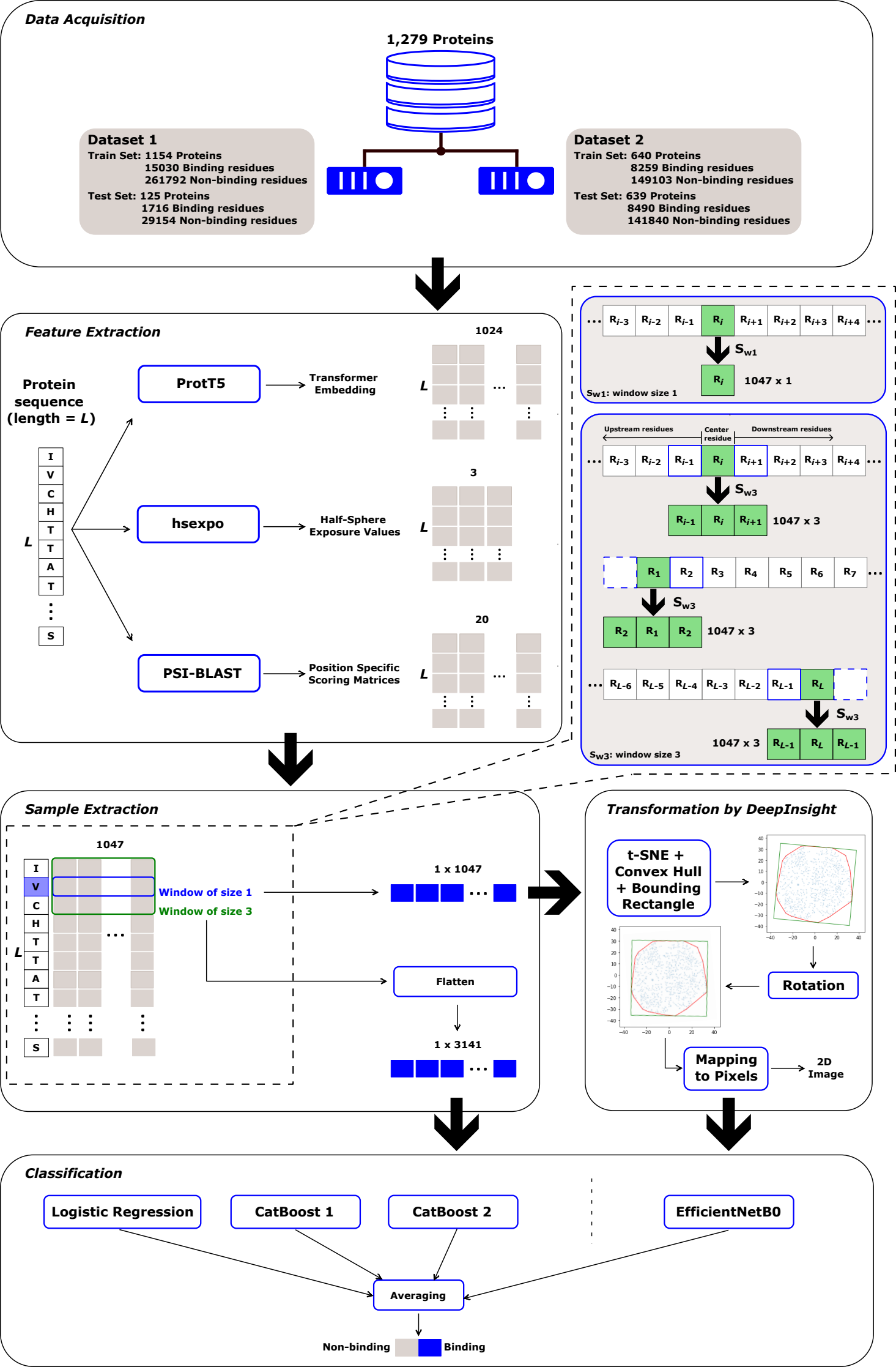
Data Acquisition
1,279 Proteins
Dataset 1
Train Set: 1154 Proteins
15030 Binding residues
261792 Non-binding residues
Test Set: 125 Proteins
1716 Binding residues
29154 Non-binding residues
Dataset 2
Train Set: 640 Proteins
8259 Binding residues
149103 Non-binding residues
Test Set: 639 Proteins
8490 Binding residues
141840 Non-binding residues
Feature Extraction
Protein sequence (length = L)
ProtT5
Transformer Embedding
1024
hsexpo
Half-Sphere Exposure Values
3
PSI-BLAST
Position Specific Scoring Matrices
20
Sw1
1047 x 1
Sw1: window size 1
Upstream residues
Center residue
Downstream residues
Sw3
1047 x 3
Sw3: window size 3
Sample Extraction
1047
Window of size 1
Window of size 3
1 x 1047
Flatten
1 x 3141
Transformation by DeepInsight
t-SNE + Convex Hull + Bounding Rectangle
Rotation
Mapping to Pixels
2D Image
Classification
Logistic Regression
CatBoost 1
CatBoost 2
EfficientNetB0
Averaging
Non-binding
Binding