Precision Medicine

Alessandra Marasi: alessandra.marasi@polimi.it

Francesca Fati: francesca.fati@polimi.it

Sofia Breschi: sofia.breschi@polimi.it

Eriberto Andrea Franchi: eribertoandrea.franchi@polimi.it

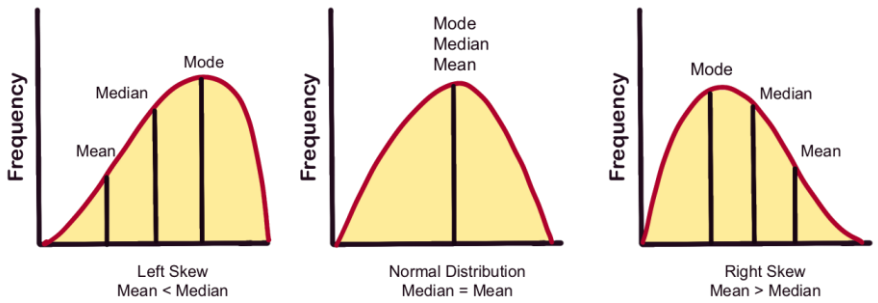
Agenda

- Precision Medicine
- Data in Healthcare
- Medical AI: Model-centric vs Data-centric AI
- **Data Processing**
- Hands-On: Heart Disease Dataset

Data Processing

Descriptive Statistics

Distribution



Central Tendency

Mean

Median

Mode

Variability

Range

Standard Deviation

Variance

Interquartile Range (IQR)

Data Processing

Descriptive Statistics

Correlation Analysis

- It refers to the degree of association or relationship between two variables.

  Measure of the extent to which changes in one variable are accompanied by changes in another variable
- Used to explore relationships between health-related variables.
- Provides insights into factors influencing health outcomes and disease progression

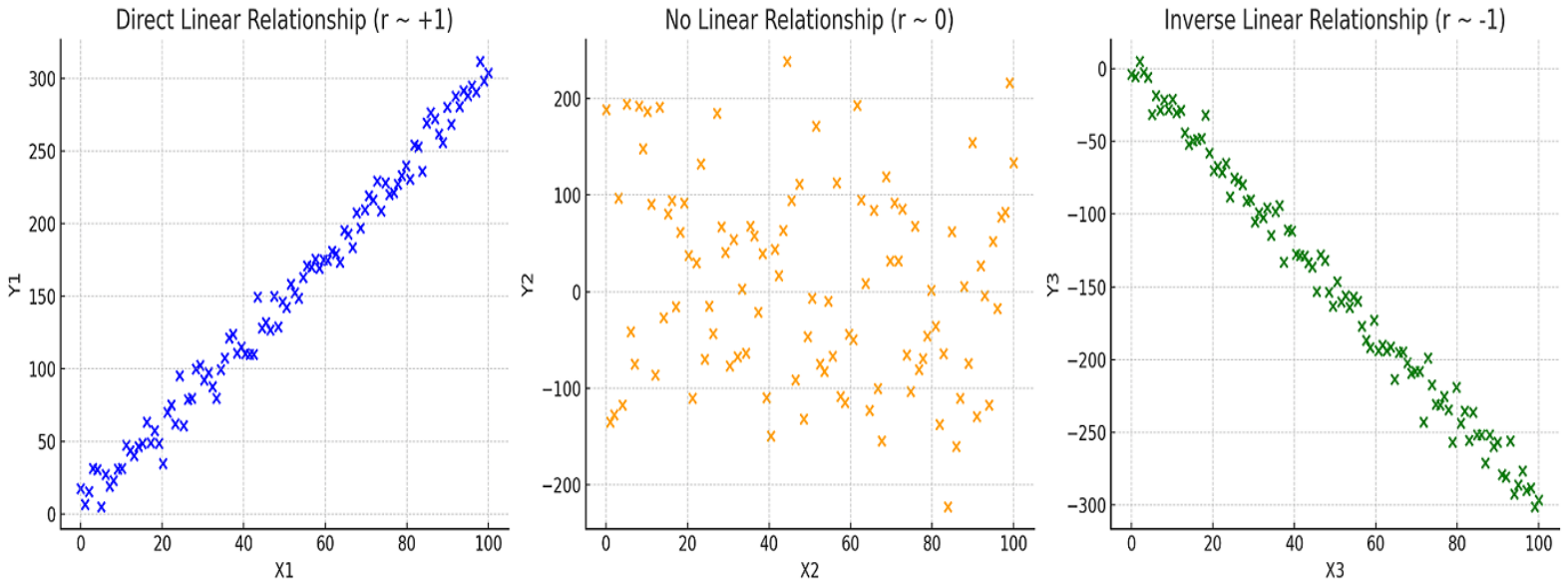
Data Processing

Descriptive Statistics

Correlation Analysis

Pearson Correlation Coefficient (r)

- Ranges from -1 (negative correlation) to 1 (positive correlation)
- r = 0: No correlation.



Data Processing

Descriptive Statistics

Correlation Analysis

Spearman's Rank Correlation Coefficient (ρ)

- Non-parametric measure of correlation.
- Assesses strength and direction of association between ranked variables.
- Suitable for ordinal or ranked data.

Kendall's Tau (τ)

- Non-parametric measure of correlation.
- Evaluates strength and direction of association based on concordant and discordant pairs.
- Useful for identifying relationships in data with tied ranks.

Descriptive Statistics

Data Distribution – Z-score

The z-score is a statistical measure that indicates how many standard deviations a data point is from the mean of the dataset

$$z = \frac{x - \mu}{\sigma}$$

X = Value of data point

μ = Mean of the dataset

σ = Standard deviation of the dataset

Data Processing

Descriptive Statistics

Data Distribution – Z-score

- A Z-score of 0 indicates that the data point is exactly at the mean of the dataset
- Positive Z-scores indicate that the data point is above the mean
- Negative Z-scores indicate that the data point is below the mean
- Z-scores help clinicians identify abnormal values and assess a patient's risk relative to the general population

Hands-On: Heart Disease Dataset

Stroke Dataset

https://tinyurl.com/3ywaw4wr

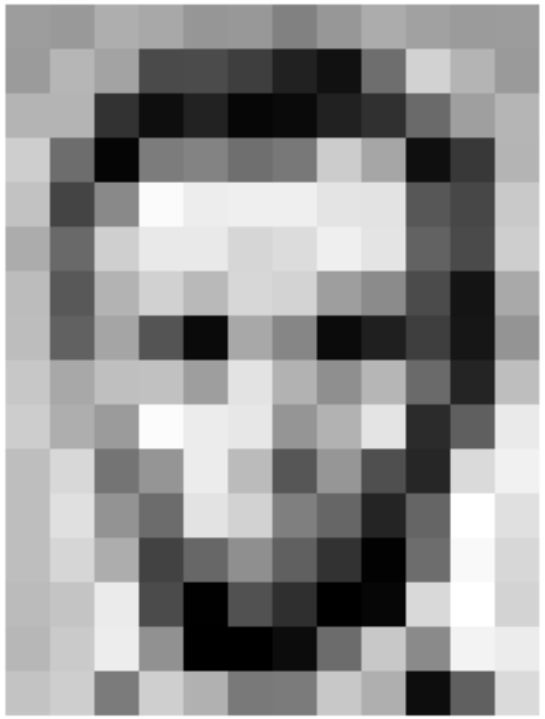
Data Processing

Medical Images

- Image Characteristics dimension (width, height, depth), pixel intensity distribution, image resolution
- Data Quality: signal-to-noise ratio, contrast, sharpness
- Pre-Processing Techniques: intensity normalization, contrast enhancement, reshape

Data Processing

Medical Images

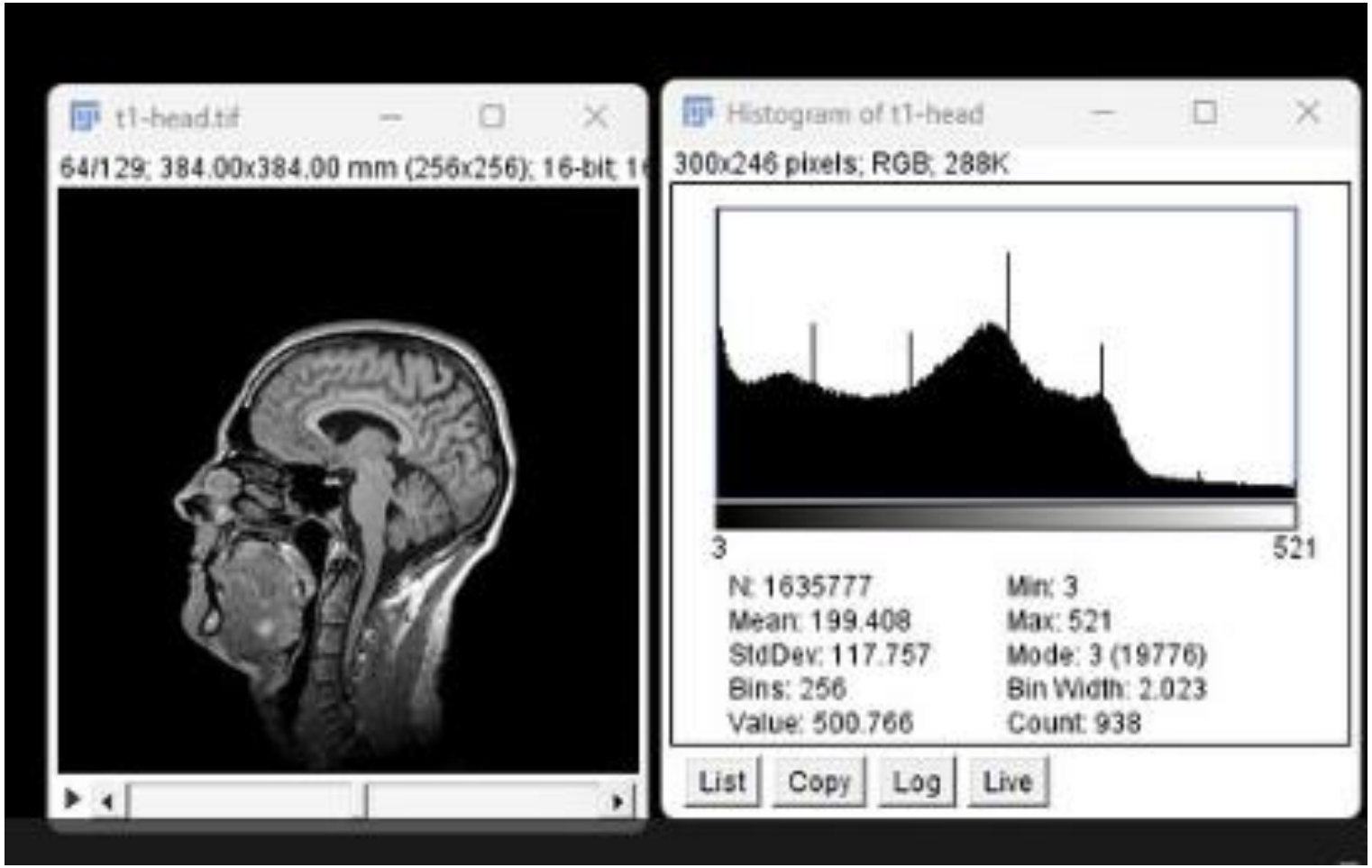


157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

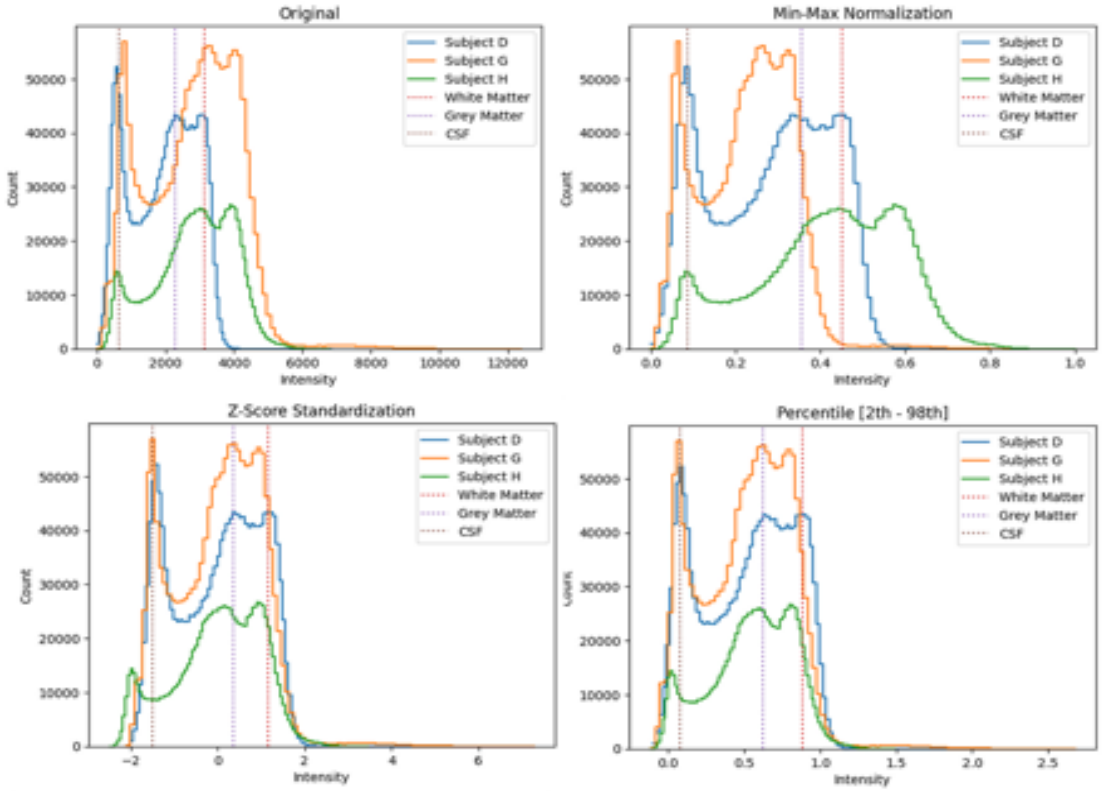
Data Processing

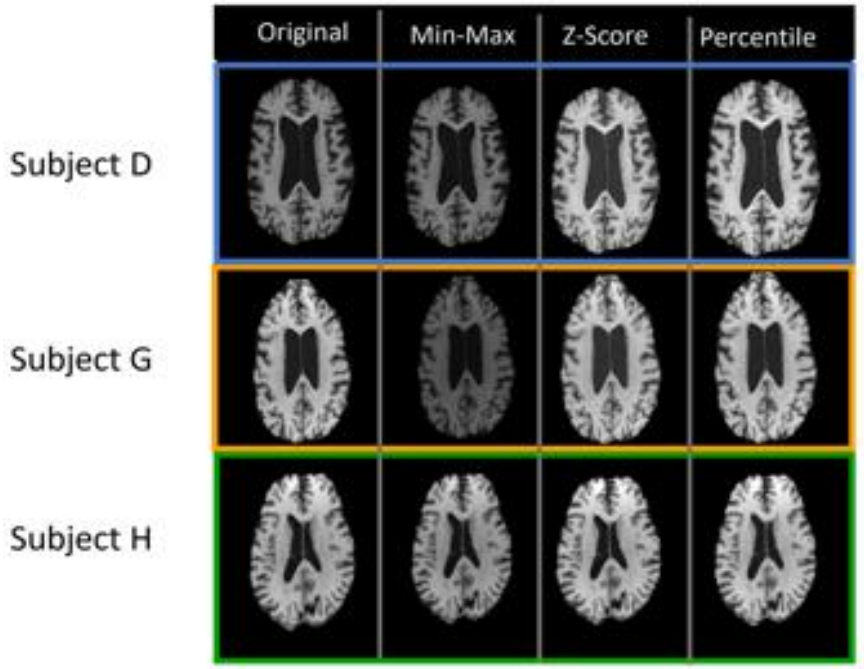
Medical Images



Data Processing

Medical Images

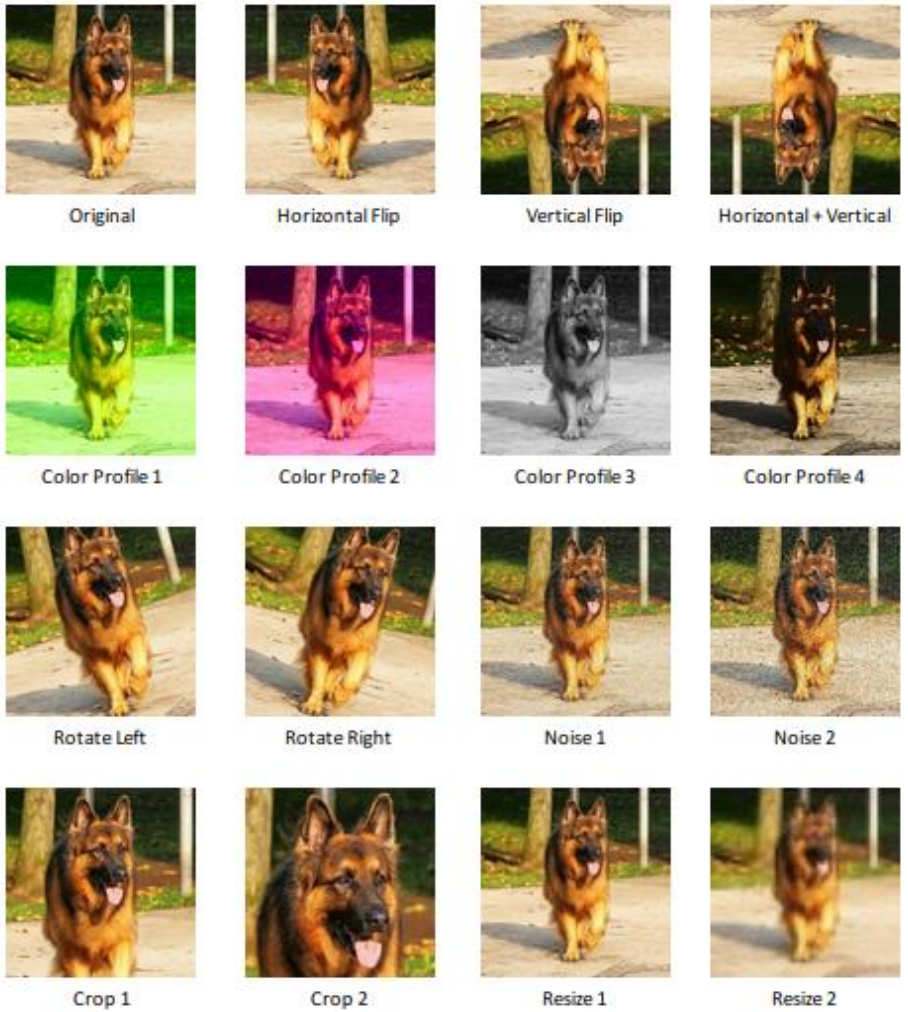




Data Processing

- Data augmentation involves creating new training samples from existing data by applying various transformations
- The synthetic variations introduced should reflect the inherent variability present in real-world data
- In medical imaging and signal processing, data augmentation techniques play a crucial role in enhancing model generalization and robustness

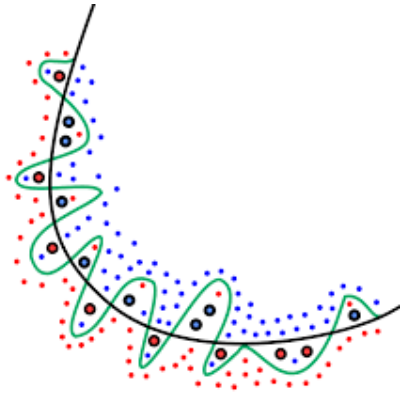
Data Processing



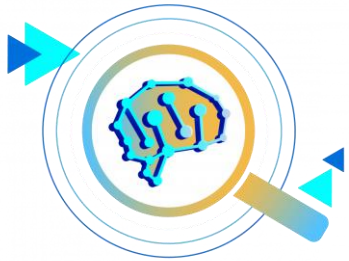
Data Processing

Purposes of Data Augmentation

Enlarge Dataset



Prevent Overfitting



Improve Model Accuracy

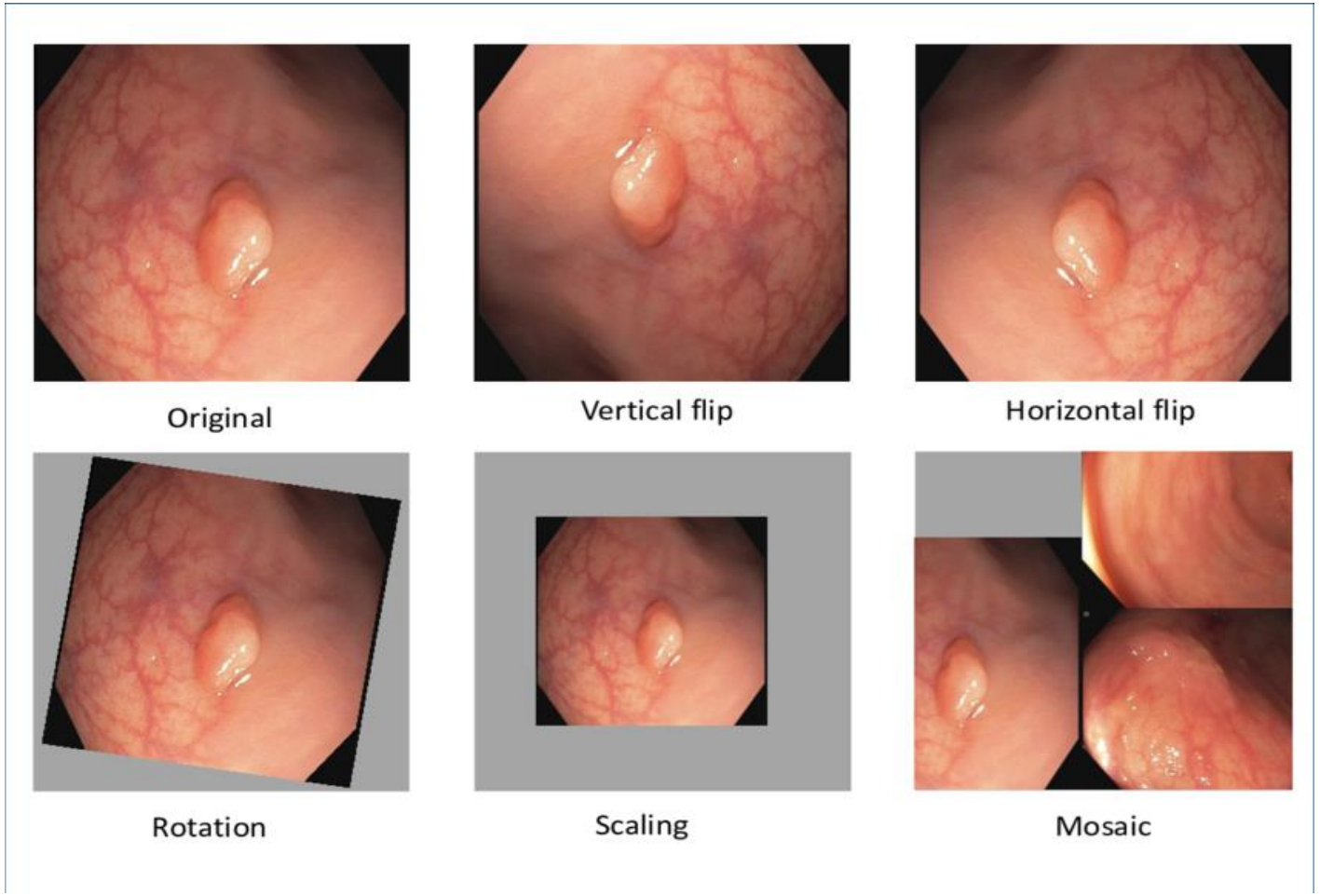
Data Processing

Endoscopy

- Endoscopic images capture internal body structures using a flexible tube with a camera.
- Data augmentation techniques include:
  - **Rotation**: Simulating different viewing angles
  - **Flip**: Mirroring images to account for variations in orientation
  - **Crop**: Cropping regions of interest to focus on specific structures
  - **Color augmentation**: Adjusting brightness, contrast, and hue to simulate lighting variations

Data Processing

Endoscopy



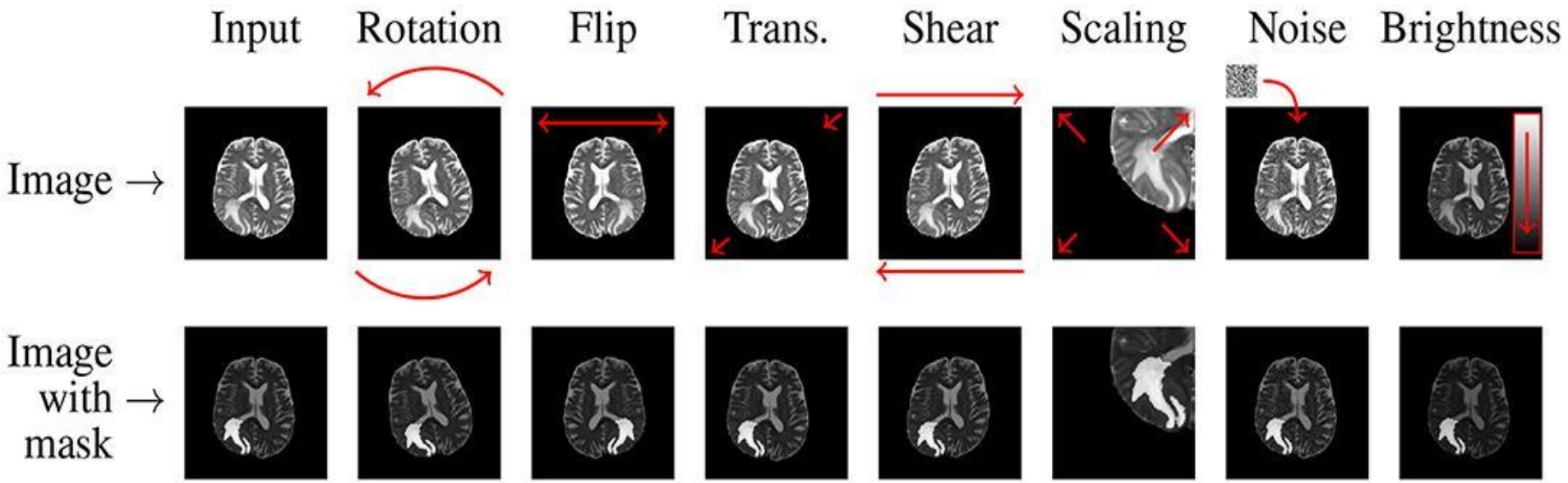
Data Processing

Medical Imaging

- CT (Computed Tomography) and MRI (Magnetic Resonance Imaging) provide detailed anatomical images.
- Data augmentation techniques include
  - **Rotation**: Rotating images to simulate different viewing angles.
  - **Flip**: Mirroring images to account for variations in orientation.
  - **Zoom**: Scaling images to simulate changes in field of view.
  - **Elastic deformation**: Distorting images to simulate anatomical variations

Data Processing

Medical Imaging



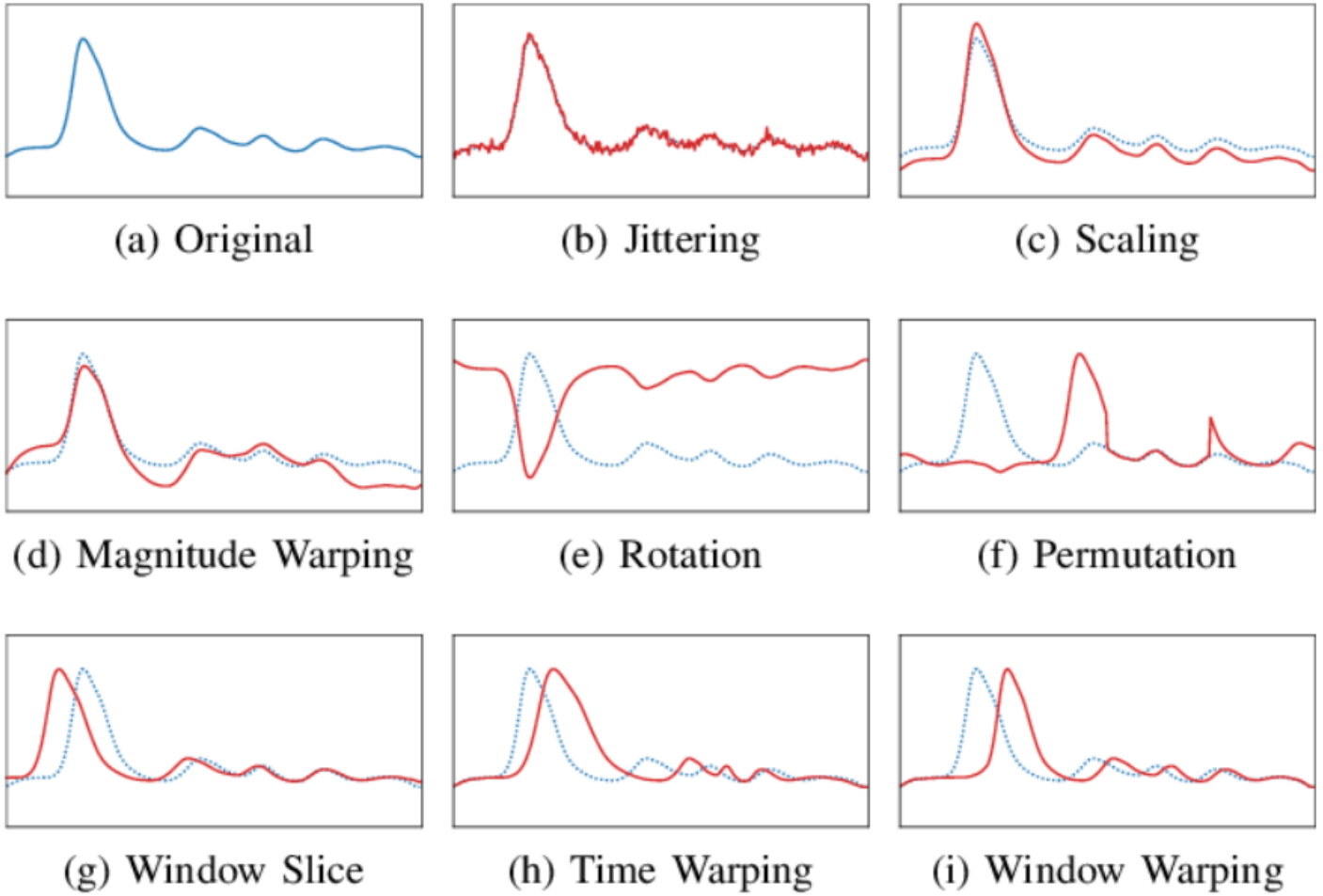
Data Processing

Biomedical signals

- Biomedical signals, such as electrocardiogram (ECG) and electromyogram (EMG), provide physiological information
- Data augmentation techniques include:
  - **Time shifting**: Shifting signals in time to simulate temporal variations
  - **Amplitude scaling**: Adjusting signal amplitudes to simulate intensity changes
  - **Noise injection**: Adding random noise to simulate measurement variability
  - **Signal cropping**: Extracting segments of signals to focus on specific events or patterns

Data Processing

Biomedical signals



Hands-On: Heart Disease Dataset

Data Augmentation: Chest Xrays

https://tinyurl.com/3ywaw4wr