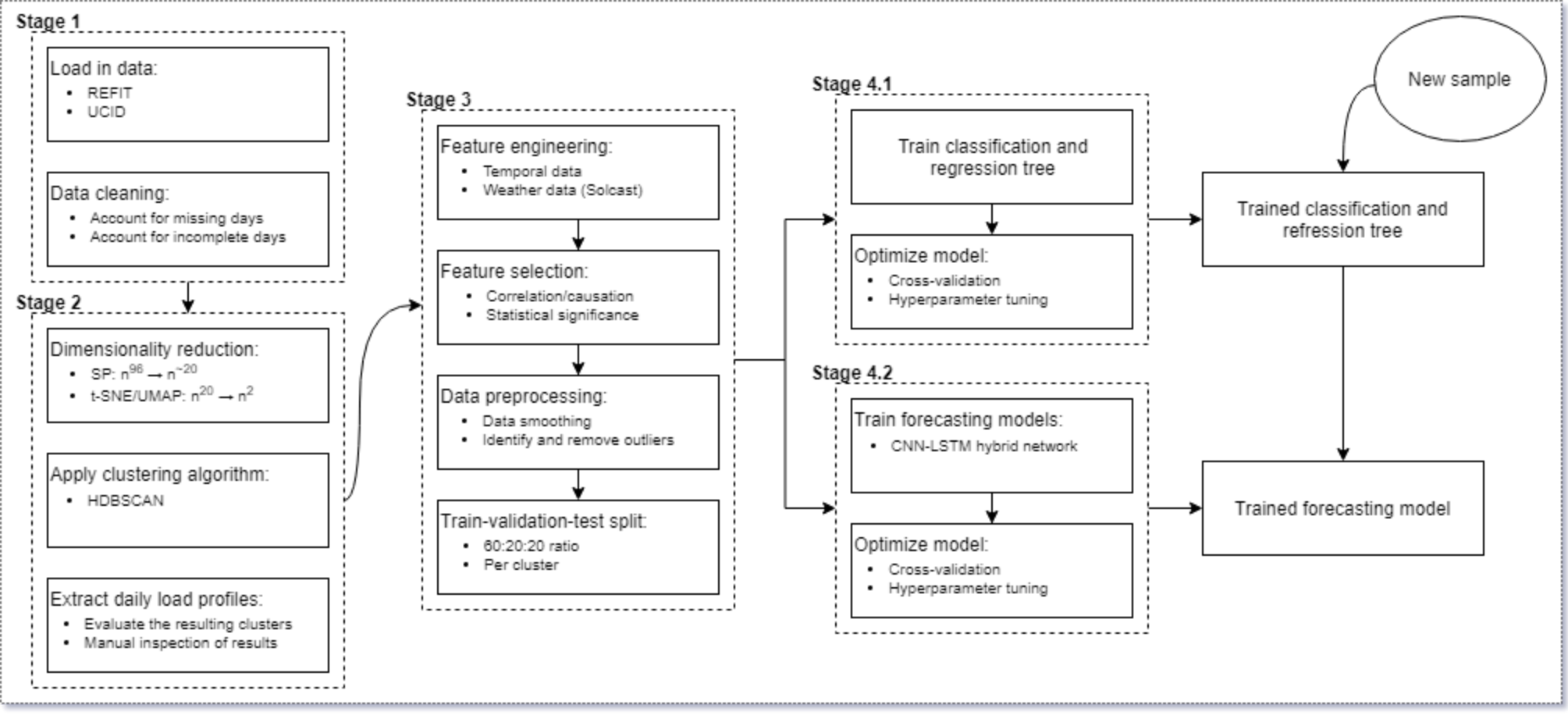
Stage 1
Load in data:
REFIT
UCID
Data cleaning:
Account for missing days
Account for incomplete days
Stage 2
Dimensionality reduction:
SP: n^96 → n^~20
t-SNE/UMAP: n^20 → n^2
Apply clustering algorithm:
HDBSCAN
Extract daily load profiles:
Evaluate the resulting clusters
Manual inspection of results
Stage 3
Feature engineering:
Temporal data
Weather data (Solcast)
Feature selection:
Correlation/causation
Statistical significance
Data preprocessing:
Data smoothing
Identify and remove outliers
Train-validation-test split:
60:20:20 ratio
Per cluster
Stage 4.1
Train classification and regression tree
Optimize model:
Cross-validation
Hyperparameter tuning
Stage 4.2
Train forecasting models:
CNN-LSTM hybrid network
Optimize model:
Cross-validation
Hyperparameter tuning
New sample
Trained classification and refression tree
Trained forecasting model