



**UNAM**



**iimas**

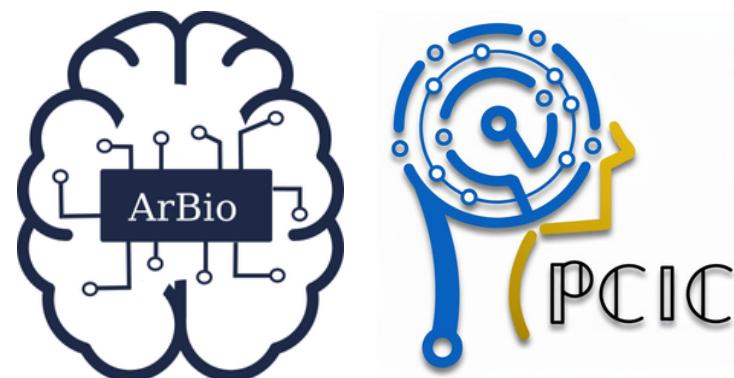
**INSTITUTO DE  
INVESTIGACIONES  
EN MATEMÁTICAS  
APLICADAS Y  
EN SISTEMAS**

## **XLVIII Congreso Nacional de Ingeniería Biomédica**

# **DATA PREPROCESSING**

**Presenta:**

**ISC. Gabriel Carcedo Rodríguez**  
**Ing. Ismael Pérez Ruiz**  
**LCC. Victoria May Balam**

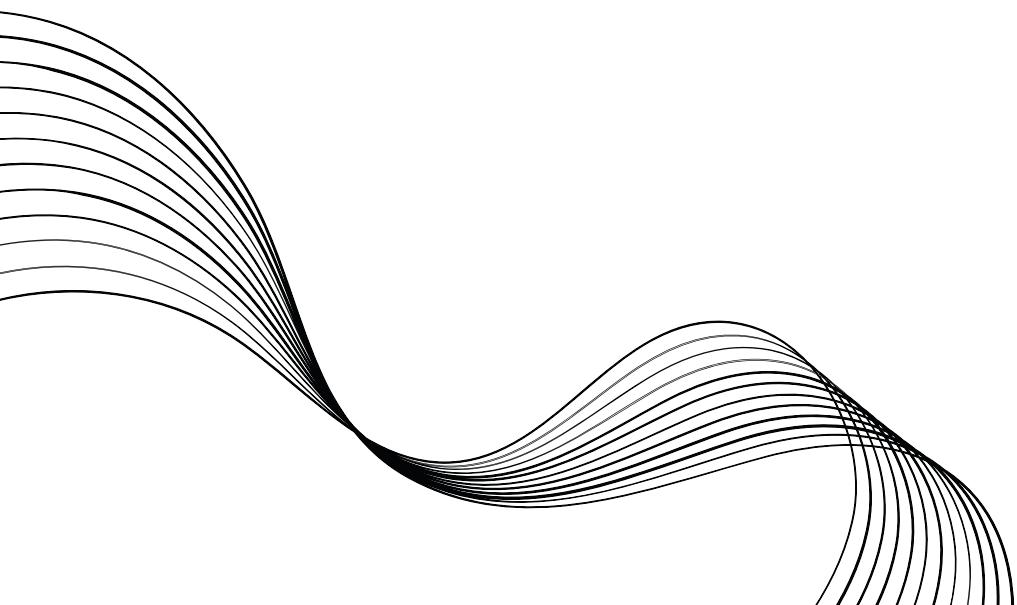


**PCIC, IIMAS, UNAM**  
**Unidad Académica en Mérida, Yucatán**

**Monterrey, 2025**

# Content

1. Data Cleaning
2. Scaling
3. Features Selection
4. Features Extraction
5. Dimensionality Reduction
6. Dataset split



# Introduction

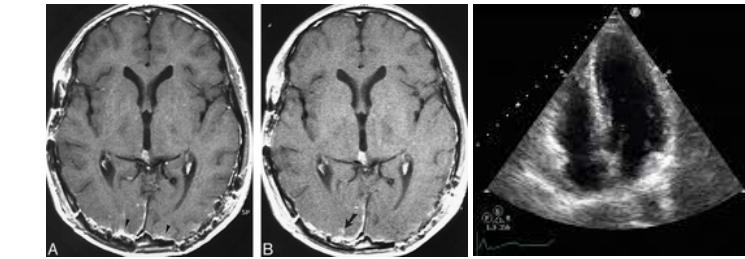
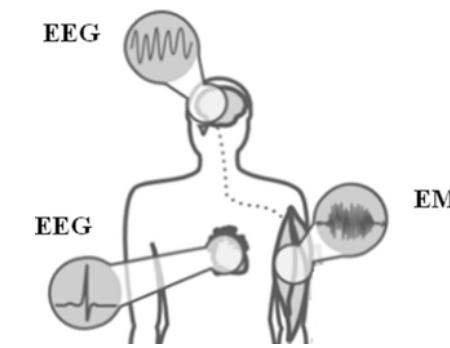
**Data analysis is  
crucial in healthcare**

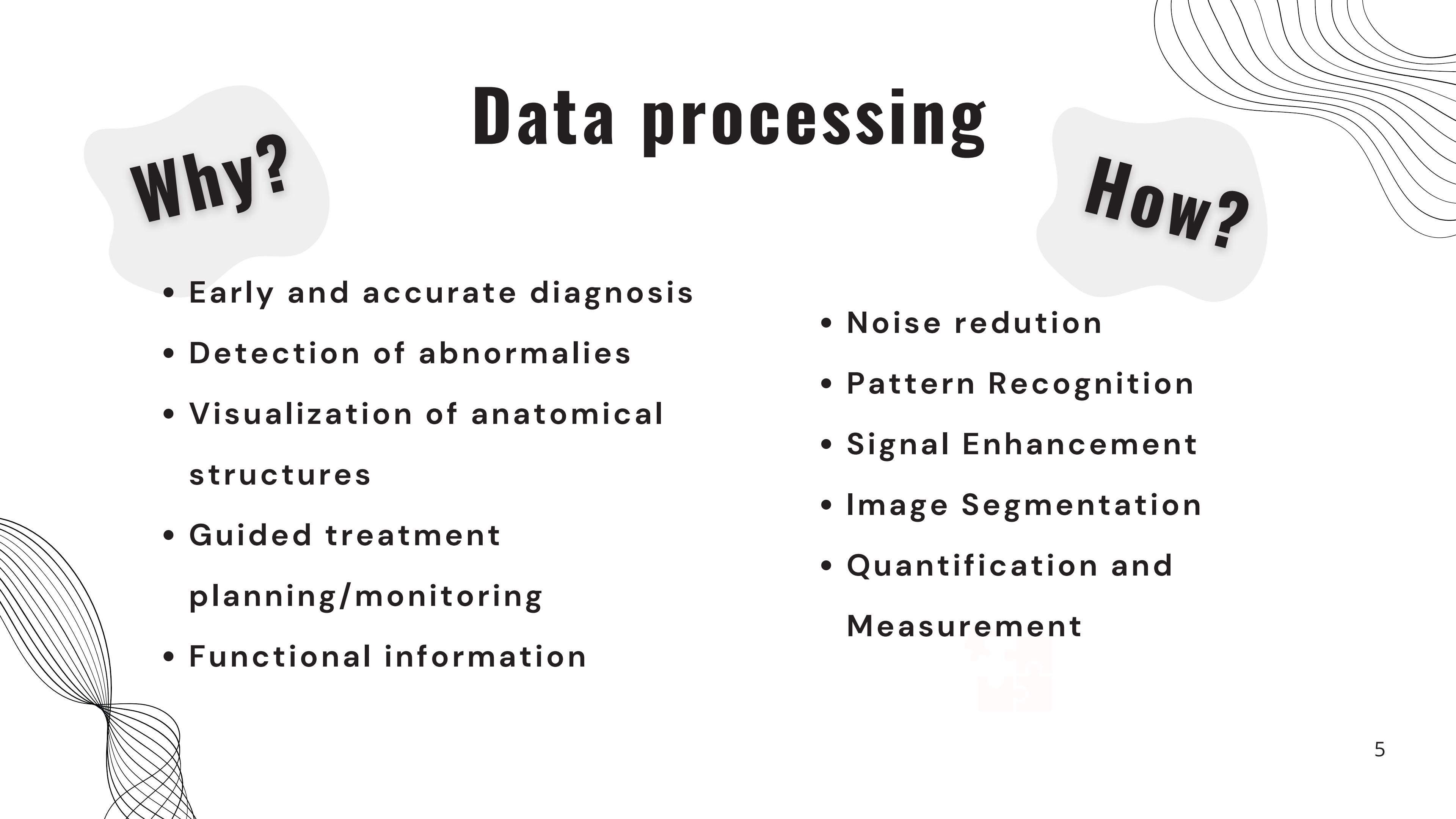


- Informed decision-making
- Early disease detection
- Predictive analytics
- Clinical research and drug development
- Healthcare cost management
- Personalized medicine/treatments
- Quality improvement

# Types of Medical Data

- Signals
- Images
- Genomic data
- Patient behavior data
- Monitoring Devices
- Biometric data
- Social / Environmental





# Why?

- Early and accurate diagnosis
- Detection of abnormalities
- Visualization of anatomical structures
- Guided treatment planning/monitoring
- Functional information

# Data processing

# How?

- Noise reduction
- Pattern Recognition
- Signal Enhancement
- Image Segmentation
- Quantification and Measurement

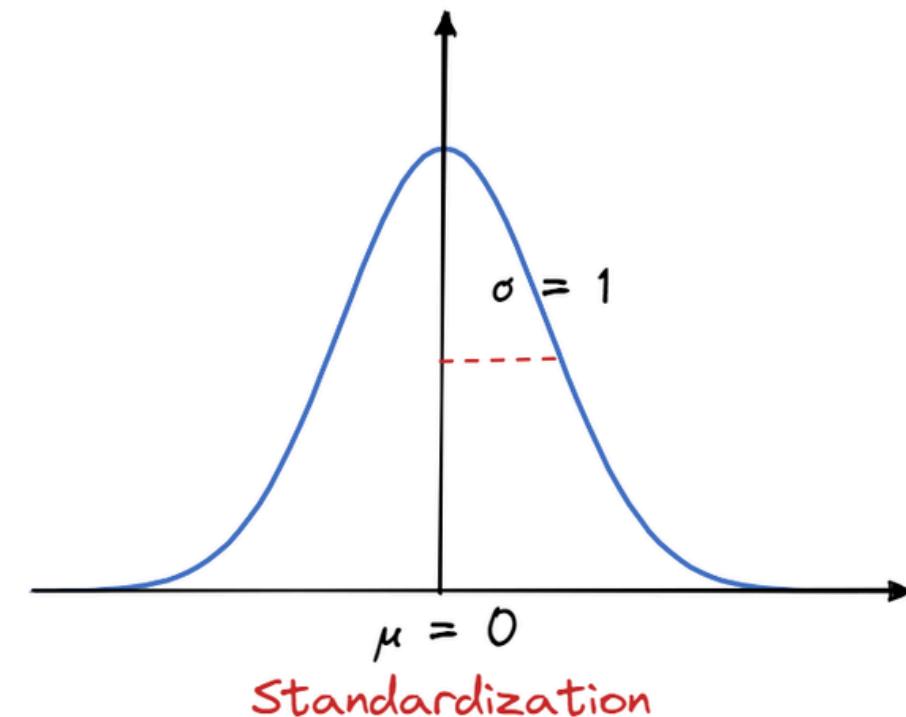
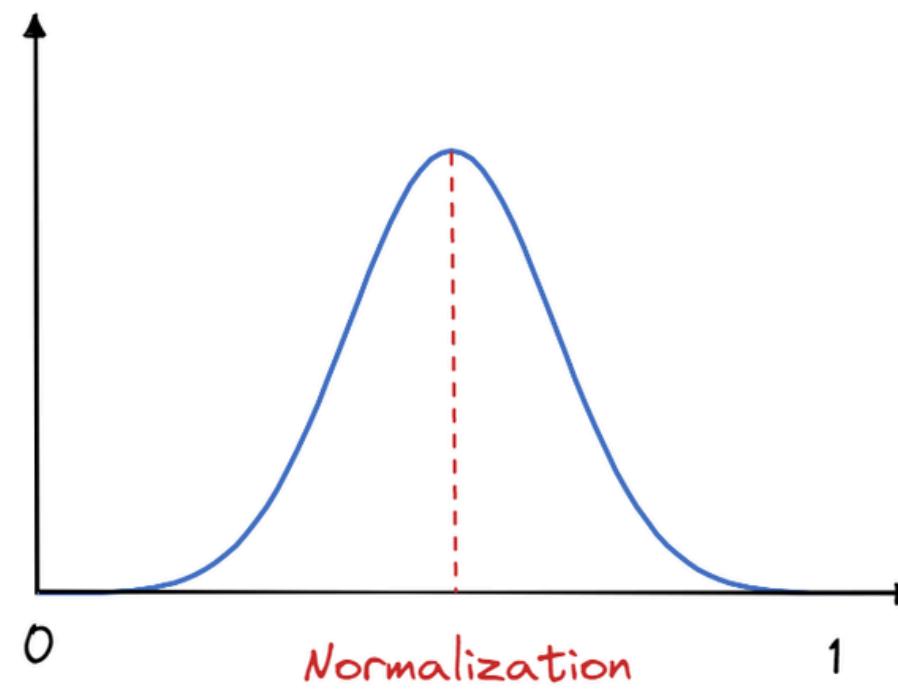
# Data Cleaning

- Missing values
- Outliers
- Inconsistent formats

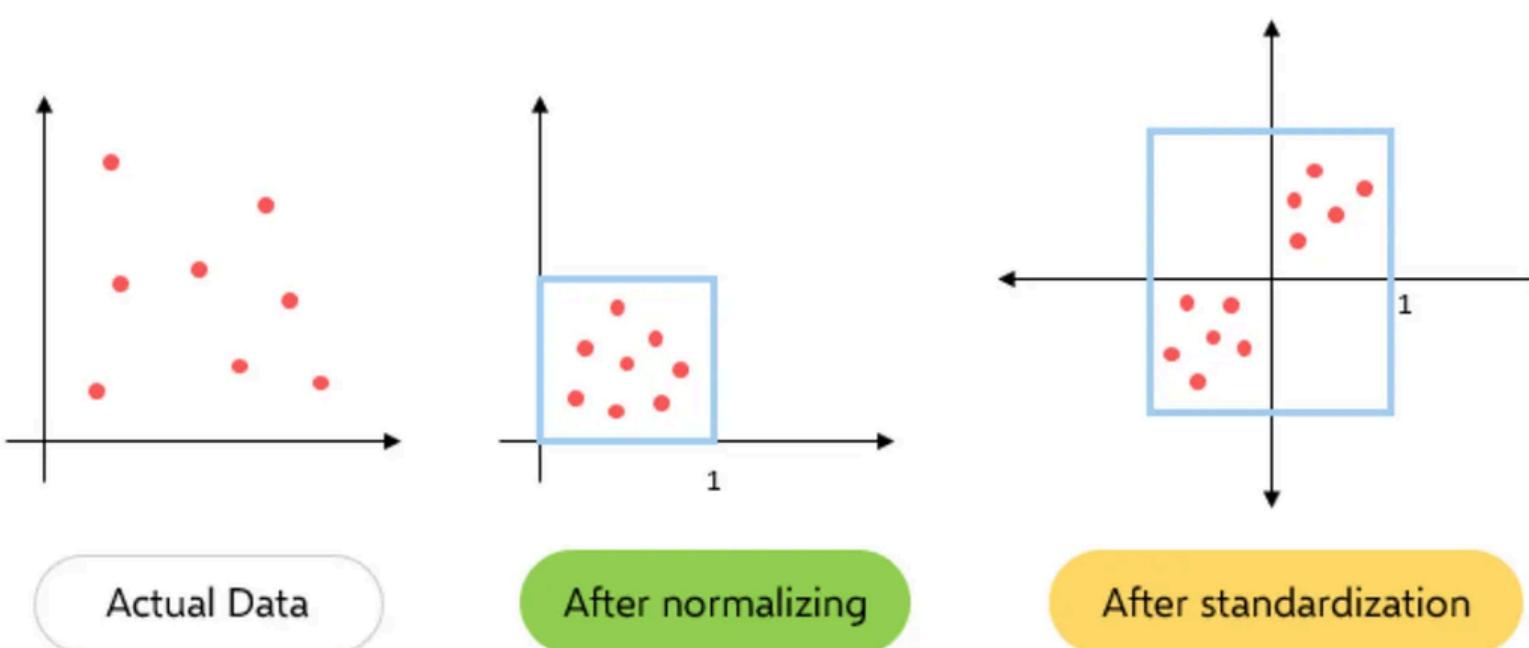


# Scaling

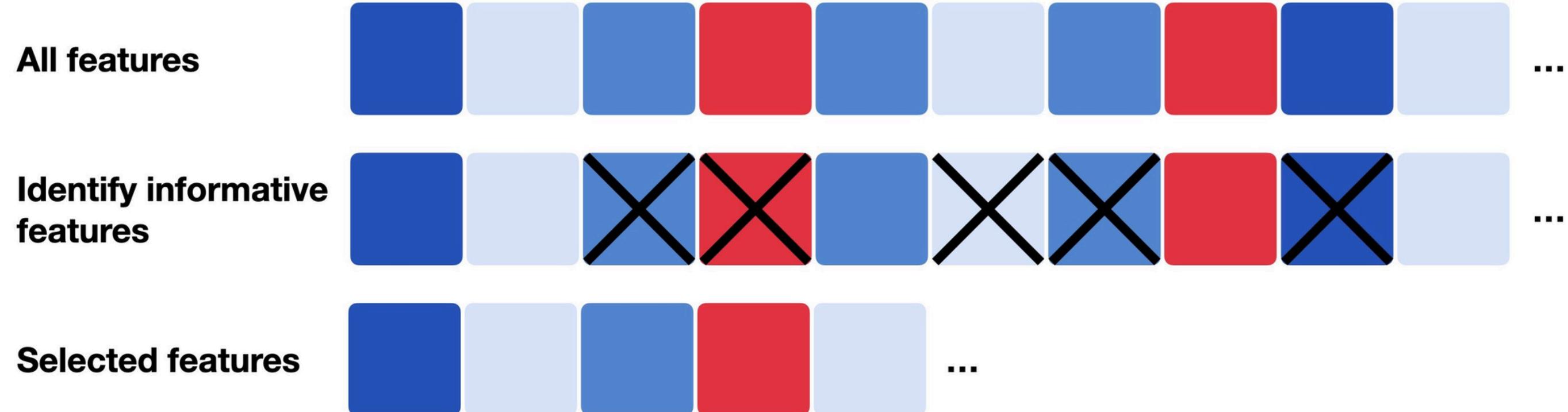
$$X_{new} = \frac{X - X_{min}}{X_{max} - X_{min}}$$



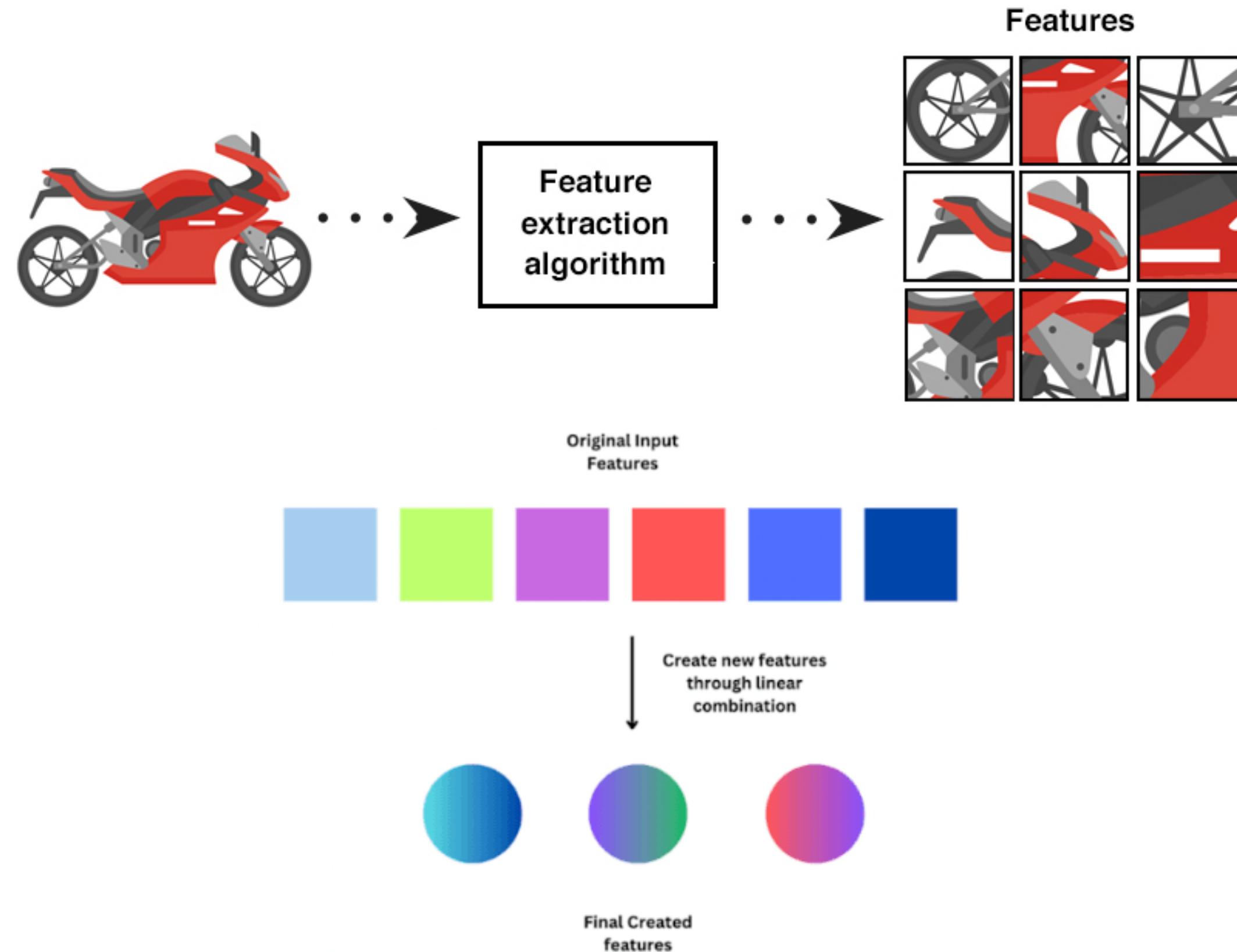
$$X' = \frac{X - \text{Mean}}{\text{Standard deviation}}$$



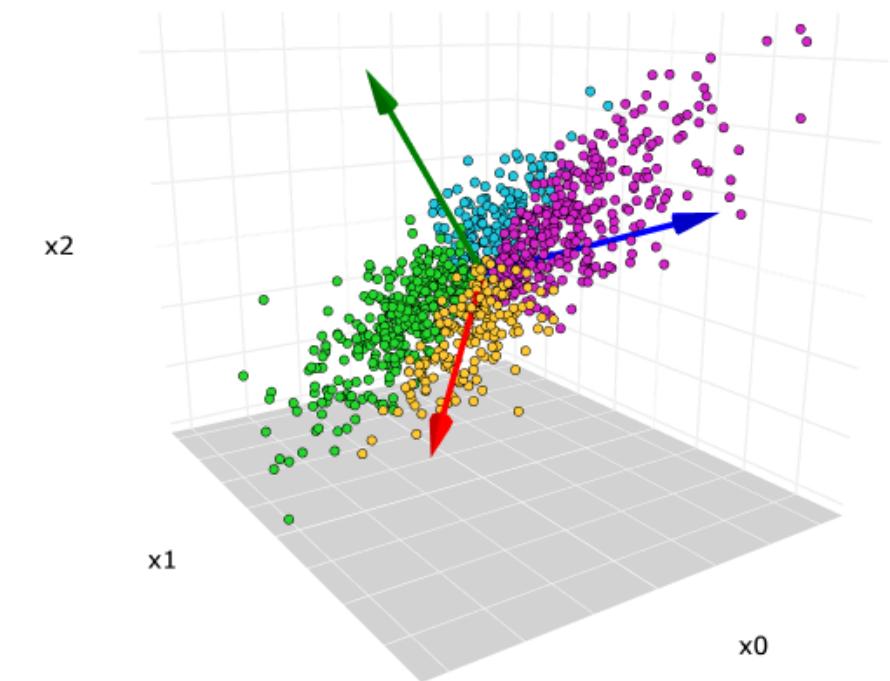
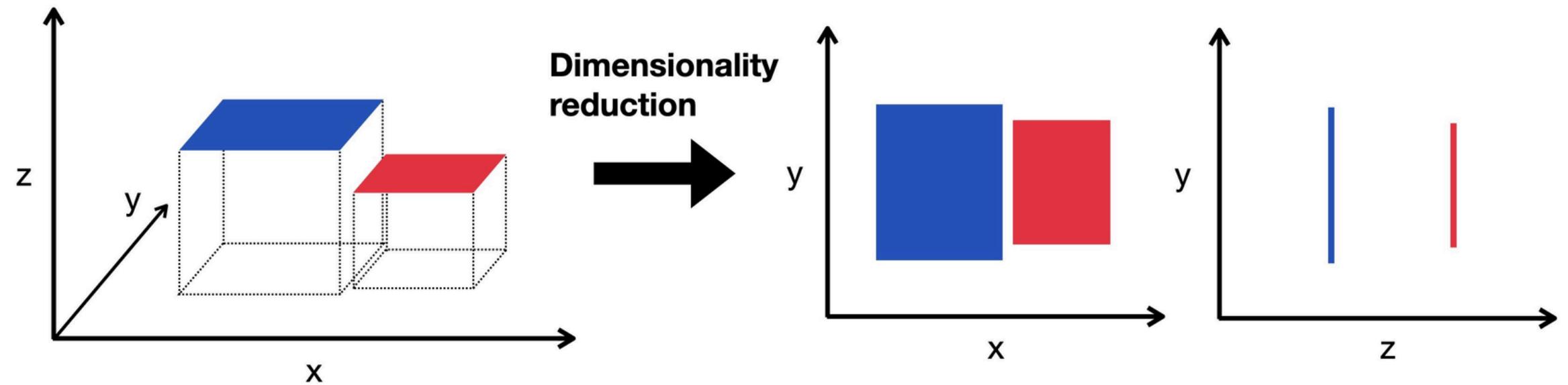
# Feature Selection



# Feature Extraction



# Dimensionality reduction



- Eigenvalues / Eigenvectors
- Singular Value Decomposition (SVD)
- CUR Decomposition
- Principal Component Analysis (PCA)

# Dataset split

