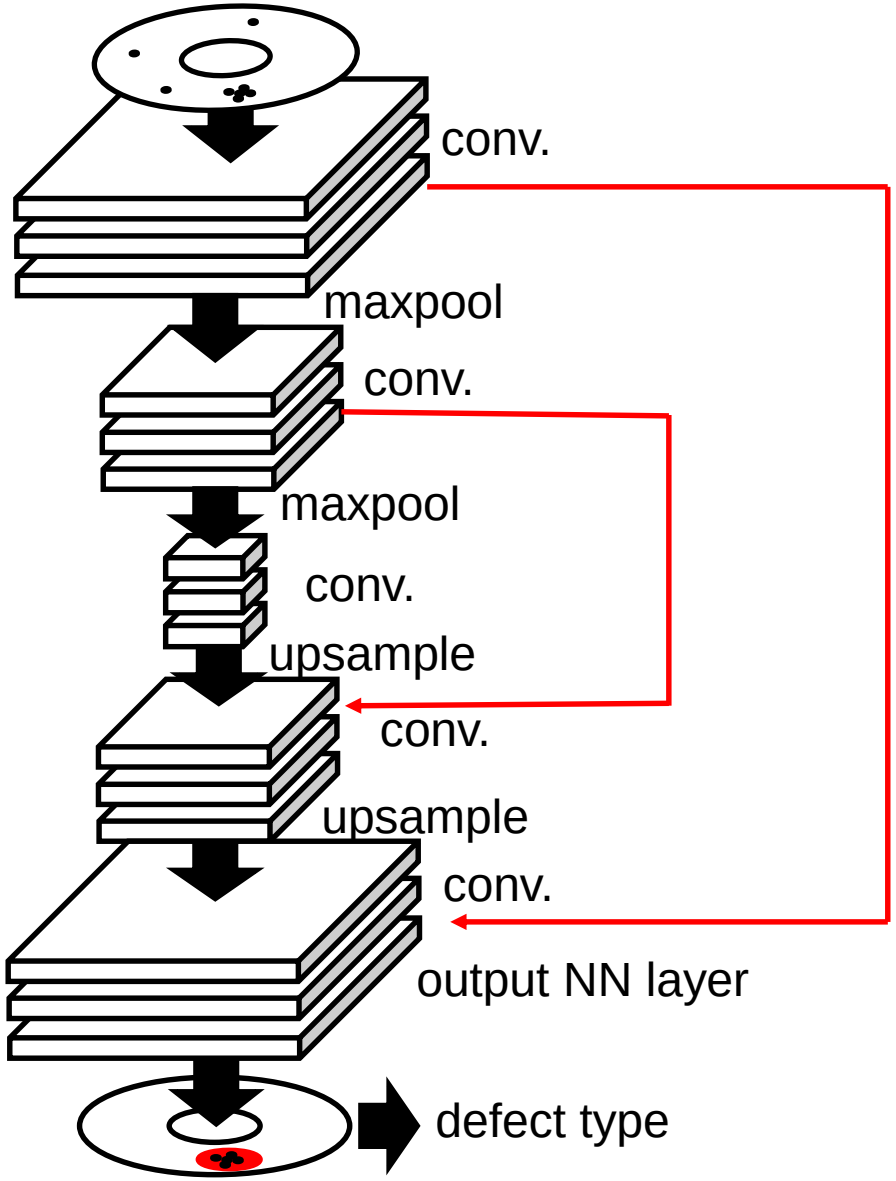
- **Basic Neural Net Classifier**
- **Parameters are Learned from Data**

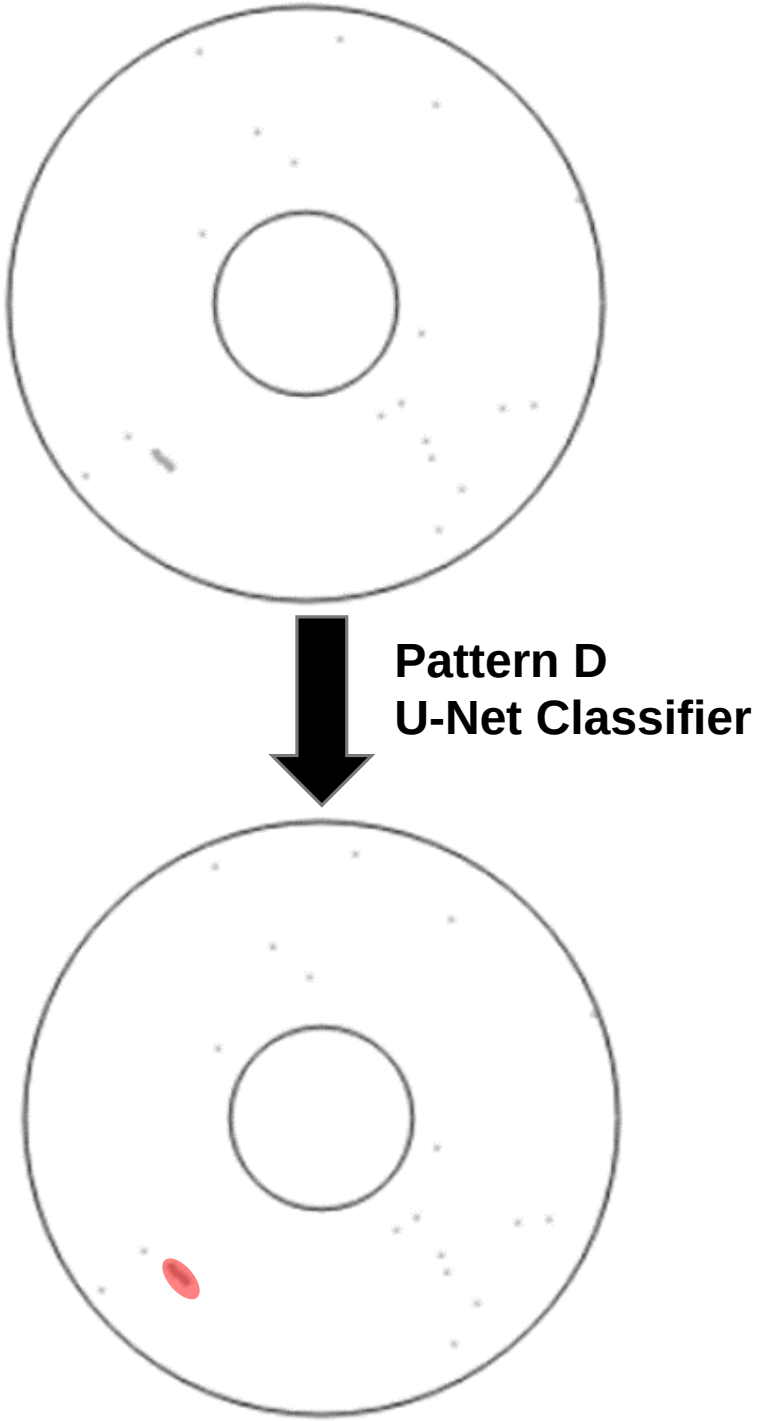


Band (0.9)
Heavy Galaxy (0.8)
S_Circ_MD (0.8)
S_Circ_OD (0.7)
Circ_Scratch (0.1)
…

U-Net Image Segmentation

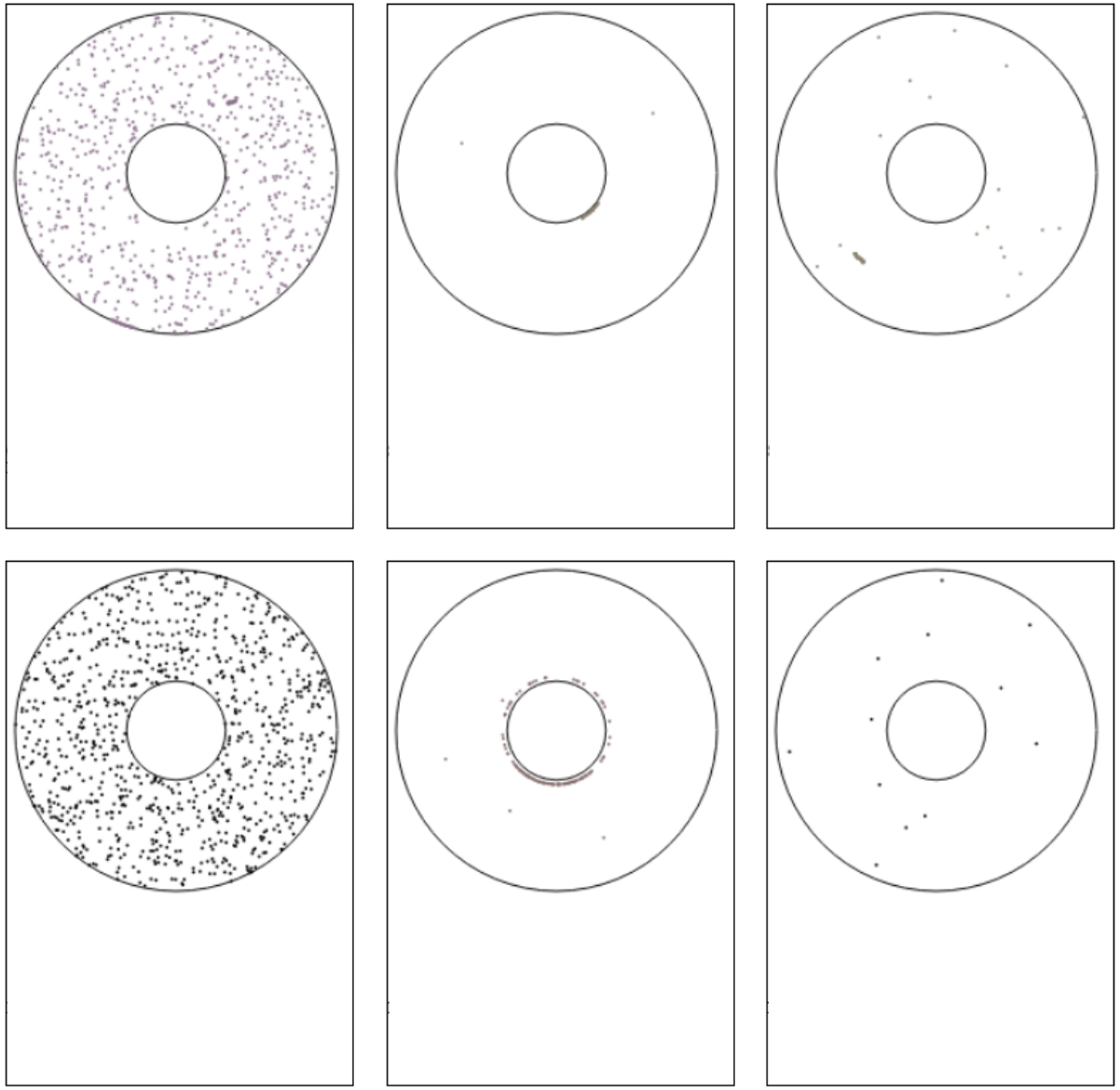
Image Segmentation





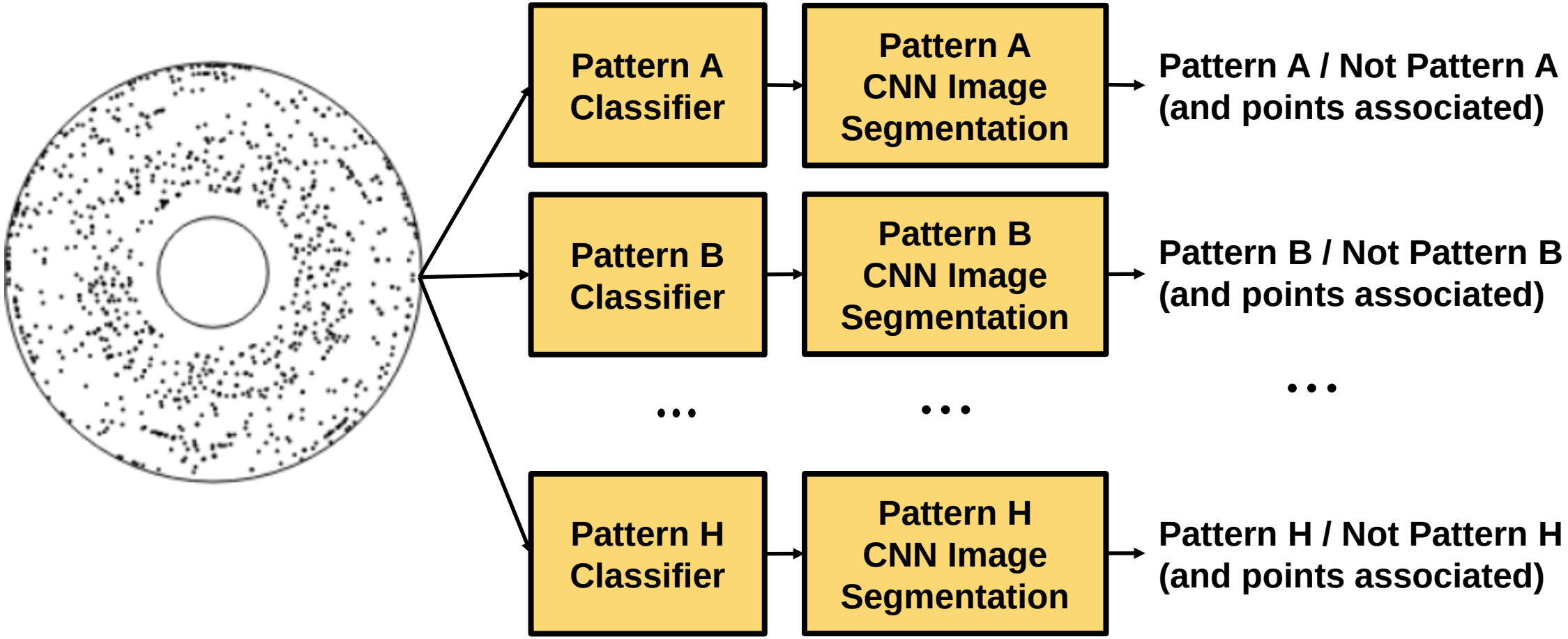
Synthetic Data Generation


```
<?xml version="1.0"?>
- <[redacted]>
    [redacted]
    <file_prefix>FILE</file_prefix>
    <file_start_index>0</file_start_index>
    <number_of_files>1000</number_of_files>
    <number_of_discs_per_file>100</number_of_discs_per_file>
    <max_pattern_per_disc>3</max_pattern_per_disc>
  - <pattern>
      <pattern_name>[redacted]</pattern_name>
      <pattern_type>[redacted]</pattern_type>
      <param>min_num_defects : 21</param>
      [redacted]
    </pattern>
  - <pattern>
      <pattern_name>[redacted]</pattern_name>
      <pattern_type>[redacted]</pattern_type>
      <param>min_num_defects : 21</param>
      [redacted]
    </pattern>
  - <pattern>
      [redacted]
```

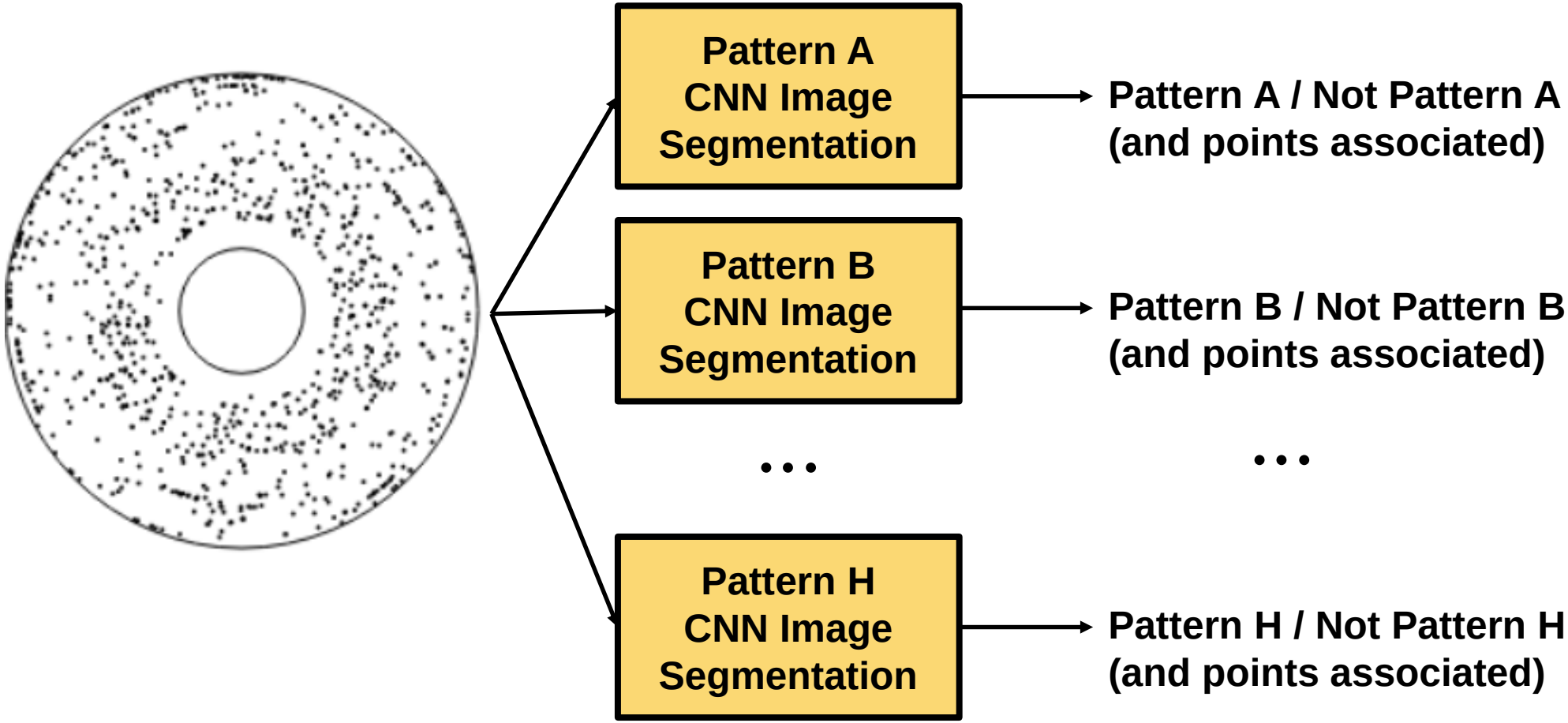
Method 1: Manual Feature Engineering

Classification Scheme

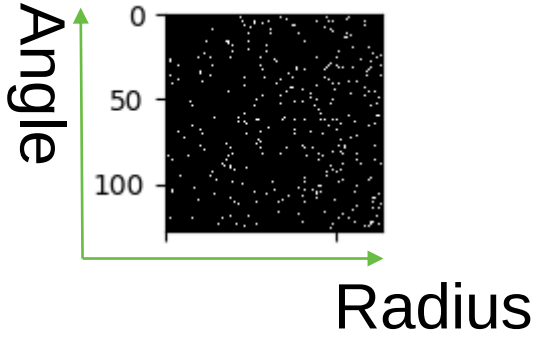


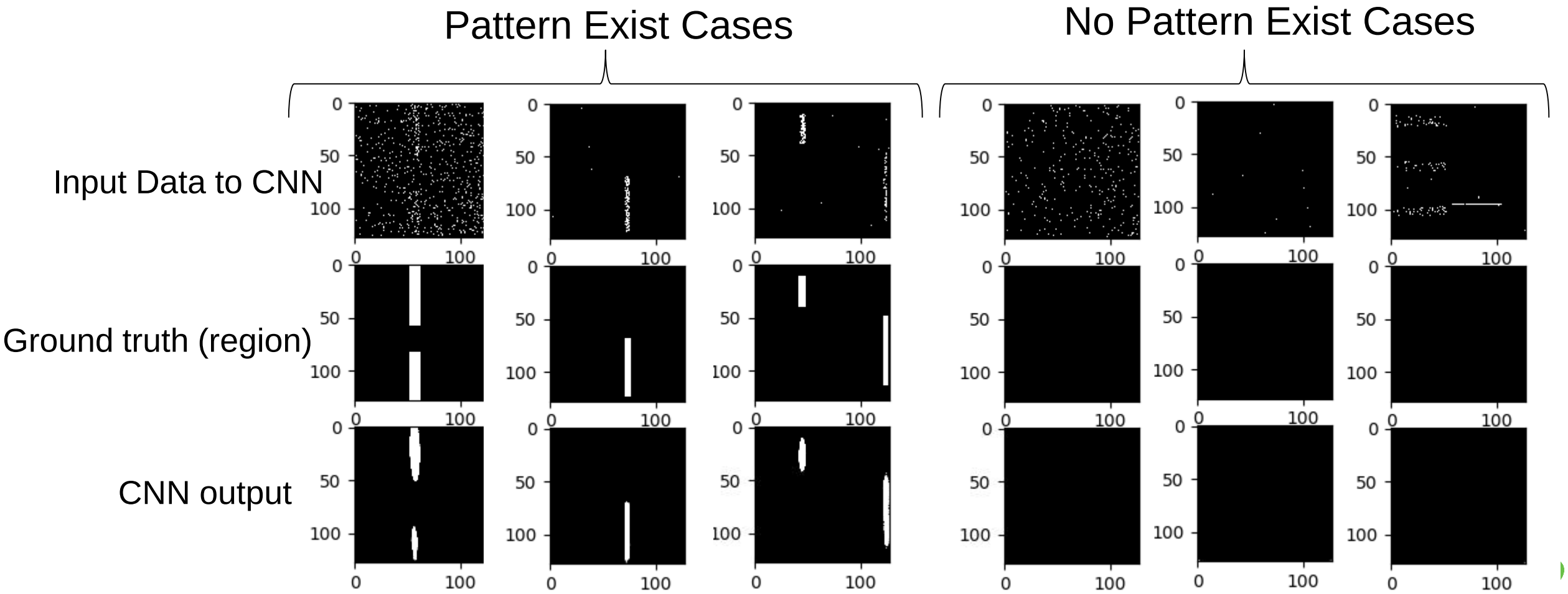
Method 2: Manual Feature Engineering

Classification Scheme



Pattern trained image segmentation

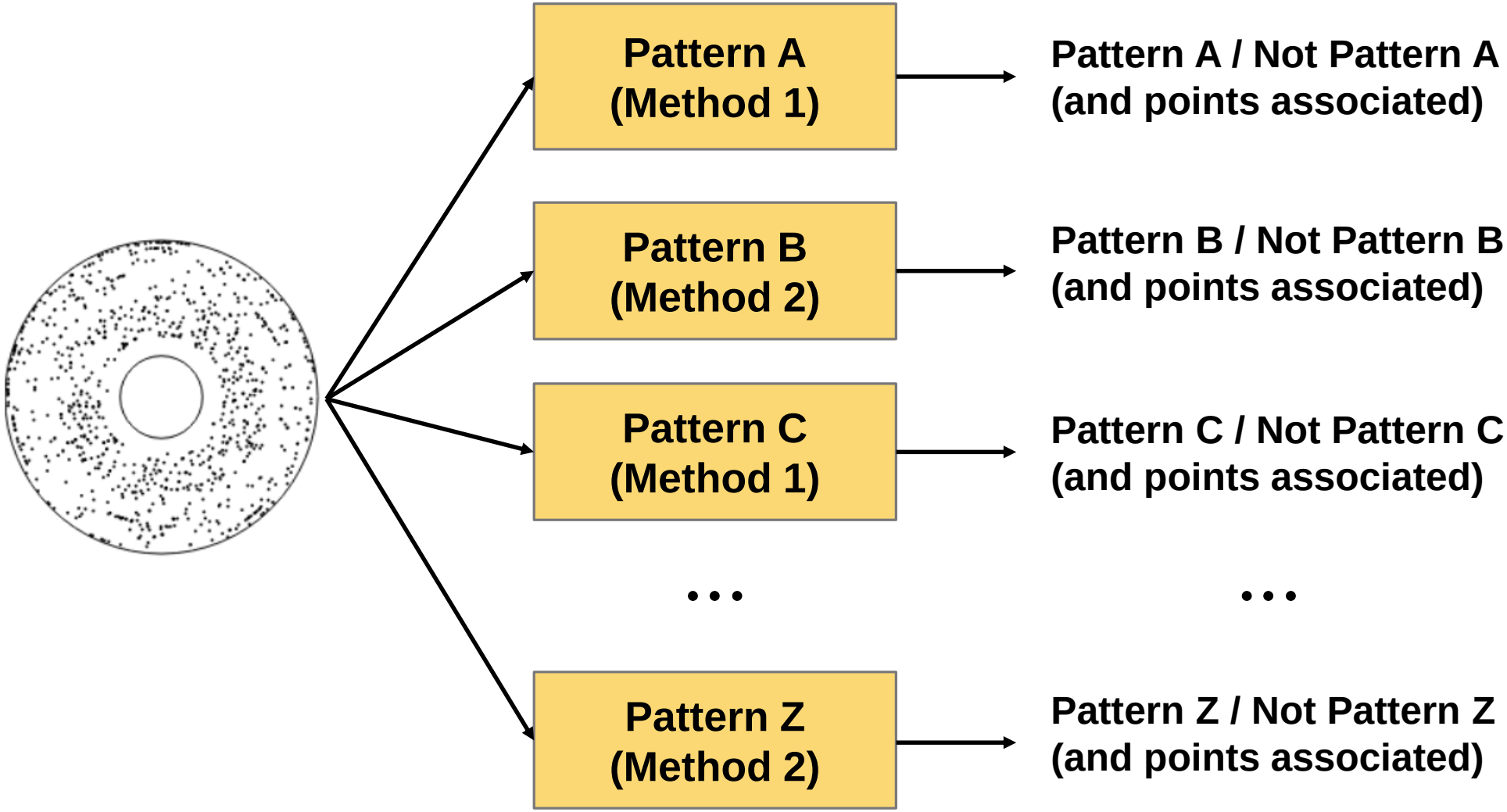




Method 2: Automatic Feature Engineering

- CNN trained with synthetic data (100K images)
- Validated with real and synthetic Data
- Simple to create models and maintain (just add/replace with new model)
- Improved accuracy with CNN
- Needs GPU or High Power CPU to perform calculations quickly

Hybrid Solution



Hardware

GPU Computer

- 2x NVIDIA Titan X Pascal GPUs (12 GB memory & 3584 cores each)
- 32 GB DDR4 3000 RAM
- 30 TB Hard Drive Space
- Intel Core i7-7700K 4.2 CPU
- 1000W Power Supply

Software

On Ubuntu 16.04

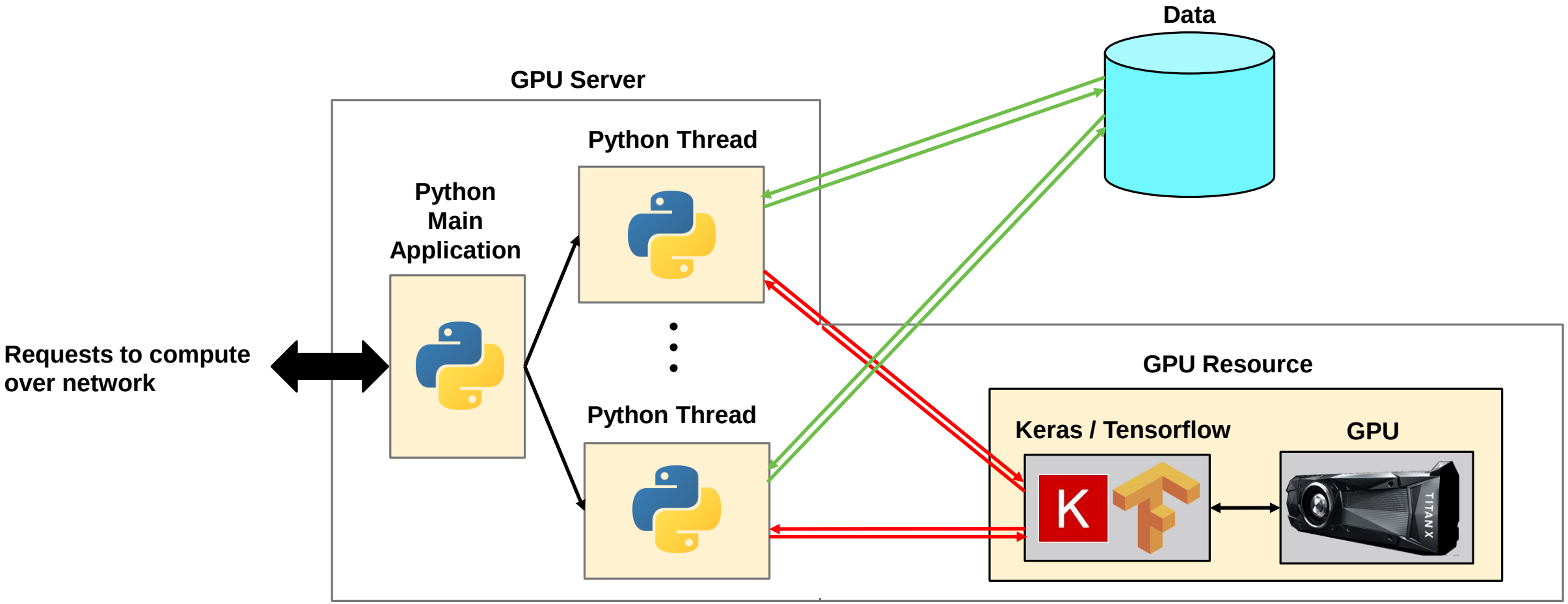
KERAS

TENSORFLOW

PYTHON 2.7.x or 3.5

NVIDIA CUDA TOOLKIT and cuDNN Library

Implementation Details



Results

Real Data Validation Results

CNN Models and Codes: 07072017-3

Pattern	Accuracy
BAND	97.76%
CIRCUMFERENTIALSCRATCH	98.13%
CLUSTER	95.26%
HALFGALAXY	96.76%
NONE	99.50%
RADIAL	90.77%
SKEW	94.89%
GALAXY	85.91%
CIRCUMFERENTIAL	78.43%

- **Synthetic data didn't work well for some defect pattern classes**
- **Method is suitable for new defect pattern classes**
- **Management of models : tradeoff between memory/storage and retraining**
- **Some defect pattern classes may not be suitable for CNN when higher resolution scans are possible**
- **Future work:**
  - **Grouping defect patterns in different models**
  - **Reducing size of models**
  - **Improve synthetic data generation for some defect patterns**

Questions?